Faculté de psychologie, sciences de l'éducation et de la formation
Université de Lille

University of Lille

École doctorale Sciences de l'homme et de la société

Trigone-CIREL

Doctoral Thesis

By Sergio Iván RAMÍREZ LUELMO

Sciences de l'Éducation et de la Formation
(Section 70)

Automatic flow (optimal learning experience) detection in a MOOC via Machine Learning

Flow & Learning Analytics

Under the supervision of Professor (PU) Jean Heutte and Professor (PU) El Mawas Nour

September 21st, 2023

Jury members

M. Mutlu CUKUROVA, AP, University College London — Examinateur
Mme. Nour EL MAWAS, PU, CREM-Pixel, Université de Lorraine — Co-encadrante
M. Fabien FENOUILLET, PU, CREF, Université Paris-Nanterre — Rapporteur
M. Jean HEUTTE, PU, CIREL-Trigone, Université de Lille — Directeur
M. Bruno POELLHUBER, Prof., CRIFPE, Université de Montréal — Président
Mme. Annelies RAES, Prof., ITEC, KU Leuven — Examinatrice
Mme. Margarida ROMERO, PU, LINE, Université Côte-d'Azur — Rapportrice

THESIS

Affidavit

Je soussigné, Ramírez Luelmo Sergio Iván, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de M. Jean Heutte (directeur) et de Mme. Nour El Mawas (co-directrice), dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect de la charte nationale de déontologie des métiers de la recherche.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Lille, le 21 septembre 2023

Affidavit

I, undersigned, Ramírez Luelmo Sergio Iván, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific direction of M. Jean Heutte (thesis director) and of Mme. Nour El Mawas (co-thesis director), in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with the French charter for Research Integrity.

This work has not been submitted previously either in France or abroad in the same or in a similar version to any other examination body.

Lille, September 21st, 2023

À mon fils

Acknowledgments

Je tiens tout d'abord à remercier mon directeur de thèse, Jean Heutte qui m'a accueilli chaleureusement au sein de l'équipe de recherche Trigone dans le laboratoire CIREL depuis mon arrivée. Tout au long de ces quatre années, j'ai pu apprécier sa vision sur la recherche autour du flow, son intégrité, sa disponibilité malgré son calendrier chargé, et sa générosité à partager ses connaissances. Je voudrais lui témoigner ici toute ma gratitude de m'avoir guidé et soutenu durant toutes ces années.

Je remercie énormément ma co-encadrante Nour El Mawas pour le vif intérêt qu'elle a porté à mon travail et ses nombreux et très pertinents conseils, mais surtout pour son soutient et sa patience durant toutes nos réunions. Cette thèse n'aurait pas vu le jour sans son inestimable contribution.

Ce projet de recherche n'aurait pas été possible sans le soutien technique de Rémi Bachelet, qui m'a facilité l'accès aux données qui font partie centrale de cette thèse, ni sans le soutien financier de l'État français dans le cadre du Programme « Investissements d'avenir » (I-SITE ULNE / ANR-16-IDEX-0004 ULNE) géré par l'Agence Nationale de la Recherche française (ANR).

Je remercie aussi les membres du jury pour avoir accepté d'examiner ce travail. Je suis très reconnaissant pour leur précieux temps de congés, sacrifié pour ce travail. Je remercie particulièrement Mme. Margarida Romero et M. Fabien Fenouillet d'avoir accepté de rapporter cette thèse. Je remercie Mme. Annelies Raes, et M. Bruno Poellhubert d'avoir accepté d'examiner ce travail.

Je voudrais aussi exprimer ma reconnaissance aux personnes avec qui j'ai eu le plaisir de partager toutes ces années de recherche, l'équipe Trigone-CIREL, l'équipe administrative du laboratoire CIREL et celle de l'école doctorale SHS.

Un très grand merci à toutes ces personnes dont je garde à toujours un précieux souvenir dans ce beau chapitre de ma vie : mes années de recherche. Nos vifs échanges ont enrichi ma vie bien au-delà de la thèse : Laëtitia, Mihaela, Melina, et Hassina.

Enfin, je souhaite exprimer toute ma reconnaissance à ma famille, qui m'a soutenu tout au long de mon cursus d'études ; à mes parents, qui m'ont toujours encouragé à apprendre.

Abstract

Flow "[...] is a gratifying state of deep involvement and absorption that individuals report when facing a challenging activity [...]." Flow has been shown to correlate reliably and positively to learning-favorable metrics. Csíkszentmihályi (2005) argued that flow promotes learning and personal development because deep and total concentration experiences are intrinsically rewarding.

However, research associating flow in MOOCs is still on the growth while live flow detection remains particularly difficult, as any artifact attempting to detect it, or measure it, inevitably contributes to disrupt it.

Because of the significance of the flow state (as a human psychological state) in the learning process, in online, distant settings, this research work proposes a flow-detecting Machine Learning model that allows for accountable, automatic, and transparent flow detection in a MOOC context. We employ Machine Learning techniques to make sense of multidimensional data without constantly requiring a human expert.

This research work differentiates itself from previous attempts in the use of a theoretical flow model and its accompanying flow measurement instrument, purposefully designed to detect flow directly without intermediate concepts, plus the use of two input datasets of participants' self-reported flow state (n~9 500), along their MOOC log data (~80GB), for a two-year long data collection period. Performance data is neither collected nor employed.

After rigorous data filtering, and pre-processing, we pair real-user data (n~1 500), to their log data aggregations (23 features such as: total number of events, diversity of events, total number of logged sessions, etc.), and to their self-reported flow state to obtain two trained Machine Learning models.

First, a Proof-of-Concept flow-detecting Machine Learning model corroborates the choices and proper coupling of methods, flow measurement instruments, and Machine Learning techniques (F1 = 0.851, AUC ROC = 0.85, Accuracy = 0.797, Precision = 0.821, Recall = 0.882).

An ensuing experiment capitalizes on this knowledge and culminates this research by creating an affordable (F1 = 0.689, AUC PR = 0.87, AUC ROC = 0.68, Accuracy = 0.605, Precision = 0.854, Recall = 0.578), fast (less than a few milliseconds per participant), of negligeable environmental impact (~0.00000237222222 g of CO_2eq per run), automatic (once setup no further intervention is needed by MOOC staff), and transparent (no intervention demanded to MOOC participants) flow-detecting Machine Learning model for use in a MOOC context.

Furthermore, neither model constitutes a black box, facilitating eventual model inspection and understanding.

Both resulting models identify flow better than no-flow: Proof-of-Concept: 58% vs. 22% of total, Prototype: ~44% vs. ~17% of total. However, the Prototype features a higher proportion of False Negatives (~32% of total) vs. True Negatives (~17% of total) when facing unprocessed, unseen data (pre-processing data improves metrics but reconstructs input data). This discrepancy can be interpreted as a cautious model preferring a no-flow classification when in doubt rather than a flow classification, which is not necessarily an unwelcome model behavior.

We hypothesize both imbalanced results are mostly due to 1) the intricate writing style employed in one of the chosen measurement instruments and thus, 2) the imbalance of our input sample. Also, both models carry a lack of granularity when detecting flow, an unsurmountable obstacle intrinsically linked to the granularity (2 moments) of the flow training data.

Perspectives for this research project comprise the successful implementation and commercialization of the Prototype model into a MOOC to 1) implement a flow dashboard; 2) personalize the MOOC's content, activities, and learning-path; and possibly 3) evaluate the incidence of flow detection in MOOC personalization when attempting to reduce MOOC dropout rates.

Résumé

Flow « [...] un état d'épanouissement lié à une profonde implication et au sentiment d'absorption que les personnes ressentent lorsqu'elles sont confrontées à des tâches dont les exigences sont élevées et qu'elles perçoivent que leurs compétences leur permettent de relever ces défis ». L'état de *flow* est positivement corrélé avec des métriques favorables à l'apprentissage. Ainsi, Csíkszentmihályi (2005) affirme que le *flow* favorise l'apprentissage et le développement personnel parce que les expériences de concentration profonde et totale sont intrinsèquement gratifiantes.

Cependant, la recherche associant *flow* dans les MOOC ne fait que débuter alors que la détection du *flow* reste particulièrement complexe, car tout artefact tentant de le détecter ou de le mesurer contribue inévitablement à le perturber.

L'importance de l'état de *flow* (en tant qu'état psychologique humain) dans le processus d'apprentissage, en ligne et à distance, nous pousse à proposer un modèle d'apprentissage automatique de détection de *flow* qui permet une détection de *flow* fiable, automatique et transparente dans un contexte de MOOC. Nous utilisons des techniques d'apprentissage automatique pour donner du sens aux données multidimensionnelles sans avoir recours à un expert humain en permanence.

Ce projet de recherche se différencie des travaux précédents par l'utilisation d'un modèle théorique de flow et de son instrument de mesure, conçus exprès pour détecter *flow* directement sans passer par des concepts intermédiaires, ainsi que par l'exploitation de deux ensembles de données d'entrée : l'état de flow auto-rapporté des participants (n~9 500), et leurs données de connexion au MOOC (~80 Go), pendant deux ans. Aucune donnée de performance n'est collectée ni utilisée.

Après des filtrages et des prétraitements rigoureux, nous couplons les états de *flow* auto-rapportés des utilisateurs (n~1 500) à leurs données de connexion agrégées (23 variables) pour obtenir deux modèles entraînés d'apprentissage automatique.

A) Un modèle dit *Proof-of-Concept* qui corrobore les choix et le bon couplage des méthodes, des instruments de mesure du *flow* et des techniques d'apprentissage automatique (F1 = 0,851, AUC ROC = 0,85, Exactitude = 0,797, Précision = 0,821, Rappel = 0,882).

B) Un modèle dit Prototype qui permet de détecter *flow* dans un contexte de MOOC de manière abordable (F1 = 0,689, AUC PR = 0,87, AUC ROC = 0,68, Exactitude = 0,605, Précision = 0,854, Rappel = 0,578), rapide (quelques ms/participant), d'un impact environnemental négligeable (~0.00000237222222 g de CO_2eq par exécution), automatique (pas d'interventions supplémentaires une fois installé), et transparente (sans intervention des participants du MOOC).

Ces deux modèles identifient mieux le *flow* que l'absence de *flow* : *Proof-of-Concept* : 58% vs. 22% du total, Prototype : ~44% contre ~17% du total. Or, le Prototype présente une proportion plus élevée de Faux Négatifs (~32 % du total) que des Vrais Négatifs (~17% du total) lorsqu'il est confronté à des données non traitées et jamais vues (le prétraitement des données améliore les métriques mais reconstruit les données d'entrée). Cet écart peut être compris comme un modèle prudent préférant une classification pas-de-*flow* en cas de doute plutôt qu'une classification *flow*, ce qui n'est pas nécessairement un comportement indésirable du modèle.

D'ailleurs, nos deux modèles souffrent d'un manque de granularité pour la détection du *flow*, un obstacle insurmontable intrinsèquement lié à la granularité (2 moments) des données d'entraînement.

Les perspectives de ce projet de recherche comprennent la mise en œuvre et la commercialisation du modèle Prototype dans un MOOC pour 1) aboutir sur un tableau de bord du *flow* ; 2) personnaliser le contenu, les activités et le parcours d'apprentissage du MOOC ; et éventuellement 3) évaluer l'incidence de la détection du *flow* dans la personnalisation du MOOC afin de réduire le taux d'abandon du MOOC.

Summary

Automatic Flow (Optimal Learning Experience) Detection in a MOOC via Machine Learning

Background

Flow "is a gratifying state of deep involvement and absorption that individuals report when facing a challenging activity and they perceive adequate abilities to cope with it" (EFRN, 2014). The phenomenon was described by Csíkszentmihályi (1975b) in order to explain why people perform activities for no reason but for the activity itself, without extrinsic rewards. During flow, people are deeply motivated to persist in their activities and to perform such activities again (Csíkszentmihályi, 1975b; EFRN, 2014; Peifer et al., 2022).

Research shows that engagement, intention, and motivation (Chuang & Ho, 2016; Goopio & Cheung, 2021; Jung & Lee, 2018; Turner & Patrick, 2008; Wang & Baker, 2018; Watted & Barak, 2018) are among the top factors to affect learners' performance in MOOCs. Furthermore, studies (Abyaa et al., 2019; Efklides & Volet, 2005; Medina-Medina & García-Cabrera, 2016) agree that the learner's psychological state carries a preponderant weight in the learning process. More specifically, the human psychological state of flow (EFRN, 2014) has shown to correlate reliably and positively to learning-favorable metrics (Csíkszentmihályi et al., 2005; El Mawas & Heutte, 2019; Heutte, 2019; Motlagh et al., 2011; Peifer et al., 2022). More specifically, Csíkszentmihályi and his mentees argued (2005) that the flow state promotes learning and personal development because deep and total concentration experiences are intrinsically rewarding, and they motivate students to repeat any given activity at progressively higher challenging levels.

However, research associating flow in MOOCs is still on the growth (El Mawas & Heutte, 2019). At the same time, researchers concur that live flow detection remains particularly difficult, as any artifact attempting to detect it, let alone measure it, inevitably contributes somewhat to disrupt it (Csíkszentmihályi, 2014; Rheinberg & Engeser, 2018).

Research Question

As such, the incentives behind this thesis are a) the significance of the flow state (as a human psychological state) in the learning process, more specifically in online, distant settings, and b) the need of an automatic and transparent flow detection mechanism for learners in a MOOC context. Both incentives are conducive to asking ourselves:

Can the human psychological state of flow be automatic and transparently detected via the digital traces left by MOOC participants?

Objective

This research work proposes a flow-detecting Machine Learning model that allows for accountable, automatic, and transparent flow detection in a MOOC context, *i.e.*, online, distant learning settings.

This research work differentiates itself both from existing attempts (G.-S. Chen & Lee, 2012; Di Mitri et al., 2017; Heutte, Kaplan, et al., 2014; Hussain et al., 2012; Moneta & Csíkszentmihályi, 1996; Pfister, 2002), and from concurrent endeavors (Ghaleb et al., 2018; Sahid et al., 2020; Sajno et al., 2022), in a) the use of a theoretical flow model when designing flow indicators, one proper to online learning contexts; b) the application of flow measurement and characterization instruments proper to online, distant, learning contexts; c) the application of the *logit* function additionally of the linear function; d) the "straight-to-flow" specific detection mechanism, as opposed to positive-emotion-to-flow, or engagement-to-flow, or motivation-to-flow, or deep-concentration-to-flow mechanisms, but most remarkably; e) the use of two input datasets (within that same context) composed of participants' answers (*n*~9 500) to pin-point flow measurement instruments, and of their MOOC log data (~80GB), spanning all in all a two-year long data collection period. After rigorous data filtering, cleaning, and pre-processing, the input dataset is constituted of real-user data (*n*~1 500), their log data aggregations (23 features), and their self-reported flow state.

Methods

First, we identify in their respective literature fields the most appropriate flow measurement instruments (FlowQ and EduFlow-2) (Csíkszentmihályi & Csíkszentmihályi, 1988; Heutte et al., 2021), Machine Learning techniques (Pedregosa et al., 2011; Ramírez Luelmo, El Mawas, Bachelet, et al., 2022; Ramírez Luelmo et al., 2021c; Raschka & Mirjalili, 2019), and trace analysis methods (Cisel, 2017; Iksal, 2012; Pierrot, 2018; Pierrot et al., 2017; Slouma et al., 2019) in a Learning Analytics context (Iksal, 2012; Peraya & Luengo, 2019; Pierrot, 2018; Siemens, 2011), while considering the research terrain (Delpeyroux & Bachelet, 2015; Verzat & Bachelet, 2020). We employ Machine Learning techniques to make sense of multidimensional data (Conati et al., 2018) without constantly requiring a human expert.

Then, after a two year long massive data collection period in a French MOOC (Delpeyroux & Bachelet, 2015; Verzat & Bachelet, 2020), we corroborate the choices and proper coupling of methods, flow measurement instruments, and Machine Learning techniques listed above in a Proof-of-Concept experiment (Ramírez Luelmo, El Mawas, Bachelet, et al., 2022) using only the MOOC's participants' answers to the two selected flow measurement instruments.

Finally, we culminate this thesis project in a Prototype experiment where we further extend the input features by pairing the MOOC's participants' answers to the two selected flow measurement instruments to their respective MOOC log data (edX Inc, 2023). This aggregated log data is composed of variables such as: total number of events, diversity of events, total number of logged sessions, mean, longest and shortest sessions' lengths, total participation time, number of events per category, number of logged sessions per weekday, among others. Performance data is neither collected nor employed.

Results

This research work proposes both a Proof-of-Concept Machine Learning flow-detecting trained model (Ramírez Luelmo, El Mawas, Bachelet, et al., 2022), and a Prototype Machine Learning flow-detecting trained model.

In one hand, the Proof-of-Concept model validates the choices of methods, flow measurement instruments, and Machine Learning techniques, and constitutes itself a flow-detecting Machine Learning trained model that allows for reliable, automatic, flow-detection in a MOOC context (online, distant learning settings) by solely applying the EduFlow-2 measurement instrument. This model is ready for integration (Ramírez Luelmo et al., 2020b) into an Open Learner Model (Bull & Kay, 2010; El Mawas et al., 2019; El Mawas, Ghergulescu, et al., 2018; Ramírez Luelmo et al., 2021a) for a MOOC.

The Proof-of-Concept model computes very acceptable metrics for self-reported data (F1 = 0.851, AUC ROC = 0.85, Accuracy = 0.797, Precision = 0.821, Recall = 0.882), without overfitting. It identifies flow better (57.86%) than no-flow (21.80%) likely due to data imbalance originating from human bias linked to the flow measurement instruments (*cf.* Discussion and Conclusions below).

In the other hand, the Prototype model allows for affordable (*cf.* metrics below), fast (less than a few milliseconds per participant), of negligeable environmental impact (~0.00000237222222 g of CO_2eq per run), automatic (once properly setup no further intervention is needed by MOOC maintainers), and transparent (no post nor prior intervention demanded to MOOC participants) flow detection in a MOOC context, *i.e.*, online, distant learning settings.

The Prototype model features affordable metrics (F1 = 0.689, AUC PR = 0.87, AUC ROC = 0.68, Accuracy = 0.605, Precision = 0.854, Recall = 0.578) without overfitting when facing human-generated, unprocessed, unseen data. The featured model is the result of several (~25) GridSearchCV tasks (Dangeti, 2017; Pedregosa et al., 2011), each assessing about 3 600 distinct Machine Learning trained models, looking for the best fit to the input data according to the selected metrics.

Additionally, ANOVA-based FeatureSelection analysis (Guyon & Elisseeff, 2003) showed variable "number of sessions started on a Monday" consistently outweighed (*k* score = 45.15) all other variables when determining feature importance for flow detection, closely followed by "number of all logged sessions" (*k* score = 39.76), and "number of events" (*k* score = 32.87).

This resulting model can be easily implemented into existing MOOC's dashboards (a "Flow detection" section) by feeding it the considered features via API calls to detect flow at any given point of the MOOC.

Discussion and Conclusions

The Proof-of-Concept Machine Learning flow-detecting trained model identifies flow better than it does no-flow (58% *vs.* 22% of total). A phenomenon that also occurs in the resulting Prototype Machine Learning flow-detecting trained model (~44% *vs.* ~17% of total). However, the Prototype model features a higher proportion of False Negatives (~32% of total) classifications *vs.* True Negatives (~17% of total) when facing human-generated, unprocessed, unseen data (pre-processing data improves metrics but reconstructs input data). This discrepancy for no-flow detection can be interpreted as a cautious model preferring a no-flow classification when in doubt rather than a flow classification, which is not necessarily an unwelcome model behavior.

We hypothesize both imbalanced results are mostly due to 1) the intricate writing style employed in one of the chosen psychometric measurement instruments which might have largely contributed to 2) the imbalance of our input sample, and 3) human bias when answering psychometric tests in distant, online, educational settings, *e.g.*, self-identification with the item's text, eventual caveats associated to the flow measurement instruments, proportion of "committed" learners during a MOOC's length, etc. Yet, if we reconsider the performance of both models under the light of this (likely) immutable data imbalance imputed to human bias, then the PR AUC instead nets a very good result of 0.87 while both the ROC AUC and the PR AUC always perform above the predefined thresholds.

Furthermore, neither flow-detecting model constitutes black box models, *i.e.*, weighs, coefficients, and operations are known, human-inspectable, and human-interpretable. Thus, they can facilitate eventual model inspection and understanding, which counterweights the underlying human bias present in the training data. However, both resulting models experience a lack of granularity when detecting flow, *i.e.*, they are unable to pinpoint a precise flow moment in time. This insuperable obstacle is intrinsically linked to the granularity of the flow training data, which is limited to two, accumulative flow detection moments.

Perspectives for this research project comprise the successful implementation and commercialization of the flow-detecting Machine Learning trained model into a MOOC to 1) implement a flow dashboard; 2) personalize the MOOC's content, activities, and learning-path; and possibly 3) evaluate the incidence of flow detection in MOOC personalization when attempting to reduce MOOC dropout rates.

Résumé Substantiel

Détection Automatique du Flow (Expérience Optimale d'Apprentissage) dans un MOOC via des Techniques d'Apprentissage Automatique (Machine Learning)

Contexte

Le *flow* « est un état d'épanouissement lié à une profonde implication et au sentiment d'absorption que les personnes ressentent lorsqu'elles sont confrontées à des tâches dont les exigences sont élevées et qu'elles perçoivent que leurs compétences leur permettent de relever ces défis » (EFRN, 2014). Le phénomène a été décrit par Csíkszentmihályi (1975b) afin d'expliquer pourquoi les gens effectuent des activités sans autre raison que l'activité elle-même, sans récompenses extrinsèques. Pendant *flow*, les personnes sont profondément motivées pour persévérer dans leurs activités et pour les répéter (Csíkszentmihályi, 1975b; EFRN, 2014; Peifer et al., 2022).

La recherche montre que l'engagement, l'intention et la motivation (Chuang & Ho, 2016; Goopio & Cheung, 2021; Jung & Lee, 2018; Turner & Patrick, 2008; Wang & Baker, 2018; Watted & Barak, 2018) sont parmi les principaux facteurs qui affectent la performance des apprenants dans les MOOC. En outre, des études (Abyaa et al., 2019; Efklides & Volet, 2005; Medina-Medina & García-Cabrera, 2016) s'accordent à dire que l'état psychologique de l'apprenant a un poids prépondérant dans le processus d'apprentissage. Plus précisément, l'état psychologique humain du *flow* (EFRN, 2014) a montré une corrélation fiable et positive avec des métriques favorisant l'apprentissage (Csíkszentmihályi et al., 2005; El Mawas & Heutte, 2019; Heutte, 2019; Motlagh et al., 2011; Peifer et al., 2022). Plus précisément , (Csíkszentmihályi et al., 2005) soutiennent que l'état de *flow* favorise l'apprentissage et le développement personnel parce que les expériences de concentration profonde et totale sont intrinsèquement gratifiantes et motivent les étudiants à répéter toute activité donnée à des niveaux de difficulté progressivement plus élevés.

Pourtant, la recherche associant *flow* dans les MOOC est encore en pleine croissance (El Mawas & Heutte, 2019). Au même temps, les chercheurs s'accordent à dire que la détection du *flow* en direct reste particulièrement difficile, car tout artefact tentant de le détecter, voire le mesurer, contribue inévitablement à le perturber (Csíkszentmihályi, 2014; Rheinberg & Engeser, 2018).

Question de Recherche

Ainsi, les motivations de cette thèse sont a) l'importance de l'état de *flow* (en tant qu'état psychologique humain) dans le processus d'apprentissage, plus spécifiquement dans les environnements en ligne et à distance, et b) le besoin d'un mécanisme de détection du *flow* automatique et transparent pour les apprenants dans le contexte d'un MOOC. Ces deux incitations nous amènent à nous poser la question :

L'état psychologique humain du *flow*, peut-il être détecté automatiquement et de manière transparente via les traces numériques laissées par les participants d'un MOOC ?

Objectif

Ce travail de recherche propose un modèle d'apprentissage automatique de détection de *flow* qui permet une détection fiable, automatique et transparente du *flow* dans un contexte de MOOC, c'est-à-dire un contexte d'apprentissage en ligne et à distance.

Ce travail de recherche se différencie des tentatives précédentes bien établies dans la littérature (G.-S. Chen & Lee, 2012; Di Mitri et al., 2017; Heutte, Kaplan, et al., 2014; Hussain et al., 2012; Moneta & Csíkszentmihályi, 1996; Pfister, 2002) mais aussi des efforts concurrents à ce projet de thèse (Ghaleb et al., 2018; Sahid et al., 2020; Sajno et al., 2022), dans a) l'utilisation d'un modèle théorique de *flow* lors de la conception d'indicateurs de *flow*, propre aux contextes d'apprentissage en ligne ; b) l'application d'instruments de mesure et de caractérisation du *flow* propres aux contextes d'apprentissage en ligne, à distance ; c) l'application de la fonction *logit* en plus de la fonction linéaire ; d) le mécanisme de détection spécifique et direct (« *straight-to-flow* »), par opposition aux mécanismes passant ou se servant des émotions positives (« *positive-emotion-to-flow* »), ou de l'engagement (« *engagement-to-flow* »), ou de la motivation (« *motivation-to-flow* »), mais surtout ; e) l'utilisation de deux ensembles de données d'entrée (du même contexte) composés des réponses des participants (*n~9* 500) aux instruments de mesure du *flow* et de leurs données de connexion au MOOC (~80 Go), couvrant en tout une période de collecte de données de deux ans. Après un filtrage, un nettoyage et un prétraitement rigoureux des données, l'ensemble de données d'entrée est constitué de données d'utilisateurs réels (*n~1* 500), de l'agrégation de leurs données de journal système (23 variables), et de leur état de *flow* auto-rapporté.

Méthodes

Tout d'abord, nous identifions dans leur littérature respective les instruments de mesure du *flow* les plus appropriés (FlowQ et EduFlow-2) (Csíkszentmihályi & Csíkszentmihályi, 1988; Heutte et al., 2021), les techniques d'apprentissage automatique (Pedregosa et al., 2011; Ramírez Luelmo, El Mawas, Bachelet, et al., 2022; Ramírez Luelmo et al., 2021c; Raschka & Mirjalili, 2019), et les méthodes d'analyse des traces (Cisel, 2017; Iksal, 2012; Pierrot, 2018; Pierrot et al., 2017; Slouma et al., 2019), dans un contexte d'analyse de l'apprentissage (Iksal, 2012; Peraya & Luengo, 2019; Pierrot, 2018; Siemens, 2011), sans perdre de vue le terrain de recherche (Delpeyroux & Bachelet, 2015; Verzat & Bachelet, 2020). Nous utilisons des techniques d'apprentissage automatique pour donner un sens aux données multidimensionnelles (Conati et al., 2018) sans avoir à faire appel à un expert humain en permanence.

Ensuite, après une période de collecte massive de données de deux ans dans un MOOC français (Delpeyroux & Bachelet, 2015; Verzat & Bachelet, 2020), nous corroborons les choix et le bon couplage des méthodes, des instruments de mesure de *flow*, et des techniques d'apprentissage automatique énumérées ci-dessus dans une expérience de démonstration de faisabilité (« *Proof-of-Concept* ») (Ramírez Luelmo, El Mawas, Bachelet, et al., 2022) en nous servant uniquement des réponses des participants au MOOC aux deux instruments de mesure de *flow* sélectionnés.

Enfin, nous terminons ce projet de thèse par une expérience de prototypage dans laquelle nous étendons encore plus les variables d'entrée du modèle en associant les réponses des participants au MOOC aux deux instruments de mesure de *flow* sélectionnés à leurs données respectives du journal système du MOOC (edX Inc, 2023). Ces données agrégées sont composées de variables telles que : le nombre total d'événements, la diversité des événements, le nombre total de sessions enregistrées, leur durée moyenne, la plus longue et la plus courte des sessions, le temps total de participation, le nombre d'événements par catégorie, le nombre de sessions enregistrées par jour de la semaine, entre autres. Aucune donnée de performance n'est collectée ni utilisée.

Résultats

Ce travail de recherche propose à la fois un modèle entraîné de détection de *flow* par apprentissage automatique démontrant avec succès sa faisabilité (« *Proof-of-Concept Machine Learning model* ») (Ramírez Luelmo, El Mawas, Bachelet, et al., 2022), et un modèle prototype entraîné de détection de *flow* par apprentissage automatique.

D'une part, le premier modèle (« *Proof-of-Concept* ») valide les choix de méthodes, d'instruments de mesure de *flow* et de techniques d'apprentissage automatique, et constitue lui-même un modèle d'apprentissage automatique de détection de *flow* qui permet une détection fiable et automatique du *flow* dans un contexte MOOC (contexte d'apprentissage en ligne et à distance) en appliquant uniquement l'instrument de mesure EduFlow-2. Ce modèle est prêt à être intégré (Ramírez Luelmo et al., 2020b) dans un modèle d'apprenant ouvert (Bull & Kay, 2010; El

Mawas et al., 2019; El Mawas, Ghergulescu, et al., 2018; Ramírez Luelmo et al., 2021a) pour un MOOC.

Ce premier modèle présente des métriques très acceptables pour les données auto rapportées (F1 = 0,851, AUC ROC = 0,85, Exactitude = 0,797, Précision = 0,821, Rappel = 0,882), sans surajustement. Il identifie mieux la présence de *flow* (57,86%) que l'absence de *flow* (21,80%), probablement en raison du déséquilibre des données d'entrée provenant lui-même d'un biais humain lié aux instruments de mesure du *flow* (*cf.* Discussion et Conclusions, dans la suite).

D'autre part, le modèle prototype permet une détection du *flow* abordable (*cf.* métriques ci-dessous), rapide (moins de quelques millisecondes par participant), d'un impact environnemental négligeable (~0,000002372222 g de CO_2eq par exécution), automatique (une fois correctement configuré, aucune intervention supplémentaire n'est nécessaire de la part des responsables du MOOC), et transparente (aucune intervention postérieure ou préalable n'est demandée aux participants du MOOC) dans un contexte MOOC, c'est-à-dire dans le cadre d'un apprentissage en ligne et à distance.

Le modèle prototype présente des métriques abordables (F1 = 0,689, AUC PR = 0,87, AUC ROC = 0,68, Exactitude = 0,605, Précision = 0,854, Rappel = 0,578) sans surajustement lorsqu'il est confronté à des données non traitées, non vues, et générées par des êtres humains. Plus spécifiquement, le modèle présenté est le résultat de plusieurs (~25) tâches GridSearchCV (Dangeti, 2017; Pedregosa et al., 2011), chacune évaluant environ 3 600 modèles entraînés par apprentissage automatique distincts, cherchant la meilleure adéquation aux données d'entrée, selon les métriques sélectionnées.

De plus, l'analyse de sélection des variables (*FeatureSelection*) basée sur l'ANOVA (Guyon & Elisseeff, 2003) a montré que la variable « nombre de sessions commencées un lundi » emportait systématiquement (score k = 45,15) sur toutes les autres variables lors de l'analyse de l'importance des variables pour la détection du *flow*, suivie de près par le « nombre de toutes les sessions enregistrées » (score k = 39,76) et le « nombre d'événements » (score k = 32,87).

Ce modèle peut être facilement implémenté dans les tableaux de bord des MOOC existants (section « Détection du *flow* ») en lui fournissant les variables considérées via des appels API pour détecter *flow* à n'importe quel point du MOOC.

Discussion et Conclusions

Notre premier modèle entraîné de détection de *flow* via l'apprentissage automatique (« *Proof-of-Concept* ») identifie mieux la présence de *flow* que l'absence de *flow* (58% *vs.* 22% du total). Ce phénomène se retrouve également dans le modèle entraîné de détection de *flow* via l'apprentissage automatique du prototype (~44% *vs.* ~17% du total). Cependant, le modèle prototype présente une proportion plus élevée de faux négatifs (~32% du total) par rapport aux vraies négatifs (~17% du total) lorsqu'il est confronté à des données réelles (générées par des humains), non traitées, et non vues au préalable (le prétraitement de ces données améliore les métriques, mais modifie les données d'entrée). Cet écart entre la classification de l'absence et la présence de *flow* peut être expliqué par un modèle « trop prudent » préférant une classification d'absence de *flow* en cas de doute plutôt qu'une classification de présence de *flow*, ce qui n'est pas nécessairement un comportement indésirable.

Nous émettons l'hypothèse que ces deux résultats déséquilibrés sont principalement dus 1) au style d'écriture complexe employé dans l'un des instruments de mesure psychométrique choisis, qui pourrait avoir largement contribué au 2) déséquilibre de notre échantillon d'entrée, et enfin 3) au biais humain lorsqu'il s'agit de répondre à des tests psychométriques dans des contextes éducatifs en ligne à distance, par exemple, l'autoidentification du répondant avec le texte de l'item, les éventuelles mises en garde connues et associées aux instruments de mesure du *flow*, la proportion d'apprenants « engagés » pendant la durée du MOOC, etc. Pourtant, si l'on reconsidère la performance des deux modèles sous la lumière d'un déséquilibre (probablement) immuable des données (imputé au biais humain), alors le PR AUC fournit un très bon résultat de 0,87, au même temps que le ROC AUC et le PR AUC demeurent toujours supérieurs aux seuils prédéfinis.

En plus, aucun des deux modèles de détection de *flow* ne constitue un modèle « boîte noire », c'est-à-dire que les poids, les coefficients et les opérations réalisées sont tous connus, vérifiables, et interprétables. Ils facilitent donc l'inspection et la compréhension du modèle, ce qui contrebalance le biais humain sous-jacent présent dans les données d'apprentissage. Cependant, les deux modèles résultants manquent de granularité lors de la détection du *flow*, c'est-à-dire qu'ils sont incapables d'identifier un moment précis du *flow* dans le temps. Cet obstacle insurmontable est intrinsèquement lié à la granularité des données d'apprentissage du flow, elles-mêmes limitées à deux moments de détection de *flow* cumulés.

Les perspectives de ce projet de recherche comprennent la mise en œuvre réussie et la commercialisation du modèle d'apprentissage automatique de détection de *flow* dans un MOOC pour 1) mettre en œuvre un tableau de bord de *flow* ; 2) personnaliser le contenu, les activités et le parcours d'apprentissage du MOOC ; et éventuellement 3) évaluer l'incidence de la détection du *flow* dans la personnalisation du MOOC dans le but de réduire le taux d'abandon des MOOC.

Index

Appendices index

Introduction

Background

Since the last decade, the world has seen a mass proliferation of Massive Online Open Courses (MOOC) (Siemens, 2015, p. xiii; Yousef, Chatti, Schroeder, Wosnitza, et al., 2014). The importance of MOOCs and online Learning Environments' (LE) was recognized as an indispensable tool to bring the classroom to the learners (Arora & Srinivasan, 2020; T. Chen et al., 2020; Doghonadze et al., 2020; Feng et al., 2020; Nabukeera, 2020; Pevneva & Edmunds, 2020). This importance was only accentuated during the worldwide pandemic of COVID-19[1], when educational institutions experienced a sudden and quick adoption rate to their increased MOOC platform attendance, affecting their offer and access, but mostly their students, all over the world (Amruta & Naik Ramgir, 2021; Kichu & Bhattacharya, 2021; Shah, 2020b, 2020c; Xiong et al., 2021).

Still, MOOCs' world success mostly goes back to their original concept: offering free and open access courses for a massive number of learners from anywhere all over the world (McAuley et al., 2010; Yousef, Chatti, Schroeder, Wosnitza, et al., 2014). But despite this global reach, their immense popularity and their -very often- low-to-none costs, MOOC learners feature very low completion rates, with most research metrics agreeing on an overall median of about 6.5% MOOC completion rate (Chuang & Ho, 2016, p. 6; Clerc et al., 2015; Jordan, 2014; MAUT, 2015; Yuan & Powell, 2013). Even when looking at fee & certification-based MOOCs, completion rates top around 60%, *i.e.*, a tenfold difference mostly explained by this economic incentive. Furthermore, research shows that engagement, intention and motivation are among the

[1] December 2019 – July 2020 in France but as of January 2023 still ongoing in some regions, *e.g.*, China.

top factors to affect performance in MOOCs (Chuang & Ho, 2016, pp. 6–7; Goopio & Cheung, 2021; Jung & Lee, 2018; Turner & Patrick, 2008; Wang & Baker, 2018; Watted & Barak, 2018).

Furthermore, researchers agree that the learner's psychological state carries a preponderant weight in the learning process (Abyaa et al., 2019; Efklides & Volet, 2005; Medina-Medina & García-Cabrera, 2016). More specifically, related studies pioneered by Csíkszentmihályi et al. (2005) and confirmed down the road by Peifer et al. (2022) have shown the human psychological state of flow to reliably and positively correlate to learning-favorable metrics, such as engagement, motivation, self-efficacy, self-determination, self-regulation, curiosity, goal-attainment, and academic achievement efficacy (Csíkszentmihályi et al., 2005; El Mawas & Heutte, 2019; Heutte, 2019; Motlagh et al., 2011; Peifer et al., 2022).

Flow is "a gratifying state of deep involvement and absorption that individuals report when facing a challenging activity and they perceive adequate abilities to cope with it" (EFRN, 2014). During flow state, people are deeply motivated to persist in their activities (as diverse as they might be) and to perform them again for their own satisfaction, without extrinsic rewards (EFRN, 2014; Rufi et al., 2014). All of these factors positively affect learning and make of flow a desired psychological state when promoting learning, specifically in online, distant settings (Skadberg & Kimmel, 2004).

However, research associating flow in MOOCs is still on the growth (El Mawas & Heutte, 2019) while automatic, real-time flow detection remains particularly difficult, as any artifact attempting to detect it, let alone measure it, inevitably contributes somewhat to disrupt flow (Csíkszentmihályi, 2014, p. 258; Rheinberg & Engeser, 2018, p. 602)[2].

Research Question

As such, the incentives behind this thesis are a) the significance of the flow state (as a human psychological state) in the learning process, more specifically in online, distant settings,

[2] Page numbering facilitates interested readers to locate and thus, corroborate the releveant passage.

and b) the need of an automatic and transparent flow detection mechanism for learners in a MOOC context. Both incentives are conducive to asking ourselves:

Can the human psychological state of flow be automatic and transparently detected via the digital traces left by MOOC participants?

Objective

The general objective of this research project is to detect flow in a MOOC in an accountable, automatic, and transparent fashion.

We intend to address this issue by using Machine Learning techniques to pair traditional flow measurement tools to the digital traces left by MOOC participants. We postulate that Machine Learning techniques will help unearth the subjacent relationship between the learners' flow state and their corresponding digital traces, up to a reasonable degree.

Thus, this research work proposes a flow-detecting Machine Learning model that allows for accountable, automatic, and transparent flow detection in a MOOC context, *i.e.*, online, distant learning settings.

While the proposed Machine Learning model materializes itself in the field of Computer Science, it is firmly established in Education and Training Sciences. This materialization exploits the notion of ideal-type (Weber, 1922), a notion akin yet distinct from that of stereotypes (Leyens et al., 1996). As such, the resulting model exists:

- First, as a simplified representation (Galindo, 2017) of the reality of flow (Csíkszentmihályi, 1975b; EFRN, 2014).
- And second, as a hypothesis that the factors characterizing flow might exist or be valid only within our specific research context (Bertereau et al., 2019).

Therefore, we adopt a positivist position (Avenier & Thomas, 2015) inasmuch as us and our resulting Machine Learning model presume an iconic and measurable representation of the phenomenon of flow, plus its accompanying bias which the model seeks to reduce as much as we can be aware of it.

Under the positivist research methodology paradigm by Churchill (1979), the resulting flow-detecting Machine Learning model issued from this research project constitutes a sort of measuring scale built upon statistical tests (Bertereau et al., 2019, p. 60) for the ideal-type construct of flow.

Because of the various notions mobilized, such as psychometry, Machine Learning and MOOCs, this research sits firmly at the frontier of three disciplines: Psychology, Computer Science, and Education and Training Sciences. The scope of this study is illustrated in Figure 0-4 in Appendix 11.

Furthermore, this research project follows a mixed, predominantly quantitative -or as we shall see later rather massive quantitative- approach.

Statement of Contribution

We believe this milestone to be of ultimate interest to our target public (MOOC designers/providers, pedagogical engineers and researchers who meet difficulties to incorporate psychological states in MOOCs) to take better informed decisions, in terms of collaborative work, learners' follow-up, and/or content's difficulty adaptation. This research work differentiates itself both from existing works (G.-S. Chen & Lee, 2012; Di Mitri et al., 2017; Heutte, Kaplan, et al., 2014, p. 20; Hussain et al., 2012; Moneta & Csíkszentmihályi, 1996, p. 292; Pfister, 2002) and from concurrent efforts (Ghaleb et al., 2018; Sahid et al., 2020; Sajno et al., 2022) in:

a. the consideration of a theoretical flow model when designing flow indicators,
b. the application of flow measurement and characterization instruments proper to online, distant, learning contexts,
c. the application of the *logit* function additionally of the linear function,

d. the "straight-to-flow" employed detection mechanism, as opposed to positive-emotion-to-flow, or engagement-to-flow, or deep-concentration-to-flow mechanisms, but most remarkably,
e. the use of two input datasets (within that same context) composed of participants' answers (*n*~9 500) to pin-point flow measurement tools, and of their MOOC log data (~80GB), spanning all in all a two-year long data collection period. After rigorous data filtering, cleaning, and pre-processing, the input dataset is constituted of real-user data (*n*~1 500), their log data aggregations (23 features), and their self-reported flow state.

Methods

Nowadays, Machine Learning is considered instrumental in addressing the issue of learning from data the knowledge that might be difficult to obtain from human experts, such as computing predictions of learners' cognitive and mental states (Conati et al., 2018, p. 22).

After a two year long massive data collection period, we pair the answers to two selected flow measurement tools (FlowQ and EduFlow-2) (Csíkszentmihályi & Csíkszentmihályi, 1988; Heutte et al., 2021) by participants in a French MOOC to aggregations of their respective MOOC log data using Machine Learning techniques. Such aggregations include features such as total number of events, diversity of events, total number of logged sessions, mean, longest and shortest sessions' lengths, total participation time, number of events per category, number of logged sessions per weekday, among others.

No performance data is collected nor employed for the design of these experiments, and thus neither is performance considered in the resulting Machine Learning flow-detecting trained model.

Results

This research work proposes one Machine Learning flow-detecting trained model that allows for affordable (*cf.* metrics below), fast (less than a few milliseconds per participant), of negligeable environmental impact (~0.00000237222222 g of CO_2eq per run), automatic (once properly setup no further intervention is needed by MOOC maintainers), and transparent (no post nor prior intervention demanded to MOOC participants) flow detection in a MOOC context, *i.e.*, online, distant learning settings.

This model features affordable metrics (F1 = 0.689, AUC PR = 0.87, AUC ROC = 0.68, Accuracy = 0.605, Precision = 0.854, Recall = 0.578) without overfitting when facing human-generated, unprocessed, unseen data (metrics improve when facing pre-processed unseen data). The featured model is the result of several (~25) GridSearchCV tasks, each assessing about 3 600 distinct Machine Learning trained models, looking for the best fit to the input data according to the selected metrics.

Additionally, ANOVA-based FeatureSelection analysis showed the variable "number of sessions started on a Monday" consistently outweighed (*k* score = 45.15) all other variables when determining feature importance for flow detection, closely followed by the variable named "number of all logged sessions" (*k* score = 39.76), and "number of events" (*k* score = 32.87).

The resulting model can be easily implemented into existing MOOC's dashboards (a "Flow detection" section) to detect flow at any given point of the MOOC; the calculations are almost instantaneous and do not require further Machine Learning training.

Discussion and Conclusions

The resulting Machine Learning flow-detecting trained model detects flow better than it does no-flow. However, the model features a higher proportion of False Positive (~32% of total) detections *vs.* True Negatives (~17% of total) when facing human-generated, unprocessed, unseen data (the closest to real-life conditions) which account for the human bias (*cf.* remarks 1 & 2 below).

We hypothesize these results are mostly due to 1) the intricate writing style employed in one of the chosen psychometric measurement tools which might have largely contributed to

2) the imbalance of our input sample, and 3) human bias when answering psychometric tests in distant, online, educational settings.

Furthermore, although the model can be applied at any given moment during the MOOC duration, we hypothesize results might be more accurate by the end of the MOOC, when the participant's actions in the MOOC provide a fuller rapport for the model to calculate. Indeed, just like during normal MOOC dropout, participants more committed to the MOOC completion answer questionnaires more accurately, hence a larger proportion of "committed" respondents, more likely to have experienced flow themselves, end up in a major representation in the final sample, when compared to those who a) did not experience flow and dropped the MOOC before reporting their lack of flow; or b) did not experience flow and thus, chose not to report their lack of flow.

Like in any other Machine Learning project, larger, quality input datasets might provide additional information on the phenomenon being detected. A different characterization of the flow state by researchers designing the flow measurement questionnaires, and/or by researchers translating these into Machine Learning features and transformations for the model in training (*i.e.*, the informed decisions behind) will deeply impact the metrics of the final model.

Finally, we point to the current phase for this research, which is the successful implementation of our Machine Learning flow-detecting trained model into a MOOC which in turn would perform the necessary steps to 1) implement a flow dashboard for the MOOC participants; 2) personalize the MOOC's content, activities, and learning-path according to the participant's detected flow state; and possibly 3) evaluate the incidence of flow detection in MOOC personalization when reducing MOOC dropout rates.

Outline

This dissertation is composed of six Chapters, grouped in three main parts: Part I – State of the art, Part II – Research method, and Part III – Our proposed approach

This Introduction served to present the current context surrounding this research, the problem statement and thus, the research question. We also described how this study contributes to the state of the art.

Part I covers the literature review framing this research work. It comprises Chapter 1 to 3 extensively detailing on the flow psychological state (Chapter 1), the phenomenon of Massive Open Online Courses (Chapter 3), and the basis of Machine Learning (Chapter 4). Fragments of Chapter 1 have been published by this thesis' author and his mentors (Ramírez Luelmo, El Mawas, Bachelet, et al., 2022; Ramírez Luelmo, El Mawas, & Heutte, 2022; Ramírez Luelmo et al., 2020a, 2021a).

Part II presents the research method employed in this thesis, organized as depicted in Figure 5-1. It announces the steps proposed to approach the research question and quickly overviews the prevalent methods of trace analysis in MOOC.

Part III concerns the experimentation that took place to approach the research question. It comprehensively describes the research-based decisions taken for each of the steps evoked during the previous Part II, *viz.* the selection of the flow measurement instruments, the experimentation terrain, and the Machine Learning method. It also details the very important steps of data collection and data processing before the actual experimentation, both extensively exploited here.

This manuscript ends with an unnumbered Conclusion & perspectives Chapter.

Part I

State of the art

*State of the Art

The first Part of this thesis describes the scientific framework surrounding this research study.

We begin by presenting the flow psychological state, its characteristics (*a.k.a.* components, facets, dimensions), its measurement instruments, along with the future lines of study that lay ahead. We conclude this flow review by explaining its importance in learning and thus, in this research work.

We proceed with a brief account of MOOCs, how they have become a viable alternative for training, especially in higher education and in the current world context. A few relevant categorizations are also presented.

As we continue connecting the dots, we review the basis of Machine Learning. Notions and terminology widely employed in Part III are covered here. A noteworthy Subsection on Machine Learning best practices by researchers in the field is also included here.

With this information under the arm, we proceed to Chapter 1 Flow.

Chapter 1. Flow

This thesis aims to detect flow in a transparent and automatic fashion in an online, distant, educational context. Without deep knowledge of what is flow, what are its components, and most importantly how to measure it, this task would not be possible.

Thus, this first Chapter presents the flow human psychological state, its scientific antecedents, its components and their evolution, how researchers have approached its measurement, and how it has been employed in educational contexts, among many other fields. Specifically linked to our research context, we review some flow measurement instruments historically employed in educational settings as we ponder employing them in this thesis. Furthermore, we present a compilation of suggestions to retain when measuring flow, by experienced flow researchers widely recognized by the flow community.

Mihály Róbert Csíkszentmihályi first described flow (1975a, 2014) in order to explain why people perform activities for the activity itself, without extrinsic rewards. Moreover, flow advent is not limited to a single subject, it may appear in any area of life (Csíkszentmihályi & Csíkszentmihályi, 1988), and the experience is the same across lines of culture, class, gender, and age, as well as across kinds of activity (Csíkszentmihályi, 2014; Moneta, 2004).

Flow is the state that people often recognize as "being in the zone", "getting into the zone", "being in the flow" (Biclar et al., 2019; Bodiam, 2017; Csíkszentmihályi, 1997; Laird et al., 2021; Milton, 2019; Shehata et al., 2021; Stamatelopoulou et al., 2018) across all spheres of life[3], or as "playing for the love of the game" in competitive contexts (Stamatelopoulou et al., 2018), or "in the groove" (Jackson & Marsh, 1996), but also characterized by expressions such as "in ecstasy" (Csíkszentmihályi, 1997), or even "a zen feeling" (Csíkszentmihályi, 1990a, p. 62).

[3] Not to be confused nor associated with "in trance".

According to Csíkszentmihályi himself, flow state is a state of optimal experience that occurs when "people are so involved in an activity that nothing else seems to matter; the experience itself is so enjoyable that people will do it even at great cost, for the sheer sake of doing it" (1990a, p. 4).

When in flow, the individual operates at full capacity (Deci, 1975; White, 1959) while in a state of dynamic equilibrium. Entering flow depends on establishing a balance between perceived action capacities and perceived action opportunities and such a balance is intrinsically fragile (Csíkszentmihályi, 2014, p. 241); *e.g.*, if challenges begin to exceed skills, an individual first becomes vigilant, then anxious; if the skill begin to exceed challenges, the individual first relaxes and then becomes bored. Experiencing anxiety or boredom presses a person to adjust his/her level of skill/challenge in order to escape the aversive state and reenter flow (Csíkszentmihályi, 2014).

Csíkszentmihályi considered flow "as a continuum, ranging from repetitive, almost automatic acts […] to complex activities which require the full use of a person's physical and intellectual potential" (1975a, p. 54) and introduced the notions of "microflow" activities (simple unstructured activities performed during the day[4]) in opposition to "macroflow" activities (complex structured activities that produce full-fledged flow experiences), as well as "shallow-flow" activities where aspects of the activity facilitate stopping the activity (making it usually short[5]) in opposition to "deep-flow" activities where there are "fewer deterrents to flow" (Csíkszentmihályi, 1975a, p. 108).

It is crucial to keep in mind that flow is neither intrinsically good nor bad but it is the surrounding social context in which it develops that qualifies the consequences attributed to individuals being in flow while performing activities: "It is an illusion to believe that any solution

[4] *e.g.*, doodling, humming, chewing gum, hair smoothing, finger tapping, smoking, etc. (Csíkszentmihályi, 1975a, p. 108).
[5] Author uses dancing as an example: it can be started and stopped at any moment at will (without "grave" consequences), and songs are "usually short".

is beneficial for all people and all times; no human achievement can be taken as the final word" (Csíkszentmihályi, 1990a, p. 70).

The following Sections "Antecedents of Flow", "Evolution of the Definition of Flow and its Conceptual Flow Model", and "Evolution of Components of Flow" are closely linked to each other and they deeply delve into the flow notion; how it has evolved since its inception in both its definition and its components. Therefore, they concern mostly flow enthusiasts, and their contents might be overwhelming to the « *non-initiés* ». It is safe to assume that the current introduction suffices to understand the flow concept in its general terms. Therefore, to ease the reading of this manuscript, the reader can skip to Section "Measurement Attempts of Flow" below and later, if needed, return here for a deeper flow read.

Antecedents of Flow

This Section traces the initial research of flow and the context in which it developed: the paradigm shifts occurring at the time from a product-based to the study of creativity, from an extrinsic motivation approach to an intrinsic motivation approach. We overview the path on creativity, enjoyment, and intrinsic motivation taken by Csíkszentmihályi that led him to flow[6].

Piniel & Albert (2020) link the birth of the flow concept to "research on creativity starting in the 1950's in the United States of America" where the focus of psychology research started to shift towards the creative process itself, in opposition to the product orientation approach then prevailing (Csíkszentmihályi & Csíkszentmihályi, 1988). This shift had been kicked off by Maslow (1965) whose conceptual framework tackled the phenomena of "peak experience" as a vector to "self-actualization" (Bernard, 2009, p. 8; Turan, 2019) whereas for Csíkszentmihályi the then known as "flow experiences" (Csíkszentmihályi & Csíkszentmihályi, 1988) were more closely linked to individual happiness and individual transcendence. Nevertheless, we invite the curious

[6] Pun intended.

reader, interested in a deeper understanding of the theoretical precursors of flow, to review the works of Engeser & Schiepe-Tiska (Engeser & Schiepe-Tiska, 2012, pp. 9–13).

As a prelude, Mihály Róbert Csíkszentmihályi (1988) explains how, during the course of his doctoral research, he was led to investigate the immersive artistic experiences of a group of painting and sculpting artists, who after spending unaccountable, immensely enjoyable hours working on their oeuvres, lost all interest as soon as it was finished: "Why, then, did they work so hard at the easel - as hard as any executive hoping for a raise or a promotion? None of the extrinsic rewards that usually motivate behavior seemed to be present" (1988, p. 4).

At the time (mid-60's), the Csíkszentmihályi couple recounted the little prevailing interest in intrinsic motivation: "few psychologists were as yet interested in intrinsic motivation; the ruling paradigm was still exclusively focused on explaining behavior in terms of extrinsic rewards" (1988, p. 4). Fortunately, things started to change and research focus broadened to consider intrinsic motivation: "The scope of investigation later broadened to include all kinds of intrinsically motivated […] activities" (Piniel & Albert, 2020, p. 580).

This booting shift was better noticed in the now classic paper of Robert White (1959), where a "different kind of motivational concept, one that would complement drives and could be the basis of a motivational theory with greater explanatory power" (Deci & Ryan, 1985, p. 5) was argued for. Such "motivational concept" would originally be named "effectance motivation" on the basis that "[…] organisms are innately motivated to be effective in dealing with their environment" (1985, p. 5). Indeed, authors Deci & Ryan (1985, p. 5) explained that, around that time, in psychoanalytic theory, this motivational force was "generally referred to as independent ego energy" (1985, p. 5), although psychologists ascribed to an "empirical tradition" (1985, p. 5) would refer to the non-drive-based motivation as "intrinsic motivation", implying that such energy is "intrinsic to the nature of the organism" and was based on the "organismic needs to be competent and self-determining" (1985, p. 5), which is how we understand the term.

Flow thus has its roots in the study of enjoyment and intrinsic motivation (Deci, 1975; Deci & Ryan, 1980, 1985) by Mihály Róbert Csíkszentmihályi (1975a). It originated from his intention to study (within his then-contemporary society context) the assumption that serious work is "grim and unpleasant"[7] and thus, to redress "this harmful situation" (1975a, p. 1).

So, Csíkszentmihályi argued that "by objectifying incentives into money and status, societies have developed a rational, universal motivational system whereby communities can produce desired behaviors predictably and can allot precisely differentiated rewards to construct a complex social hierarchy" (1975a, p. 2). Thus, this led society to assume that "extrinsic rewards like money and status are basic human needs". Nevertheless, evidence showed him that individuals chose "to expend energy for goals that carry no conventional material rewards" (1975a, p. 3)., and he hoped to learn from these individuals the "inner working and relationships of intrinsic motivation.

Csíkszentmihályi did put in evidence the (somewhat still) then-prevailing cultural differentiation on work and leisure and how inexorably there is a link of the latter with enjoyment although not so with the former:

> […] in our culture […] what one **must** do cannot be enjoyable. So we have learned to make a distinction between 'work' and 'leisure': the former is what we have to do most of the time against our desire; the latter is what we like to do, although it is useless. (Csíkszentmihályi, 1975a, p. 3)

He punctuated that we "therefore feel bored and frustrated on our jobs, and guilty when we are at leisure", further distinguishing that assumed roles with extrinsic rewards do not necessarily lead to more [life] enjoyment: "the more a person complies with extrinsically rewarded roles, the less he enjoys himself, and the more extrinsic rewards he needs" (Csíkszentmihályi, 1975a, p. 4). On this matter, he resolved that the only escape at hand from this vicious circle

[7] Paradoxically, in the later findings of Csíkszentmihályi & LeFevre (1989), flow experiences were reported "when working, not when in leisure" although motivation was higher "in leisure than in work".

would be by making the roles themselves more enjoyable, hinting thus to the need of an intrinsic motivation for the role itself, and finally concludes that "when a social system learns to rely exclusively on extrinsic rewards, it creates alienation among its members […]" (1975a, p. 4). Yet, avoiding being viewed from a reductionist approach, Csíkszentmihályi takes the time to explain that any experience, by being enjoyable, is not simply pleasurable. Instead, a holistic approach considering an individual's goals, abilities, its subjective evaluation of the experience at hand, and (most importantly) the complex interaction of these subjective elements is what is needed to determine the difference between an enjoyable and a pleasurable experience (1975a, p. 6), *e.g.*, the difference between taking pleasure in eating, which everybody can do, and enjoying food, which is more difficult (Csíkszentmihályi, 1990a).

Thus, in a quest to determine what makes an activity enjoyable, Csíkszentmihályi (1975a) was set on exploring activities that appeared "to contain rewards within themselves, that do not rely on scarce material incentives" to bridge the gap between 'work' and 'leisure' (or productive and unproductive activities, as he implied), among which 'play' had a reconciliating role (1975a). Such activities were called autotelic activities[8], from the Greek *auto* = self and *telos* = goal, purpose. Accordingly, Csíkszentmihályi first defined an activity as autotelic if it "required formal and extensive energy output on the part of the actor, yet provided few if any conventional rewards" (1975a, p. 10), putting the accent on the energy cost and the little reward associated.

We make a parenthesis to clarify that Csíkszentmihályi did not wander about this "intrinsically" motivated behavior in a scientific void and that other researchers also took an interest into this phenomenon (Engeser & Rheinberg, 2008). Here follows a brief account of two related works:

In his article, Hebb (1955) looks at motivation, as it relates to the conceptual nervous system (from two different time periods: pre-1930s and between 1930 and 1950), and writes on

[8] *vs.* exotelic activities: "activities done for external reasons only" (Csíkszentmihályi, 1990a, p. 67)

the phenomenon of "work for its own sake" and how it emerges on the conditions that 1) 'work' does not refer only to "activity imposed", and 2) if "the timing is controlled by the worker himself". He bluntly (for today's standards) concludes on his research that "animal data show that it is not always a matter of extrinsic reward; risk and puzzle can be attractive in themselves, especially for higher animals such as man" (1955, p. 250).

Contemporary to Csíkszentmihályi's book, known researchers Deci & Ryan (1980) found that performance-contingent rewards actually enhance intrinsic motivation if administered in such a way that an emphasis is placed on effective performance rather than on acquiring the reward. At the time, their literature review on the nature of intrinsic motivation "highly" supported the "competence and [the] self-determination formulation of intrinsic motivation" (1980, p. 76). This research crystallized later in 1985 on their subsequent works on Self-Determination Theory (Deci & Ryan, 1985) -distinguishing on the different types of motivation, of which intrinsic and extrinsic are but just one basic distinction, and on Cognitive Evaluation Theory -a sub theory of self-determination theory- which argues that rewards and/or feedback conduce towards feelings of competence (Ryan & Deci, 2000), neither of which we cover here. Nevertheless, Deci & Ryan (1985) reviewed Csíkszentmihályi's works on flow as "another strand of the current perspective" on intrinsic motivation research (1985, p. 29). They saw flow as a net occurrence of intrinsic motivation, sharing central characteristics, but mostly a cause-effect relationship:

> In sum, interest and excitement are central emotions that accompany intrinsic motivation, and the concept of flow represents a descriptive dimension that may signify some of the purer instances of intrinsic motivation. When highly intrinsically motivated, organisms will be extremely interested in what they are doing and experience a sense of flow. (Deci & Ryan, 1985, p. 29)

We close the parenthesis now and back to our telling on flow.

Csíkszentmihályi (1975a) noticed that for an individual to engage in an autotelic activity, the individual needed to be "responsive to intrinsic rewards" (1975a, p. 21). This led him in turn to update his definition of what an autotelic activity is to a "pattern of action which maximize immediate, intrinsic rewards to the participant" (1975a, p. 21), without negating the activity's possible extrinsic rewards., *i.e.*, agreeing that the same activity could potentially provide both intrinsic and extrinsic rewards.

Thus, autotelic activities were characterized by a sense of discovery, exploration, problem-solving; individuals explore their own limits and try to go further: "The underlying similarity that cuts across these autotelic activities, regardless of their formal differences, is that they all give participants a sense of discovery, exploration, problem solution –in other words, a feeling of novelty and change" and "[…] they are all exploring the limits of their abilities and trying to expand them" (1975a, p. 30).

Clearly involving the individual[9] in all his proposed definitions, Csíkszentmihályi (1975a) added that "an autotelic person is one who is able to enjoy what he is doing regardless of whether he will get external rewards for it" (1975a, p. 22).

Further on, the results of his research led him to conclude that:

1. The bridge between (autotelic) activity and (autotelic) people was the autotelic experience, which he defined as a "psychological state, based on concrete feedback, which acts as a reward in that it produces continuing behavior in the absence of other rewards. The reality of this experience permits us to conceive of autotelic activities and [autotelic] persons." (Csíkszentmihályi, 1975a, p. 23)

2. Flow **is** the autotelic experience: "Flow is what we have been calling 'the autotelic experience'." (Csíkszentmihályi, 1975a, p. 36)

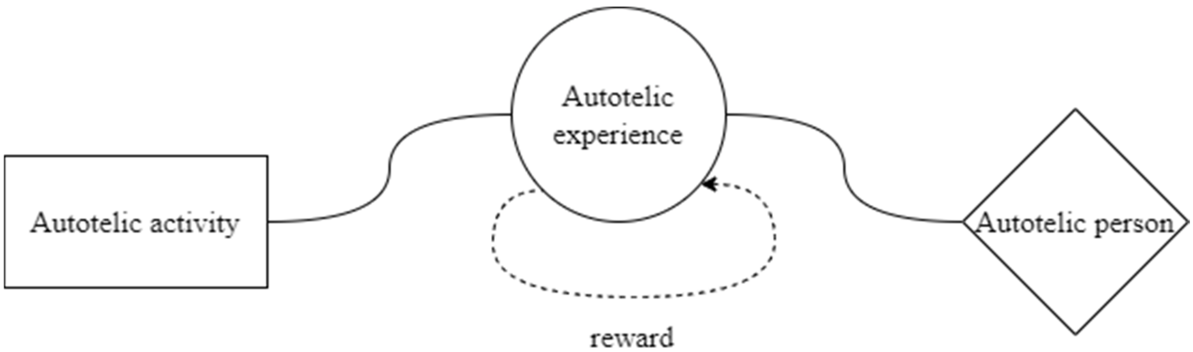
[9] Also named "participant" or "person" by Csíkszentmihályi.

3. The data from these findings appeared to support the then-existing theories of intrinsic motivation (1975a, p. 32).

We illustrated the (1) bridge in Figure 1-1, where the autotelic experience is represented as a self-rewarding experience linking the autotelic activity and the autotelic person[10]:

Figure 1-1

Autotelic Experience (Flow) Bridging the Autotelic Activity and the Autotelic Person



Note: Employed shapes carry no meaning and serve only to differentiate the elements depicted in the illustration.

It is thus during the study of the autotelic experience, which brings enjoyment out of itself, while propelled by an intrinsic motivation that a first (and initial) definition of flow materialized: "Flow is the autotelic experience, a psychological state, based on concrete feedback, which acts as a reward in that it produces continuing behavior in the absence of other rewards" (Csíkszentmihályi, 1975a, p. 36).

In this Section we saw the context in which the initial research of flow was situated: the paradigm shifts from a product-based to the study of creativity; from an extrinsic to an intrinsic motivation. We reviewed how Csíkszentmihályi came about his findings on flow when working on creativity, enjoyment, and intrinsic motivation. These findings (Csíkszentmihályi, 1975a) are

[10] The models of autotelic personality (Baumann, 2012; Nakamura & Csíkszentmihályi, 2002; D. C. Tse et al., 2022) do not concern this thesis.

noteworthy in that firstly, they relate the autotelic experience to a psychological state; secondly, they connect all three autotelic definitions (activity, people, and experience) into a single model in a rather unassuming manner; and thirdly, they define flow and its (initial) components as elements of enjoyment.

Evolution of the Definition of Flow and its Conceptual Flow Model

The definition of flow has continuously changed since its inception by incorporating findings of dedicated researchers over the years. This development has also encompassed an actualization on the conceptual model[11] of the flow state, and on flow theory[12]. Furthermore, flow's applicability on practically all areas of life led quasi inevitably to various flow models adapted to these specific fields. In this Section we follow the evolution of the general definition of flow and of its accompanying conceptual flow model (Moneta, 2021, p. 64), leaving aside the flow models related to specific fields. For the parts[13] composing flow, we dedicate Section "Evolution of Components of Flow" to track their progression.

From the previous Section "Antecedents of Flow", we gather that in 1975, the father of flow Mihály Róbert Csíkszentmihályi, defined flow as the autotelic experience: "Flow is what we have been calling 'the autotelic experience'" (Csíkszentmihályi, 1975a, p. 36). So far at that point in his life, he had just defined the autotelic experience (from the Greek *auto* = self and *telos* = goal, purpose) as a "psychological state, based on concrete feedback, which acts as a reward in that it produces continuing behavior in the absence of other rewards" (1975a, p. 23).

Furthermore, he continued on to re-define flow as a peculiar dynamic state; "the holistic sensation that people feel when they act with total involvement" (1975a, p. 36), highlighting its

---

[11] Flow researchers named it "conceptual flow model" or "theoretical flow model" but mostly simply employ the term "flow model" (Csíkszentmihályi, 1975a, 1975b) or "Flow Model" (initial uppercase). In this thesis, we follow the flow researchers' convention and, otherwise noted *e.g.*, Machine Learning flow-detecting model, we imply the term "conceptual flow model according to its authors" when referring to a "flow […] model". Also, we respect the authors' original uppercase writing when citing primary sources, *e.g.*, "Model of the Flow State".
[12] *A.k.a.* Flow Theory, a subject out of the scope of this thesis.
[13] *A.k.a.* elements, components, characteristics, dimensions, features, traits, …

temporal continuity ("[a] flowing from one moment to the next"), its characteristic sense of control ("in control of his actions"), the loss of self-consciousness associated ("[…] little distinction between self and environment, between stimulus and response, or between past, present, and future"), and its defining autotelic trait ("flow is what we have been calling 'the autotelic experience'") (1975a, p. 36).

Csíkszentmihályi (1975a, 1975b) presented a first "Model of the Flow State" (shown in Figure 1-2) explaining (in depth) that the state of flow sits as a narrow band at the proportional balance between the skills of an individual and the perceived challenges this individual faces. When the perceived challenges are beyond what the person can handle, the individual enters a state of worry. If the perceived challenges increase even more, the individual enters a state of anxiety. Similarly, when the individual possesses great skills but few opportunities to put them in motion, boredom ensues. If the opportunities become even less challenging, anxiety will again set foot. In contrast, the state of flow is experienced when people perceive challenges evenly matched by their skills:

> When a person is bombarded with demands which he or she feels unable to meet, a state of anxiety ensues. When the demands for action are fewer, but still more than what the person feels capable of handling, the state of experience is one of worry. Flow is experienced when people perceive opportunities for action as being evenly matched by their capabilities. If, however, skills are greater than the opportunities for using them, boredom will follow. And finally, a person with great skills and few opportunities for applying them will pass from the state of boredom again into that of anxiety. It follows that a flow activity is one which provides optimal challenges in relation to the actor's skills. (Csíkszentmihályi, 1975a, p. 50)

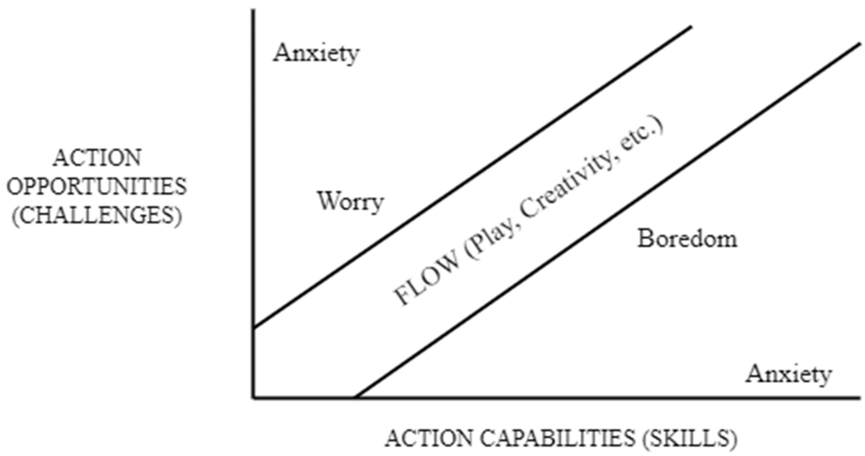
This first model relies on the following two axioms:

1. People are aware of the "finite number of opportunities which challenge them to act" (1975a, p. 50, 1975b, pp. 55–56).

2. People are aware of their skill, as in the “capacity to cope with the demands imposed by the environment” (1975a, p. 50, 1975b, pp. 55–56).

Figure 1-2

Model of the Flow State (Csíkszentmihályi, 1975a, 1975b)



This first model also highlights the difficulty of determining the flow state, as the nature of the challenges and the skills depend entirely on the individual’s perception of these challenges and skills:

> […] [flow] does not depend entirely on the objective nature of the challenges present or on the objective level of skills. In fact, whether one is in flow or not depends entirely on one’s perception of what the challenges and skills are. (Csíkszentmihályi, 1975a, p. 50, 1975b, p. 56)

We will come back to this point often (*cf.* “The Challenge-Skill Balance” below & “Measurement Attempts of Flow” below) as it importantly relates to the work presented in this thesis.

For the sake of completeness, we take a moment to approach the two initial frameworks of flow theory. Up to this moment, flow had been defined in terms of the “optimal experience”, the “autotelic experience”, or the “challenge-skill balance” but in the pivotal book by the

Csíkszentmihályi couple the terms "psychic negentropy"[14] and "emergent teleonomy of the self" are introduced (Csíkszentmihályi & Csíkszentmihályi, 1988). Psychic negentropy defines flow in terms of happiness and enjoyment, plus of its autotelic nature:

> [Psychic negentropy (flow) is] when all the contents of consciousness are in harmony with each other, and with the goals that define the person's self. These are the subjective conditions we call pleasure, happiness, satisfaction, enjoyment. Because the tendency of the self is to reproduce itself, and because the self is most congruent with its own goal-directed structure during these episodes of optimal experience, to keep on experiencing flow becomes one of the central goals of the self. (Csíkszentmihályi & Csíkszentmihályi, 1988, p. 24)

While "teleonomy of the self" defines flow as "the goal-seeking tendency that shapes the choices we make among alternatives" and "a set of goals that have been freely chosen by the individual" (1988, p. 24; Nakamura & Csíkszentmihályi, 2002, p. 91), *i.e.*, a motivational system leading to "reorganization and growth in the order and complexity of consciousness" (Moneta & Csíkszentmihályi, 1996, p. 277), experienced as enjoyable. Neither definitions are required to comprehend this thesis which does not tackle the ontology of flow (Moneta & Csíkszentmihályi, 1996; Nakamura & Csíkszentmihályi, 2002) but its applications. Back to our tale.

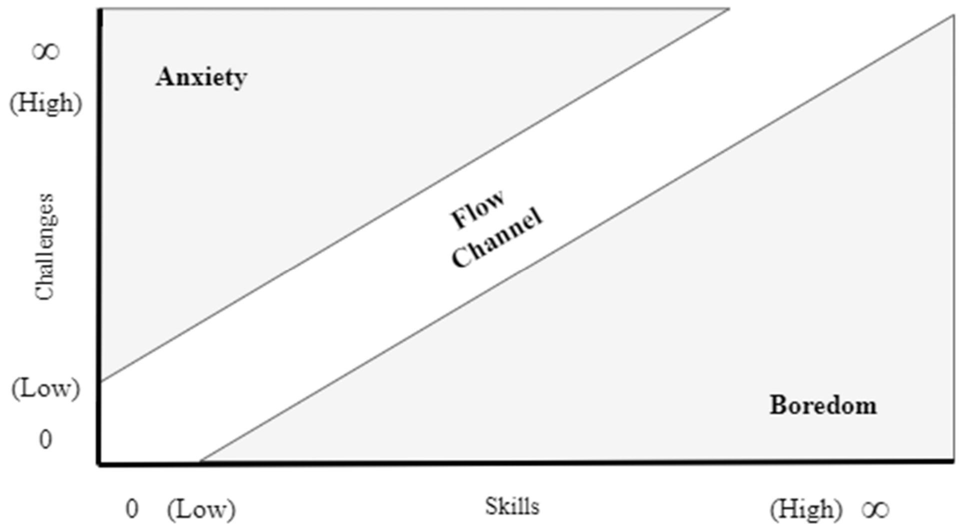
Given the clear oblique shape where flow is located, this first Model of the Flow State was reimagined in a successive Csíkszentmihályi publication (1990a, p. 74) as the Flow Channel Model (Peifer et al., 2022), shown in Figure 1-3. In this diagram, axis labels are now clearly identified as Challenges and Skills (instead of opportunities and capabilities), ranging from zero to infinity (low and high skills and challenges, respectively), making of flow a linear and growing function of skills and challenges. Flow is thus represented as an area; more precisely a strip delimited by the four points consisting of the combinations of minimal and maximal skills and

[14] "Negative entropy or a state of order"(Csíkszentmihályi & Csíkszentmihályi, 1988).

challenges. Just like in its predecessor, flow is not a fixed point in this space, but a broad state achieved by a multitude of possible combinations of skills and challenges. Flow is surrounded by anxiety and boredom, where the balance of skills and challenges is no longer (shaded areas).

Figure 1-3

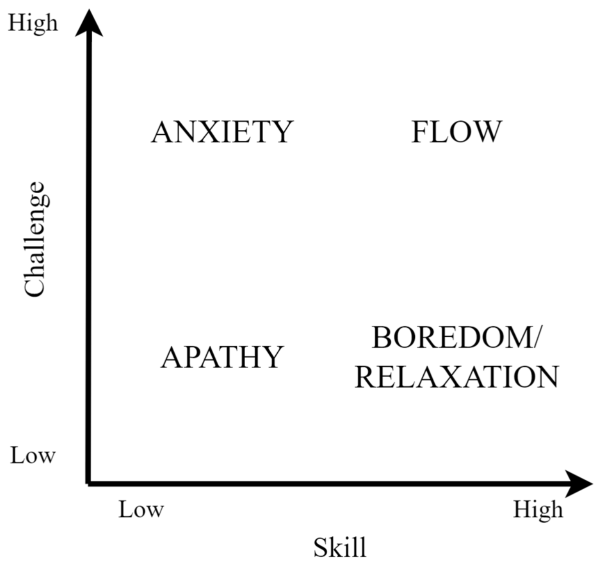
The "Flow Channel" Diagram (Csíkszentmihályi, 1990a, p. 74)



In more specific terms, this diagram (and its predecessor) can be seen as divided into four possible quadrants, as per the low/high combinations of challenges and skills, as shown in Figure 1-4: apathy, anxiety, boredom/relaxation, and finally flow ([low-skill, low-challenge], [low-skill, high-challenge], [high-skill, low-challenge], and [high-skill, high-challenge]). This quadrant splitting followed the rectification by Massimini & Carli (1988) made to the previous models in that the notion of skill stretching was inherent to the flow concept: flow could not exist at the bottom, close to the origin, of the Channel Model because "Activities providing minimal opportunities for action do not lead to flow, regardless of whether the actor experiences a balance between perceived challenge and skill" (Nakamura & Csíkszentmihályi, 2002, p. 94).

Figure 1-4

Quadrant Model of Flow (Csíkszentmihályi & Csíkszentmihályi, 1988)



Note: This model is an adaptation (Engeser & Rheinberg, 2008) of a previous work by Massimini & Carli (1986).

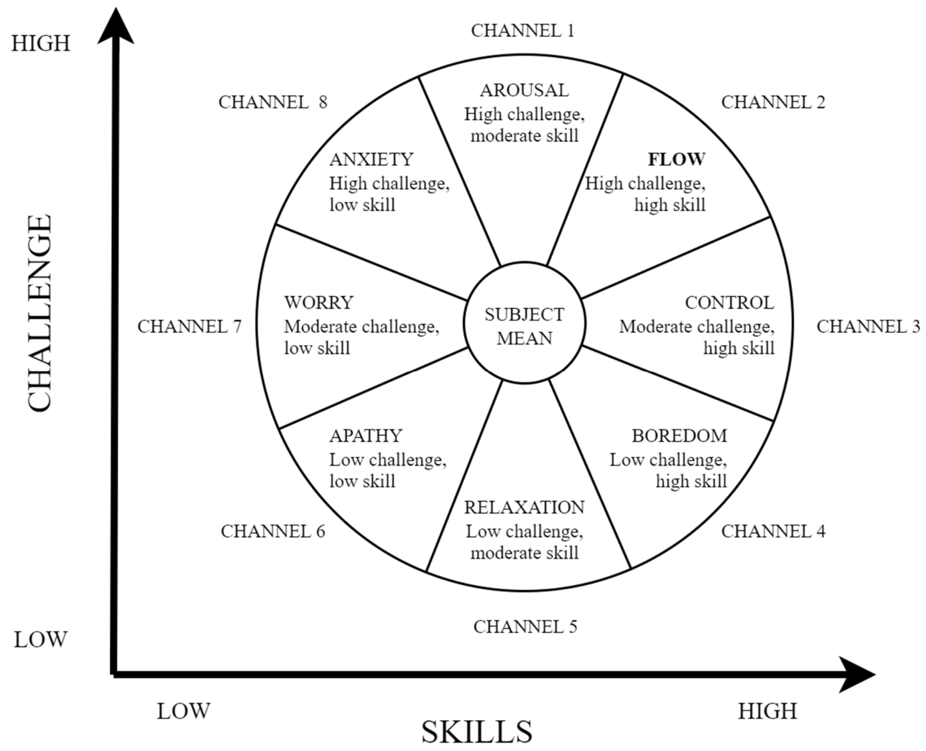
Indeed, since 1988, researchers Massimini & Carli (1988) argued that if flow was defined as a balance of challenges and skills, such balance had to happen at above average levels for the individual and not simply at any given intersection, e.g., watching TV (a low challenge activity balanced by a low skill) is not conducive to flow. Certainly, this premise revealed the existence of a not-yet-tackled state of "apathy" associated with low challenges and correspondingly low skills, a milestone conclusion that persisted into more than a decade later (Nakamura & Csíkszentmihályi, 2002, p. 95).

Thus, Massimini & Carli (1988) presented a "Model for the analysis of experience" (shown in Figure 1-5) comprising eight channels superseding the previously mentioned four quadrants (1988, p. 270). These channels arise from the following combinations of the low/moderate/high combinations of challenges and skills, labeled as (starting with Channel 1): arousal (high challenge, moderate skill), flow (high challenge, high skill), control (moderate challenge, high skill), boredom (low challenge, high skill), relaxation (low challenge, moderate skill), apathy (low challenge, low skill), worry (moderate challenge, low skill), and anxiety (high challenge, low skill). A major contribution of this model is the revelation that apathy is the inverse of the flow state: "[…] apathy, associated with low challenges and correspondingly low skills […] is a sphere of stagnation and attentional diffusion, the inverse of the flow state" (Nakamura & Csíkszentmihályi, 2002, p. 95).

Figure 1-5 shows self-perceived skills on the *x*-axis and the self-perceived difficulty of the challenge on the *y*-axis. The center of the entire figure is determined by "the average level of the individual's weekly challenges and skills", flow is on Channel 2, and the study control is situated in Channel 3 (moderate challenge, high skill) (Csíkszentmihályi & Csíkszentmihályi, 1988; Massimini & Carli, 1988, p. 270).

Figure 1-5

A Model for the Analysis of Experience (Massimini & Carli, 1988, p. 270)



In the meantime, in a joint journal article Csíkszentmihályi & LeFèvre (1989), the flow experience is said to be the "optimal experience" citing Csíkszentmihályi's book[15] (1975a). This definition is revisited in the Csíkszentmihályi's 90's publication, where flow is called "the process of total involvement"[16] first, and later called "the optimal experience": "[…] for this reason that we called the optimal experience 'flow'" (1990a, p. 53).

Shortly after, Csíkszentmihályi & Rathunde (1992) employed the "analysis of experience model" described above for "determining flow" in a longitudinal study on teenagers considering stopping/continuing to develop their talents (1992, p. 69). They argued that the eight channels present in the model (*cf.* Figure 1-5) were meant to adjust the definition[17] to study flow, *i.e.*, the states surrounding flow and what an individual might experience if not in flow. In this same publication, flow was defined in terms of total involvement in the activity, and of its autotelic nature:

> Flow is a subjective state that people report when they are completely involved in something to the point of forgetting time, fatigue, and everything else but the activity itself. […] The depth of involvement is something we find enjoyable and intrinsically rewarding. (Csíkszentmihályi & Rathunde, 1992, p. 59)

In 1996, flow is then defined "as a psychological state in which the person feels simultaneously cognitively efficient, motivated, and happy" (Moneta & Csíkszentmihályi, 1996, p. 277). Furthermore, they recognized the importance of the challenges-skill balance: "In situations characterized by the simultaneous presence of high perceived challenges and high perceived skills, the person experiences flow in consciousness and the overall quality of subjective experience is the highest" (1996, p. 277).

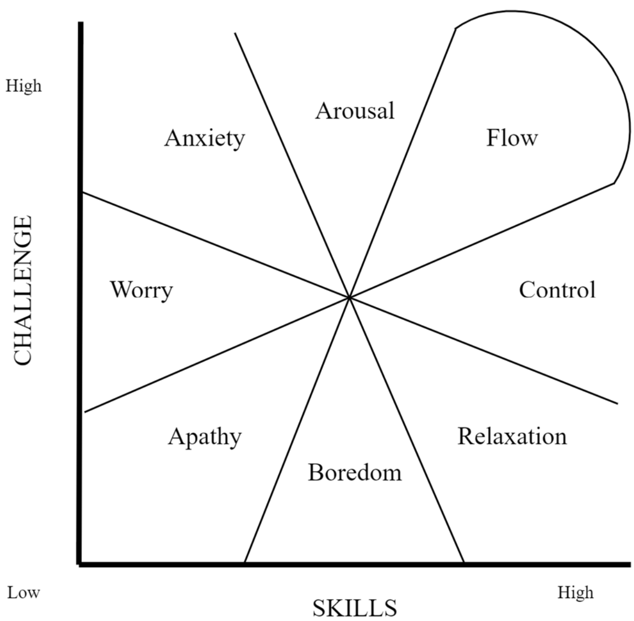
---

[15] Although such definition is not found in that specific citation, the view of flow as an "optimal situation" did appear in that book, and it is described as "when the challenges match skills" (Csíkszentmihályi, 1975a, p. 66).
[16] This first, quick definition, followed by the already-familiar definition, might obey to the non-academic nature (according to the author) and vulgarization intent behind the book.
[17] In this context, a more appropriate term would be 'resolution'.

The next year, an adaptation to this model is published in Csíkszentmihályi's book (1997) and it is named "Quality of experience", shown in Figure 1-6. Notice that the segments corresponding to Relaxation and Boredom in Figure 1-6 have switched places compared to Figure 1-5. Originally (1975), it was thought that low challenges and high skills would result in boredom but here, Csíkszentmihályi (1997) acknowledged that many studies, notably those published in his then decade-old book (Csíkszentmihályi & Csíkszentmihályi, 1988), among others, showed that "people report feeling relaxed in [low challenges and high skills] situations, whereas boredom tends to occur more when both challenges and skills are low" (Adlai-Gail, 1995; cited by Csíkszentmihályi, 1997).

Figure 1-6

Quality of Experience (Csíkszentmihályi, 1997, p. 31)

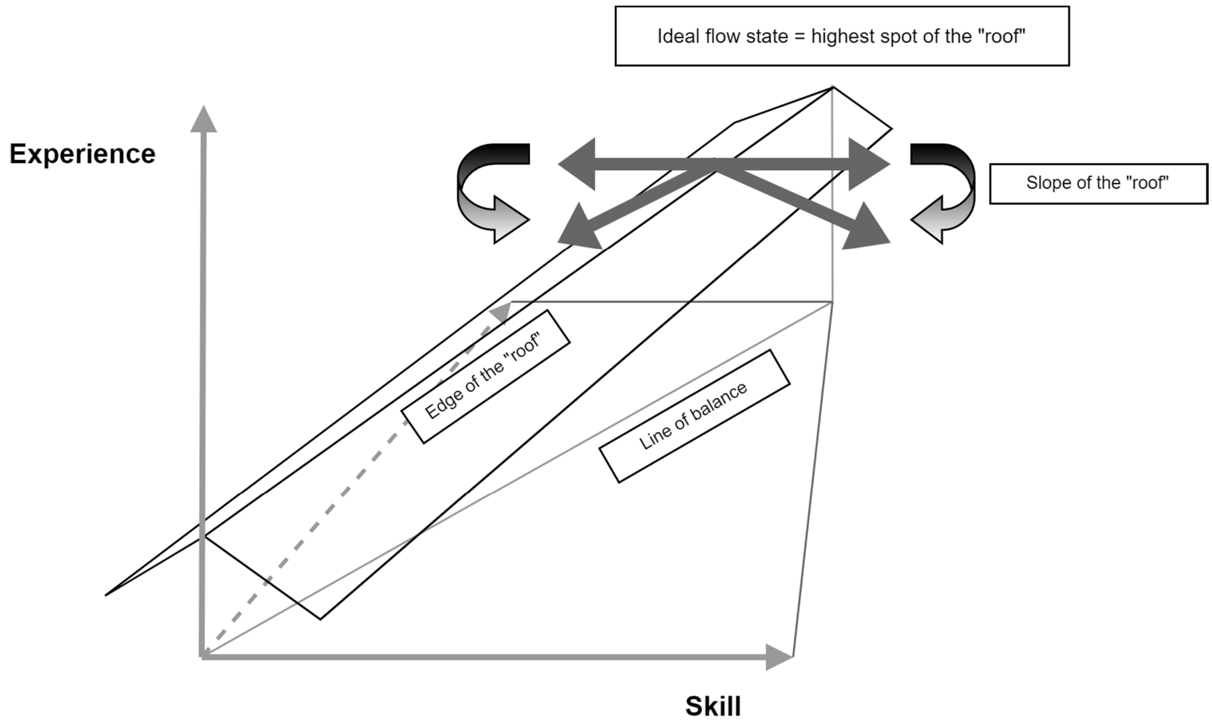


Already since 1996, flow researchers (Moneta & Csíkszentmihályi, 1996) evoked the metaphor of an action happening on the edge of a reclined roof: although failure or success depend on few factors, accounting for the optimal challenge-skill balance ratio as well as for the individual's previous experience, the individual would perceive progressively the goal as likely reachable. Thus, for the individual to remain at the edge of this roof-like surface (where the optimal challenge-skill balance ratio occurs), newer and more challenging goals would be set up by the individual. So, this hypothesized roof-like surface represents in their model the "ideal case in which challenge and skill have identical and positive coefficients and the absolute difference of challenge and skill has a negative coefficient" (Moneta, 2012a, p. 37).

Moneta (2012a) presents a "three-dimensional representation of the absolute difference regression model of the flow state" (2012a, p. 38; adapted from Moneta & Csíkszentmihályi, 1996), shown in Figure 1-7. Notice that the usual plane of challenges and skills is now lying flat, and a new axis 'Experience' arises perpendicularly. In this model case, the top-most edge of the roof-like surface represents the optimal challenge/skill ratio, and then, its highest point would represent "the ideal flow state" (2012a, p. 37).

Figure 1-7

3D Representation of the Regression Model of the Flow State (Moneta, 2012a, p. 38)



Note: This model is an adaptation (Moneta, 2012a, p. 38) of a previous work by Moneta & Csíkszentmihályi (1996).

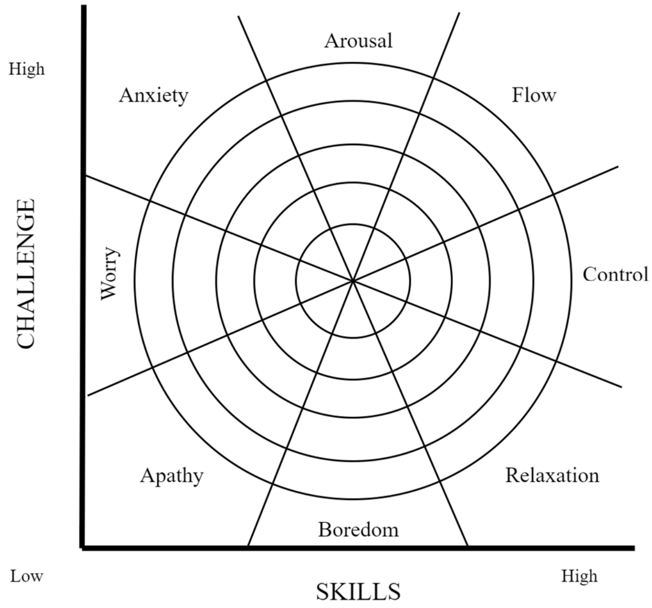
Through the various works of Ceja & Navarro (2009, 2011, 2012) this model would eventually evolve into the "Cusp catastrophe model of flow"[18] (Moneta, 2021, p. 26). Contrary to the previous theoretical flow models, these previous two non-linear models would predict that, as the entire system gets further away from an "equilibrium point", the expected behavior of the model would become increasingly unstable (instead of continuous and smooth) up to the point where additional change in the input variables would lead to an abrupt and discontinuous change (akin to a 'jump').

Based on previous, widely recognized milestone works (Csíkszentmihályi, 1990a, 1997; Massimini et al., 1988), Nakamura & Csíkszentmihályi (2002, p. 95) presented the "current model of the flow state", shown in Figure 1-8. A series of concentric rings placed over the previous model and emanating from the starting point of the segments represented the intensity of the experience, which increases from the center of the concentric circles, *i.e.*, outer rings concern deep flow experiences whereas rings closer to the center relate to microflow activities. The center of the figure is determined by the individual's average level of challenges and skills (2002, p. 95).

[18] The "Cusp catastrophe model of flow" is not shown here because of its complex representation.

Figure 1-8

Model of the Flow State (Nakamura & Csíkszentmihályi, 2002, p. 95)



Note: This model is an adaptation (Nakamura & Csíkszentmihályi, 2002, p. 95) of a previous work by Csíkszentmihályi (1997).

Findings derived from the application of this model kept pointing towards the initial and "essential insight" that "perceived challenges and skill must be relative to a person's own average levels" (Nakamura & Csíkszentmihályi, 2002, p. 95), confirming the subjective nature of the flow state. However, Benjafield & Moneta (2023) recently suggested a Flow Immersion Model grounded in neuroscience aiming to study flow objectively and not as a subjective phenomenon (dependent on the individual's perception), in terms of neural and cognitive processes. This position intends to challenge the commonly accepted so-called "subjective-experiential" (2023, p. 2) paradigm adopted by the flow research community. Instead, their model emphasizes the productive outcome of flow states to conclude on its "optimal" quality by quantitatively testing the "objective-economic usefulness of flow" (2023, p. 2). Yet, authors of this model admit, as per its novelty, it requires "extensive empirical testing" (2023, p. 4) and thus, remains largely unassessed.

Ten years after the confirmation of the subjective nature of the flow state by Nakamura & Csíkszentmihályi, Engeser & Schiepe-Tiska (2012) clearly pointed at the direction to follow when, in the very first paragraph of Chapter 1, bearing no introduction, immediately stated the ensuing Concept of Flow:

> Flow is a state in which an individual is completely immersed in an activity without reflective self-consciousness but with a deep sense of control. (Engeser & Schiepe-Tiska, 2012, p. 1)

Almost a decade later, again, bearing no introduction, in the very first paragraph of their Chapter 1, Engeser, Schiepe-Tiska, & Peifer (2021, p. 2) reprise it almost identically[19]:

> Flow is a state in which an individual is completely <u>absorbed</u> in activity without reflective self-consciousness but with a deep sense of control. (Engeser et al., 2021, p. 2)

Despite the multiple takes on flow along the many years since its birth, Engeser *et al.* (2021; 2012) underlined the high level of agreement on the flow definition since its inception in 1975, and also agreed on the advantageous flexibility of such multidimensional definition. Flexibility allowed room for elements to be added or removed without completely altering the meaning of the flow notion, a trend that must be by now clear to the reader: "the definition of flow, with different components, provides the flexibility to pronounce a particular component or add new components without completely changing the definition" (Engeser & Schiepe-Tiska, 2012, p. 2).

Nevertheless, the notion of flow needed a commonly agreed definition. The European Flow-Researchers' Network[20] (EFRN) "was founded in 2012 with the aim to reach a common understanding of the concept of flow, its antecedents and consequences" (Engeser et al., 2021),

[19] The minor difference between both definitions is <u>underlined</u>.
[20] https://efrn.eu/

based on rigorous scientific standards (EFRN, 2014). With the backing of Csíkszentmihályi himself as a founding member, one of their first points in the agenda was to agree on a scientific definition of flow, by researchers.

The resulting flow definition[21] that emerged out of these scientific debates spans from its origins to its multiple repercussions in daily life, while accounting for what now are considered flow's main components. Scientifically, it encompasses the notions of focused attention, challenge-skill balance, intrinsic motivation, and autotelic experience:

> Flow was first described by Mihaly Csikszentmihalyi in his book Beyond Boredom and Anxiety in 1975. It is a gratifying state of deep involvement and absorption that individuals report when facing a challenging activity and they perceive adequate abilities to cope with these challenges. Flow is described as an optimal experience during which people are deeply motivated to persist in their activities. Research shows that flow experiences can have far-reaching implications in supporting individuals' growth, by contributing both to personal well-being and full functioning in everyday life. (EFRN, 2014)

In this statement, flow is first and foremost, a human state[22] that must involve gratification[23]; it is characterized by an intense concentration, focus, and merged awareness on the activity, experienced individually[24]; it appears when the subjective, self-perceived abilities of the individual seem sufficient to successfully handle the challenges[25] presented by the activity. As such, flow depends on the individual' subjective experience of the activity, which is to be the optimal experience; it lasts a time span non null; and it involves a strong intrinsic motivation to continue the activity, even with increasing levels of difficulty.

---

[21] While the definition correctly attributes flow fatherhood to Csíkszentmihályi, Hungarian diacritics are missing in the original 2014 source quoted here.
[22] Leaving room for psychological and physiological studies of flow.
[23] Replacing initial recurrent terms such as 'enjoyment' and 'pleasure', *cf.* Section "Antecedents of Flow".
[24] At least firstly.
[25] Also, "demands".

We do not forget to include a very recent proposed flow definition by Peifer & Engeser (2021b) resulting from the merging of thirteen flow components coming back all the way from the "early years" into three "core components" (*cf.* Table 1-1 in "Evolution of Components of Flow" below). In their proposal "flow can be defined as the enjoyable experience of full absorption in an activity in which the demands are perceived as optimally compatible with one's skills" (2021b, p. 424). Please notice that in this proposal the challenge-skill balance is the primary precursor of flow (*cf.* "The Challenge-Skill Balance" below) while enjoyment and immersion, and thus flow, are the result of this "optimal" compatibility.

Now, in this thesis, we consider the EFRN's as a comprehensive and scientifically accepted flow definition. Nevertheless, we do not forego the multiple considerations of researchers on what traits might be considered the most significative for characterizing flow. We keep the noteworthy remark of Keller & Landhäußer (2012, p. 52) who, along with Engeser & Rheinberg (2008) prefer the use of 'demands' *vs.* 'challenge': "[…] we consider the term 'demands' much more appropriate than the term 'challenge'".

This decision arose from two reasons; first, the results of the Pfister study (Pfister, 2002) hinted at the term being insufficiently precise, according to Engeser & Rheinberg (2008). In this study, swapping the term 'challenge' for 'difficulty' in 'challenge-skill' and 'difficulty-skill' determined that "the participants reported similar experiences, and one could therefore argue that it makes no (empirical) difference whether one asks about challenge or difficulties" (2008, p. 159). Second, when assessing the balance between challenge and skill "by asking whether the **demands** of the task are too low, just right or too high" (2008, p. 161), individuals were able to report the perceived balance "more accurately" than with the two "abstract variables of difficulty and skill" (Ellis et al., 1994), of which 'difficulty' had already proven unprecise (Engeser & Rheinberg, 2008, p. 161).

In this sense, although we fully agree on the more adequate term ‘demands’, for the sake of consistency we keep the often-employed in the literature term ‘challenge’, notably for the dimension historically named ‘challenge-skill balance’, which will come often in this thesis.

Thus, we consider in this thesis flow as “a gratifying state of deep involvement and absorption that individuals report when facing a [demanding] activity and they perceive adequate abilities to cope with [the] challenges [presented by this demanding task]” (EFRN, 2014) characterized by an autotelic experience.

In this Section we surveyed the evolution of the flow notion and its accompanying theoretical flow model: flow was initially seen mostly in terms of an “autotelic experience” (Csíkszentmihályi, 1975a, p. 36, 1975b, pp. 53–55), characterized by a challenge-skill balance (Csíkszentmihályi, 1975a, p. 36, 1975b, pp. 55–58, 1990a, p. 74), which led to the initial Model of the Flow State and the Flow Channel diagrams (*cf.* Figure 1-2 & Figure 1-3). This challenge-skill balance was subsequently seen by Massimini & Carli (1986, 1988) as requiring above-average level challenges for above-average levels of skills of the individual instead of a generic balance. The Quadrant Model of Flow (*cf.* Figure 1-4) hinted first at this difference, which was later reflected on the model for analysis of experience, and the Quality of experience model (*cf.* Figure 1-5 & Figure 1-6).

In his 90’s book, Csíkszentmihályi (1990a), flow is said to be the “optimal experience”, then characterized by a complete involvement in the activity, which is of autotelic nature (Csíkszentmihályi & Rathunde, 1992, p. 59). Moneta & Csíkszentmihályi regarded flow as a psychological state, and acquired the dimensions of cognitive efficiency and happiness (Moneta & Csíkszentmihályi, 1996, p. 277). By the turn of the century, the Model of the flow state (*cf.* Figure 1-8) is considered the current theoretical model of flow. Finally, the works of the EFRN (2014) brought up the currently accepted definition of flow, which is the one primarily employed in this thesis, in a general context.

Such evolution also echoed in the number and meanings behind the components of flow, which we trace in the following Section.

An attentive reader will remark that the evolution of the definition of flow roughly told in this Section is highly reliant on Csíkszentmihályi's collaboration and/or endorsement and that is not a coincidence. Certainly, beyond his passing in October 2021, Mihály Róbert Csíkszentmihályi is still regarded, not only in academic circles, as highly contributing to the development of Positive Psychology (Seligman & Csíkszentmihályi, 2000; *Thinker of the Year - 2000: Mihaly Csikszentmihalyi*, 2000, para. 3) but also in pop culture (GoogleDoodle, 2023; Just Dance 2024 Edition [@justdancegame], 2022) as a leading researcher and an influential figure.

Evolution of Components of Flow

In this Section we survey what flow researchers (starting with Csíkszentmihályi) historically considered to be part of the flow experience. We ask the reader to bear with us in what would appear to be recurrent terms listing but our aim is to convey the difficulty in coming to terms at the currently accepted research framework employed in this thesis. We try to diminish the volume of descriptions as we move on in the text and as the reader gets acquainted with their definitions.

As you will see, although the number of flow components, their listing order, and their respective definitions have evolved, the general structure is noticeably consistent, with terms coming repeatedly, *e.g.*, "optimal experience", "autotelic experience" (both are considered flow), and "challenging activity". Nevertheless, to better understand the ensuing, we consider important to remind a key definition discussed in Section "Antecedents of Flow" above:

> The term "autotelic" derives from two Greek words, *auto* meaning self, and *telos* meaning goal. It refers to a self-contained activity, one that is done not with the expectation of some future benefit, but simply because the doing itself is the reward. (Csíkszentmihályi, 1990a, p. 67)

On the terminology employed, we clarify that Csíkszentmihályi originally named these "Elements of Flow Experience", and often used the terms "characteristic", "component", "trait", or "quality" interchangeably when referring to them[26]. In a fidelity effort, we keep the original term employed by its authors whenever it does not conflict with another notion in this text.

In his first production and based on interviews on the subject, Csíkszentmihályi (1975a, pp. 36–49) listed and named six flow elements (order matters). We quickly describe them after their name, citing the people's experiences that led him to name and characterize them so:

1. Merging of action and awareness, where a person, focused on an activity, has merged (for short periods of time) into his actions but disassociated from his consciousness of the merging: "aware of his actions but not of the awareness itself". While defining this element, Csíkszentmihályi noticed that a precondition seemed to be needed: the ability to perform. We acknowledge here the outline of what would become later an additional flow element: the challenge-skill balance.

2. Centering of attention on a limited stimulus field ("narrowing of consciousness", "giving up the pas and the future"), where participants perceive external stimuli diminished or blocked out, affecting memory ("as if my memory input has been cut off"), hearing ("I don't seem to hear nothing"), sight ("I see only the positions"), and problem-awareness ("Problems are suspended for the duration of the tournament"). Csíkszentmihályi also assures this element to be a pre-condition of the merging of action and awareness element ("is made possible by").

3. Loss of self-consciousness ("loss of ego", "loss of self-consciousness", "self-forgetfulness", or even "transcendence of individuality" or "fusion with the world") although this is not to be misinterpreted as a losing of touch of one's physical reality but instead, as the (temporary) losing of the considerations for the self, in favor of the activity.

[26] Other terms found in the literature are "dimension" or "feature".

4. Sense of control of his actions and of the environment ("feeling of control", "being in control", "being merged with the environment", "sense of control"), where the possibility of lack of control does not constitute a worry.
5. Coherent, noncontradictory demands for action and clear, unambiguous feedback, where the flow experience distinguishes itself from "everyday reality" in that the flow experience "contains ordered rules" which make of the possible actions and of the valuation off such actions "automatic and hence unproblematic", *e.g.*, if during a football game (from a player point-of-view) the referee adds three goals in favor of the rival team "the self reappears to negotiate between the conflicting definitions [of the rules of the game]" and the steps to take, and flow is interrupted.
6. The "autotelic nature" of the flow experience, where the activity "appears to need no goals or rewards external to itself" and no goal but the activity in and of itself: "The purpose of the flow is to keep on flowing, not looking for a peak or utopia but staying in the flow". (Csíkszentmihályi, 1975a, pp. 36–49)

It is important to underline that in this very first approach, Csíkszentmihályi (1975a) considered these elements "linked [..] and dependent on each other", but most significantly he emphasized that, among them all, he considered the merging of action and awareness as the "clearest sign of flow" (1975a, p. 38).

Later on, Csíkszentmihályi came to distinguish seven elements of the "phenomenology of enjoyment"[27], plus an autotelic experience (Csíkszentmihályi, 1990a, pp. 49–70). His general description on people's reflection on that experience goes like this:

> First, the experience usually occurs when we confront tasks we have a chance of completing. Second, we must be able to concentrate on what we are doing. Third and fourth, the concentration is usually possible because the task undertaken has clear goals

[27] *Cf.* Abuhamdeh discusses on the relationship between flow and enjoyment (2021b).

> and provides immediate feedback. Fifth, one acts with a deep but effortless involvement that removes from awareness the worries and frustrations of everyday life. Sixth, enjoyable experiences allow people to exercise a sense of control over their actions. Seventh, concern for the self disappears, yet paradoxically the sense of self emerges stronger after the flow experience is over. Finally, the sense of the duration of time is altered; hours pass by in minutes, and minutes can stretch out to seem like hours. The combination of all these elements causes a sense of deep enjoyment that is so rewarding people feel that expending a great deal of energy is worthwhile simply to be able to feel it. (Csíkszentmihályi, 1990a, p. 49)

Nevertheless, a merger of elements third and four happened in an latter prose description (1990a, pp. 49–70), which led instead to the following list of eight Elements of Enjoyment:

1. A Challenging Activity That Requires Skills, where challenge is defined as a "bundle of opportunities for action", within any given activity. The challenge is to provide enjoyment and must be perceived by the individual as equal to his capabilities: "Enjoyment appears at the boundary between boredom and anxiety" (1990a, p. 52).

2. The Merging of Action and Awareness. Once the necessary skills are recalled for handling the challenging activity, the individual's attention is completely devoted to the activity, leaving no attention left to process any external information but the one issued from the challenging activity. This element is recognized to be "one of the most universal and distinctive features of [the] optimal experience" (1990a, p. 53) and the reason to call the optimal experience flow: "[…] for this reason that we called the optimal experience 'flow'" (1990a, p. 53).

3. Clear Goals and Feedback. Often[28], the challenging activity has a clear, unambiguous set of rules and goals; its feedback on how actions logically affect the activity should provide enjoyment.

4. Concentration on the Task at Hand. Hinted as a consequence ("by-product") of the flow experience, the focus on the flow activity is such that other irrelevant aspects of life disappear temporally from consciousness ("one is able to forget all the unpleasant aspects of life") (1990a, p. 58).

5. The Paradox of Control. Described as a "lacking the sense of worry about losing control" found in other normal life experiences, it implies no worry of failing at the activity, and the possibility of exerting control over it, rather than its actual control (exercising control *vs.* being in control) (1990a, p. 61).

6. The Loss of Self-Consciousness. Akin to the second element (Merging of Action and Awareness), in that the individual's attention is completely devoted to the activity leaving no attention left to process any external information, the sense of self (and its accompanying self-scrutiny concerns) is also removed from processing ("a Zen feeling"). Instead, a union with the activity's relevant environment ("becoming one flesh") is experienced, whether it is other participants, the physical conditions surrounding the activity, and even the individual's own memories and body (1990a, pp. 63–64).

7. The Transformation of Time. During activities that do not depend critically on time, the passing of time is perceived as accelerated or decreased, but definitely changed ("hours seem to pass by in minutes") and with a tendency to be experienced much faster (1990a, p. 66).

[28] Creative endeavors, like painting or music composing, rely on the individual's own sense of intention to set and recognize goals and feedback indicators among the initial vague goals and gauges of feedback.

8. The Autotelic Experience. Finally, what is considered to be the "key element of an optimal experience" is that it is the activity itself (and not its by-products or consequences) that is an end in itself, even if it started with other ends in mind (1990a, pp. 49–70).

Later on, Csíkszentmihályi & Rathunde (1992, p. 60) considered the "Flow Experience" to be composed of eight "Characteristic Dimensions", which have switched order compared to two years prior (*italic* emphasis is from the original text):

1. *Clear goals*: it is clear what should be done; *immediate feedback*: one knows how well one is doing.
2. The opportunities for action are relatively high, and they are met by one's perceived ability to act; *challenges = skills*.
3. *Action and awareness merge*; one-pointedness of mind.
4. *Concentration of the task at hand*; irrelevant stimuli disappear from consciousness, worries and concerns are temporarily eliminated.
5. A sense of potential control.
6. *Loss of self-consciousness*, transcendence of ego boundaries, a sense of growth and of being part of some greater entity.
7. *Sense of time altered*; usually time seems to pass faster.
8. *Experience becomes autotelic*-if several of the previous conditions are present, what one does becomes autotelic, or worth doing for its own sake. (Csíkszentmihályi & Rathunde, 1992, p. 60)

During the validation of their Flow State Scale (*cf.* "The Flow State Scale 2 (FSS2) & the Dispositional Flow Scale (DFS-2)" below), Jackson & Marsh (1996) remarked the frequent changes in the works of Csíkszentmihályi concerning the sense of control dimension: "the labeling of this dimension by Csíkszentmihályi had shifted from being 'in control'

(Csíkszentmihályi, 1975a, p. 44), to the 'paradox of control' (Csíkszentmihályi, 1990a, p. 59), to 'sense of control' (Csíkszentmihályi, 1993, p. 181)". As a result, they refined the description of the term in their definitions of the flow dimensions (1996, pp. 18–20):

1. Challenge-Skill Balance: "In flow, the person perceives a balance between the challenges of a situation and one's skills, with both operating at a personally high level."
2. Action-Awareness Merging: "Involvement in the flow activity is so deep that it becomes spontaneous or automatic. There is no awareness of self as separate from the actions one is performing."
3. Clear Goals: "Goals in the activity are clearly defined (either set in advance or developed out of involvement in the activity), giving the person in flow a strong sense of what he or she is going to do."
4. Unambiguous Feedback: "Immediate and clear feedback is received, usually from the activity itself, allowing the person to know he or she is succeeding in the set goal."
5. Concentration on Task at Hand: "Total concentration on the task at hand occurs when in flow." Intense and focused concentration on what one is doing in the present moment. This is one of the most frequently mentioned flow dimensions (Csíkszentmihályi, 1990a).
6. Sense of Control: "A sense of exercising control is experienced, without the person actively trying to exert control." "What seems critical to this dimension is that it is the potential for control, especially the sense of exercising control in difficult situations, that is central to the flow experience."
7. Loss of Self-Consciousness: "Concern for the self disappears during flow as the person becomes one with the activity." "The absence of preoccupation with self does not mean the person is unaware of what is happening in mind or body, but rather is not focusing on the information normally used to represent to oneself who one is."

8. Transformation of Time: "Time alters perceptibly, either slowing down, […] or speeding up, giving the perception that the event was "over so fast" […]. Alternatively, time may simply become irrelevant and out of one's awareness."

9. Autotelic Experience: "An autotelic experience is an intrinsically rewarding experience. This dimension is described […] as the end result of being in flow. […] An activity is autotelic if it is done for its own sake, with no expectation of some future reward or benefit." (Jackson & Marsh, 1996, pp. 18–20).

An important milestone in flow research came in the joint Nakamura & Csíkszentmihályi (2002, p. 89) publication where a distinction was made among the existing elements of flow, splitting them along the lines of predecessors and qualities named "conditions of flow" and "characteristics of flow". The former included (but not limited to):

- Perceived challenges, or opportunities for action, that stretch (neither overmatching nor underutilizing) existing skills; a sense that one is engaging challenges at a level appropriate to one's capacities.
- Clear proximal goals and immediate feedback about the progress that is being made. (Nakamura & Csíkszentmihályi, 2002, p. 90)

These conditions reiterated the required perceived challenge-skill balance but aggregated the clear goals and unambiguous feedback (nowadays considered separate dimensions) into a single requisite.

In consequence and "under these conditions", the subjective state of flow would be experienced with the ensuing characteristics (2002, p. 90):

- Intense and focused concentration on what one is doing in the present moment.
- Merging of action and awareness.
- Loss of reflective self-consciousness (*i.e.*, loss of awareness of oneself as a social actor).

- A sense that one can control one's actions; that is, a sense that one can in principle deal with the situation because one knows how to respond to whatever happens next.
- Distortion of temporal experience (typically, a sense that time has passed faster than normal).
- Experience of the activity as intrinsically rewarding, such that often the end goal is just an excuse for the process. (Nakamura & Csíkszentmihályi, 2002, p. 90)

Notice that in this view, the distortion of time emphasizes on the "faster than normal" passing of time, when compared to the Jackson & Marsh (1996, pp. 18–20) listing.

On their part, Engeser & Rheinberg (2008) recognized instead six components of flow, with the notable absence of the "autotelic nature" of flow:

> (1) A balance between perception of one's skills and the perception of difficulty of the activity (task demand). […] (2) The activity has coherence, contains no contradictory demands, and provides clear, unambiguous feedback. (3) The activity seems to be guided by an inner logic. (4) A high degree of concentration on the activity due to undivided attention to a limited stimulus field. (5) A change in one's experience of time. (6) The self and the activity are not separated, leading to a merging of the self and the activity and the loss of self-consciousness. (Engeser & Rheinberg, 2008, p. 158)

The definitions of these six components reflected important contributions by Engeser & Rheinberg (2008) of which we present here what we consider most apply to the current research context, notably on the challenge-skill component:

- The term 'demands' substituting the term 'challenges', because of its proven imprecise nature (*cf.* Section "Evolution of the Definition of Flow and its Conceptual Flow Model").

- Flow = challenge-skill balance, always. Engeser & Rheinberg (2008) argue that previous theoretical flow models (*cf.* Figure 1-3 & Figure 1-4) raised some questions regarding the presumption of flow when the challenge-skill balance was met, *i.e.*, the admitted assumption that the challenge-skill balance is binding to flow while it might not always be the case. Their literature review showed that such balance would not automatically be conducive to flow: "[the characterization of flow by the perceived challenge-skill balance] does not necessarily mean that flow is always experienced when this balance is present" (2008, p. 159).

- The presumption of valid flow indicators. Researchers (Ellis et al., 1994; Engeser & Rheinberg, 2008) have highlighted that the need to review the construct validity of the chosen indicators of flow: "Researchers have done little to examine [the] construct validity of the indicators of flow […]" (1994, p. 342) and "[one should] examine the construct validity of the indicators of flow; instead […] data are considered to be ecologically valid" (2008, p. 159).

The definitions of these components were updated in the Rheinberg & Engeser (2018, p. 601) book as follows:

1. "Feeling of optimal challenge: feeling of being in control despite high situational demands (demands and skills are in balance at a high level).

2. The demands of the activity and feedback are perceived as clear and unambiguous; people in flow intuitively know what to do, and how to do it, at any given moment.

3. The pursuit of the action is experienced as smooth. One step flows into the next, as if guided by some inner logic. (This component presumably inspired the term "flow.")

4. There is no need for effortful and volitional concentration; rather, concentration occurs of its own accord, like breathing. Awareness is shielded from all cognitions that do not relate directly to the activity at hand.

5. The sense of time changes: people in flow usually lose all track of time; hours fly by like minutes.

6. People in flow feel a part of what they are doing and become completely absorbed in it ("merging" of action and awareness): loss of self-reflection and self-consciousness". (Rheinberg & Engeser, 2018, p. 601)

Even more recently, based on a "deeper analysis" of existing flow components, differentiating "an accompanying phenomenon, an antecedent, or a consequence of flow", Peifer & Engeser (2021b, p. 422) proposed an integration of the "components of flow in early years" into three components named Absorption, Perceived demand-skill balance, and Enjoyment. Authors suggest that these three resulting "meta-components" or "core experiences" would describe flow while the original components would be kept as 'aspects' of flow to "to more holistically define, describe, and measure flow so as not to miss subtle aspects of the experience of flow as outlined by Csíkszentmihályi" (2021b, p. 424). This proposed integration is shown in Table 1-1:

Table 1-1

A proposal of core flow components (Peifer & Engeser, 2021b, p. 423)

Components of flow in early years	**Proposal of core components**
Merging of action and awareness	
Centering of attention	
Loss of self-consciousness	Absorption
Absorption	
Distortion of temporal experience of time	
Feeling of control	
Experience of coherent, non-contradictory demands	
Experience of clear goals	Perceived demand-skill balance
Experience of unambiguous feedback	
Experience of challenge-/demand skill balance	
Autotelic nature	
Intrinsic motivation	Enjoyment
Enjoyment	

Finally, in a research project involving over 10 European flow researchers from varied universities in different countries, Peifer *et al.* (2022) presented a very much needed, agreed-upon list of characteristics of flow: "(1) challenge-skill-balance, (2) merging of action and awareness, (3) clear goals, (4) unambiguous feedback, (5) concentration on the task, (6) sense of control, (7) loss of self-consciousness, (8) time transformation, and (9) autotelic experience" (2022, p. 1).

While we find the previously mentioned works of Peifer & Engeser (2021b, p. 422) extremely comprehensible, and without doubt their validity shall soon be corroborated by the flow researchers community, we choose to keep instead the list of characteristics of flow admitted in the works by the EFRN and Peifer *et al.* (2014; 2022). Among these components, the challenge-skill balance has been regarded since the beginnings of flow as its most determinant characteristic and thus, the challenge-skill balance is found often at the foundations of many theoretical flow models. In the following Subsection (a) we elucidate the specific relevance of the challenge-skill balance component. Next, Subsection (b) reviews a brief description of the known major obstacles to flow.

The Challenge-Skill Balance

As the reader might have noticed so far, the naming, number, and definition of flow components have remained almost invariable since Csíkszentmihályi's inception on the notion (Engeser et al., 2021). Relatively speaking, "minor" modifications have been brought on a few selected terms, although among those the challenge-skill balance component stands out ("gained much attention in flow research" (Peifer et al., 2022)) mostly because of its importance as a condition to flow.

As recently seen in Section "Evolution of the Definition of Flow and its Conceptual Flow Model" above, the challenge-skill balance characteristic was fundamental in the conceptualization of the theoretical flow model (*cf.* Figure 1-2 & Figure 1-3) since the origins of flow, and its importance was later on confirmed by many other researchers. Piniel & Albert (2020, p. 581) recently recapped on the challenge-skill balance as a determinant "condition" for flow, before proceeding to consider the clear set of goals, and immediate feedback conditions' importance:

> Probably the most important of these [conditions] is that there has to be a balance between perceived skills and perceived challenges. […] Another condition which enhances the likelihood of flow is the presence of a clear set of goals which provide direction and purpose for behavior. […] individuals need to have clear and immediate feedback about their actions, which themselves appear easier to achieve relative to clearly set goals. (Piniel & Albert, 2020, p. 581)

Furthermore, flow researchers make the point that both latter conditions might be implied in the challenge-skill balance condition:

> […] not only are the clear setting of goals and feedback dependent upon each other, but since the balance between skills and challenges cannot be interpreted without clear goals and immediate feedback, these two antecedents can be said to be superfluous. (Keller & Landhäußer, 2012, pp. 52–53; 2020, p. 581)

Equally, Engeser *et al.* (2021, p. 128) name the "balance of challenge and skill" a "key component" of flow theory and incite researchers "[…] not to leave [it] aside […]". Furthermore, Abuhamdeh (2021a, p. 151) describes the link between the challenge-skill balance and enjoyment: "Optimal challenges may promote attentional involvement, and this would promote enjoyment. Optimal challenges may also heighten suspense, which has been linked to enjoyment" (2021a, p. 151).

Thus, the importance of the challenge-skill balance dimension in flow research relies on its dependence on the individual's subjective skills and challenges' view more than it depends on the task itself. An empirical study aiming to examine "relationships between measured flow antecedents, flow experiences and flow consequence-course satisfaction" (Shin, 2006) showed that it is the students' perceptions of their level of 'skill' and 'challenge' -and not the difficulty of the task itself- to be "critical to determining the level of flow" (2006). While discussing flow activities, Csíkszentmihályi assured that "It is the subjective challenges and subjective skills, not objective ones, that influence the quality of a person's experience" (Csíkszentmihályi, 2014, p. 242 originally published in 2002).

Certainly, the Csíkszentmihályi (1988) couple have long suggested that the "ability to experience flow may be due to individual differences that are in part inborn, but it certainly can be learned" and that learning such skills makes "much easier to achieve the necessary balancing of challenges and skills" considered a primary condition for flow emergence, insinuating the importance of such balance and its individual and subjective nature.

In the same note, Nakamura & Csíkszentmihályi (2002, p. 91) add that such individual experience is to be understood as a system accounting for the environment as well, highlighting the importance of the "dynamic system composed of person and environment […]" and of their interactions, which we stress, characterize the individual, unique and subjective nature of the flow experience: "[those] who routinely find deep enjoyment in an activity illustrate how an organized set of challenges and a corresponding set of skills result in optimal experience" (2002, p. 91). They further add that applications of their 1996 flow model kept pointing towards the initial and "essential insight" that "perceived challenges and skill must be relative to a person's own average levels" (2002, p. 95).

Nevertheless, and contrary to the common notion that only a privileged few number of people can attain this state (*e.g.*, elite athletes, artists), a journal article by Burt & Gonzalez (2021) shows that ordinary people are "learning to achieve their maximum potential and live optimal lives", citing the study examples of a manuscript (Ruiz-Martínez et al., 2021) on work-life balance, and a book chapter on psychotherapy and mental health rehabilitation (E. Riva et al., 2016), highlighting the "ability to find flow in everyday life was connected with individuals' well-being and reduced symptomatology" (2016).

Finally, to even further accentuate the importance of the challenge-skill balance dimension when facing other experiential notions, a journal article by Jackson & Eklund (2002) affirmed that "other experiential dimensions such as anxiety, apathy, and boredom are also predicted via the challenge-skill ratio".

Thus, we would like to conclude this Subsection with an important consideration to keep in mind in flow research:

> [any] given individual can find flow in almost any activity – working a cash register, ironing clothes, driving a car, etc. Similarly, under certain conditions and depending on an individual's history with the activity, almost any pursuit – a museum visit, a round of golf, a game of chess – can bore or create anxiety. It is the subjective challenges and subjective skills, not objective ones, that influence the quality of a person's experience. (Csíkszentmihályi, 2014, p. 242 originally published in 2002)

Known Obstacles to Flow

In this Section we quickly look at some of the confirmed obstacles to flow, which can be social or individual (Csíkszentmihályi, 1990a, p. 85).

Leaving aside the terrible social conditions that obviously prevent any experience of enjoyment, such as slavery, oppression, exploitation, etc., we present the ones Csíkszentmihályi (1990a, p. 86) deemed relevant: *anomie* ("lack of rules") & alienation[29].

Csíkszentmihályi considers *anomie* as the condition in society "When it is no longer clear what is permitted and what is not, when it is uncertain what public opinion values, behavior becomes erratic and meaningless" (1990a, p. 86), *e.g.*, an economic crash, exception state, failed state, cultural destruction, a global pandemic, etc. During such social situations it becomes unclear what activities are worth for the individuals to invest themselves in, inhibiting investment in possible flow activities.

[29] *Cf.* R. G. Mitchell (1988) for a more comprehensive literature review on these two concepts and how they relate to flow.

Similarly, Csíkszentmihályi sees alienation as the condition when "people are constrained by the social system to act in ways that go against their goals" (1990a, p. 86), *e.g.*, in this case: attributed gender roles, reaching puberty, forced migration, labor, etc.

Regarding the individual obstacles, according to (1990a, p. 84), some people might be "constitutionally" unable to experience flow. A condition named *anhedonia* ("lack of pleasure"), proper of schizophrenics, is characterized by an inability to concentrate and then, paradoxically they tend to notice all surrounding stimuli, relevant or no.

In the same note, individuals suffering from excessive "self-consciousness" and "self-centeredness", *i.e.*, constantly worrying about how others perceive them, or in fear of committing a *faux-pas*, for the former, and those who see the world in terms of how anything or anyone can be useful to their desires for the latter ("[…] a man or a woman who cannot advance one's interests does not deserve further attention" (1990a, p. 85)), might be devoid of the required attentional fluidity to relate to flow activities: "it is difficult to become interested in intrinsic goals, to lose oneself in an activity that offers no rewards outside the interaction itself" (1990a, p. 85).

When studying the challenge-skill balance, published findings (Engeser & Rheinberg, 2008, p. 165) suggest that the individual's willingness to take on challenges (hope of success *vs.* fear of failure) increases the likelihood of the flow experience: "[…] flow is higher for individuals high in the implicit achievement motive 'hope of success'. The reverse pattern holds true for the explicit achievement motive of 'fear of failure'" (2008, p. 165).

Moreover, the model employed[30] in their experiments points out that, contrary to intuition, these individual "high in the implicit achievement motive" might prefer medium-difficulty tasks (where the challenge-skill ratio is balanced) instead of high-difficulty tasks, whereas individuals with a strong fear of failure "even avoid tasks of medium difficulty" (2008, p. 165). This underlines the importance not only of the challenge-skill balance but also of the

[30] Risk-taking model of Atkinson (1958; cited by Engeser & Rheinberg, 2008).

challenge's medium perceived difficulty as a not-so-recognizable factor when designing learning resources intended for facilitating flow.

We would like the reader to retain from this Subsection what we consider might constitute an all-encompassing definition of a personal obstacle to experience flow, according to Csíkszentmihályi (1990a, p. 87). He related what he considered an association between people who needed more external information to form representations of reality in consciousness and their self-reported daily-life lower intrinsic motivation. He concluded that people needing less extrinsic information (to conceive reality representations) would have an easier time experiencing flow:

> Individuals who require a great deal of outside information to form representations of reality in consciousness may become more dependent on the external environment for using their minds. They would have less control over their thoughts, which in turn would make it more difficult for them to enjoy experience. By contrast, people who need only a few external cues to represent events in consciousness are more autonomous from the environment. They have a more flexible attention that allows them to restructure experience more easily, and therefore to achieve optimal experiences more frequently. (Csíkszentmihályi, 1990a, p. 87)

We present these obstacles to first, confirm that flow's emergence depends on the subjective characteristics of the individual although it can be greatly impeded by social contexts. Second, this confirms us that the subjective and individual challenge-skill balance (*cf.* Section "Evolution of the Definition of Flow and its Conceptual Flow Model" above) is a major flow determinant, accounting the lack of individual obstacles previously cited.

Thus, we are aware that social conditions might play a role in the general ability of the participants of this study to experience flow, but we can somewhat comfort the reader that, during this study's data collection period, no major war abrupted that directly affected a group of participants, neither did any major economic crash. Still, while the global pandemic (and the drastic measures taken to overcome its spread) could be considered a major disruptor of flow, we did not see it as an impending one for the following two reasons:

1. The pandemic was global: Although we are aware of its effects on people are different, and that people also experienced it differently, depending mostly on their government taken steps (Carroll et al., 2020), we posit everybody had some concise awareness of something unusual happening. Thus, the pandemic did not subject any obviously clear group of participants to a flow-acquiring disadvantage compared to any other clear group which would have a clear advantage, unless of course the participant fell ill.

2. MOOCs happen online, and are followed usually at home, *i.e.*, the conditions in which individuals participated in this study would have been similar if no global pandemic existed, which are an educational, online, asynchronous context requiring an internet-connected computer and a few hours of "at-home" work (*cf.* Chapter 3 below).

A global economic crisis ensued shortly after the global pandemic. Nevertheless, this post-pandemic European crisis happened after this study data-collection period was over. If anything, this crisis would have affected the rate of fee-based certifications in MOOCs (*cf.* Chapter 3 below), which this thesis does not address.

In this Subsection we looked at what Csíkszentmihályi considered obstacles to flow, separated into social and personal obstacles. Despite these obstacles, we asserted the notion that flow depends intensively on the subjective individual challenge-skill balance and thus, any given activity will represent a different challenge to distinct individuals according to each person's own skills (or perception of). Finally, we highlighted what social obstacles could have played a negative role in this study and how we saw fit to consider them.

Measurement Attempts of Flow

This Section traces some approaches for measuring flow. We bring out the difficulties acknowledged by researchers when attempting to measure flow. We intend to illustrate this scientific endeavor with evidential proof that, given some concessions, acceptable flow measuring instruments can be attained. We begin by stating that, despite a general academic consensus on flow's definition, there is very much room for improvement when proposing flow measurement approaches.

Engeser & Peifer (2021a) argued that the definition of flow has suffered few modifications since Csíkszentmihályi coined (and refined) its innermost structure (Csíkszentmihályi, 1975a, 1975b, 1990a): "The original definition provided by Csikszentmihalyi (1975) was only marginally modified over the years" (Engeser, 2012; Peifer & Engeser, 2021a, p. 2). Still, while such time-proven definition enjoy a strong agreement among researchers, a certain level of difference persist on how flow should be measured, while pointing out that no agreed-upon measurement standard has been found yet (Moneta, 2021):

> […] there is a certain level of disagreement among researchers as to how flow should be measured. […] researchers have kept developing and validating new measurement tools for flow, and modifying and re-validating established ones, which indicates that a gold measurement standard for flow has yet to be achieved. (Moneta, 2021, p. 1)

Undeniably, this collection of "measurement tools for flow" obeys the crucial reality of flow's subjective and fragile nature. Truly, flow (*cf.* Section "Evolution of the Definition of Flow and its Conceptual Flow Model" above) is a subjective psychological state that depends on the individual's perception on its abilities to tackle the demands of the task and the difficulty of the not-so-easy task; where a balance is to be met (EFRN, 2014). Two essential questions have confused flow researchers' minds:

- How to scientifically measure what would shift from one individual to another when faced with the same, objective situation?

Since its inception in 1975 the difficulty of determining the flow state was ascertained: the nature of the challenges and the skills depended entirely on the individual's perception of these challenges and skills:

> […] [flow] does not depend entirely on the objective nature of the challenges present or on the objective level of skills. In fact, whether one is in flow or not depends entirely on one's perception of what the challenges and skills are. (Csíkszentmihályi, 1975a, 1975b)

When Csíkszentmihályi (1990a, p. 61) evoked a "lacking the sense of worry about losing control" found in other normal life experiences, again, he asserted that the individual's subjective perception on the possibility of exerting control is what seemed to contribute to flow instead of the actual, objective control over the challenging activity. This underlined that the subjective perception of the individual applied not only to the challenge-skill balance (*cf.* Section "The Challenge-Skill Balance" above) but also to the perception of the possibility of control.

The Jackson & Marsh (1996) publication recognized a few difficulties on this matter: the then-existing scarcity on psychology research with flow as a variable "due to the difficulty in measuring the concept". Or, that "Research of flow has lagged behind experiential awareness of the state due to the inherent difficulties of applying empirical methods to phenomenological experiences" (1996, p. 17).

- How to scientifically measure a per-definition, unstable phenomenon?

Again, Csíkszentmihályi (2014, p. 258 originally published in 2002) acknowledged that "interrupting deep flow […] destroys the phenomenon". The Jackson & Eklund (2002) publication conceded that "disrupting performance during the activity is another obstacle to using the ESM approach". Rheinberg & Engeser (2018, p. 602) listed "frequent interruptions" as conditions that inhibit flow.

Consequently, over the last 47 years, flow researchers have shown ingenuity to tackle this seemingly impossible task. Trade-offs and concessions have been made and therefore multiple measurement instruments have been developed for different research fields, each accounting for specific flow dimensions (*cf.* Section "Evolution of Components of Flow" above). We present the main phases in flow measurement and how they were confronted.

As previously mentioned, currently any on-the-hands attempt to detect or measure flow inevitably contributes somewhat to flow disruption. To confront this hurdle, researchers first employed measurement instruments that attempted to elude or circumvent this situation, such as semi-structured interviews (Csíkszentmihályi, 1975a, p. 35; Jackson et al., 2008, p. 562), or the famous Experience Sampling Method (*cf.* Section "Experience Sampling Method (ESM) and the ES Form" below) (Csíkszentmihályi, 2014, pp. 35–54; Larson & Csíkszentmihályi, 2014; Moneta, 2021; Nakamura & Csíkszentmihályi, 2009; Rheinberg et al., 2003).

For example, upon receiving an alert[31], the four questionnaires employed initially by Csíkszentmihályi (1975a) surveyed the participants' instant, real-time perception of vary diverse feelings such as mood, sleep, hunger, headaches, itches, happiness, irritability, reason, "general psychological state", social contact, relaxation, "speed", among many others (Csíkszentmihályi, 1975a, app. Tests and Procedures Used in Microflow Experiments). However, as these questionnaires were to be filled up upon receiving the beeper's alert, they could potentially interrupt a flow activity. Aiming to reduce the possibility of a possible flow interruption, in 1996 Moneta & Csíkszentmihályi (1996, p. 279) listed a classification method for measuring experience according to the timing of measurement (Wheeler & Reis, 1991):

> (a) "interval contingent", where participants are required to respond at regular intervals (as in the end-of-day diaries), (b) "signal contingent", where participants respond when signaled (for example, by a pager), and (c) "event contingent", where participants respond when a specified event occurs. (Moneta & Csíkszentmihályi, 1996, p. 279)

Additionally to their associated potentially-in-flow interrupting behavior, techniques such as the ESM are costly, as they require not only the purchasing and preventive/corrective maintenance of the alerting devices but they also involve a minimal training on the operation of such device, plus the logistics of distributing and gathering them at the beginning/end of the

[31] The Experience Sampling Method relies on the participant answering a form at random moments of the day prompted by a "beeper".

experiment: Rheinberg & Engeser (2018, p. 603) admitted this approach as a "time and cost-intensive technique" but also extremely advantageous on the validity of data collection.

At a second moment, researchers employed self-reported measure instruments, *i.e.*, post-event questionnaires, which present the advantage that they do not disrupt flow and can be applied to many individuals, in a variety of contexts (online/offline or distant/presential settings) at a minimal cost. This approach usually relies on a componential construct of flow to conceive standardized scales.

In agreement with this approach, academics have developed, validated, modified, and re-validated unidimensional and multidimensional measure instruments for flow, *cf.* the very complete literature reviews contained in several publications (Peifer et al., 2022; Moneta, 2021, p. 20; Hoffman & Novak, 2009; Moneta, 2012a). Conforming to this position, Jackson & Marsh (1996) insisted on the hypothetical construct nature of flow and that "its usefulness must be established by investigations of construct validity" (1996, p. 21).

For instance, Csíkszentmihályi (1975a) first employed a combination of observations, casual discussions, interviews, adaptations of standardized tests, and questionnaires (which at the moment, did not constitute a standardized flow-detecting questionnaire).

Hoffman & Novak (2009) surveyed and analyzed more than 20 different conceptual and structural models of flow employed in diverse contexts[32] such as creative or performing arts, work (Bakker, 2008; Bakker & Demerouti, 2008)[33], music (Bloom & Skutnick-Henley, 2005), ecommerce (Hoffman & Novak, 2009; Rheinberg et al., 2003), sports (Jackson & Eklund, 2002; Rufi et al., 2014), eLearning (Heutte & Fenouillet, 2010; Skadberg & Kimmel, 2004), and/or video gaming (D.-S. Choi et al., 2000; Fu et al., 2009).

---

[32] *Cf.* (de Moura Jr & Porto Bellini, 2019; Pels et al., 2018; Tan & Sin, 2021) for field-specific literature reviews of flow measurement in music, work, and social flow, and/or (Rosas et al., 2023) for a compiled list of 34 validated flow measurement instruments in English (Non-ESM scales).

[33] Also surveyed by Nakamura & Csíkszentmihályi (2009, p. 198)

Off the record, Lonczak (2019) additionally sampled 22 flow and flow-related assessment tools in different validation stages, distributed among varied domains, such as work (Bakker & Leiter, 2010; Schaufeli et al., 2006b, 2006a; Schaufeli & Bakker, 2010; Seppälä et al., 2009) human-computer interaction or technological environments (Guo & Poole, 2009; Redaelli & Riva, 2011), games (D. Choi & Kim, 2004; Kiili & Lainema, 2008), as well as more global measures (Magyaródi et al., 2013; Martin & Jackson, 2008; Oláh, 2005; Payne et al., 2011).

Lastly, several studies have attempted to measure flow via proxies (de Moura Jr & Porto Bellini, 2019) to prevent flow interruptions, more specifically via physiological indicators as previously suggested by Rheinberg & Engeser (2018, p. 609): "Such an interruption could be avoided if flow could be measured with physiological indicators (resp. correlates) during the activity". A comprehensive literature review on the psychophysiology of flow is addressed in two of Peifer works (2022, p. 8; 2021) but here we present a few examples:

Using a large twin people sample, a study (Butkovic et al., 2015, p. 137) delved into the relationship between music practice and the proneness to experience psychological flow (among other personality variables) using genetic analyses, *i.e.*, hereditary traits . Their findings showed "that openness and music flow are important predictors of music practice, and their associations are largely due to shared genes", although they admitted that their results account for about "a quarter of the variance in music practice".

Following this approach, Cheron (2016) attempted to "catch the flow" with Electroencephalogram (EEG) dynamics and Electromyographic signals (EMG), specifically by quantifying the "the (1) power and (2) the phase of the different frequency EEG oscillations ranging from delta, theta, beta and gamma bands occurring before, during and after the flow", for the EEG prediction (2016, p. 3).

Another approach gathering considerable momentum makes use of cardiac features (heart rate, variability, etc. obtained via wearables) for classifying flow in real time using Machine Learning techniques (Hussain et al., 2012; Rissler et al., 2018, 2020). Such results show the potential for detailed flow classification and automatic assessment of flow in Information Technologies.

Finally, as a last example, Peifer (2012, p. 153) explained that the closeness between the notions of flow and stress would suggest likewise a connection between the cortisol hormone to flow. Yet, findings in this area seem to be contradictory with different studies showing "positive association, no association and a negative effect […] on flow" (2022, p. 8).

As the reader can imagine, such measurement methods remain too complex to replicate in other domains, and even within their own specific research fields, although they testify to the ingenuity and dedication of the flow community.

We have seen how researchers have tried to tackle the subjectivity and the instability natures of the flow concept to measure it: they have employed observations, interviews, beepers, questionnaires (standardized scales of the componential approach), and finally complex proxy-based protocols.

Yet, concessions have been granted when designing such measurement protocols and therefore, researchers agree not a single measurement instrument fully captures flow. So, although there are "popular" measurement methods currently and validly employed by the academic community, no single agreed-upon measurement standard method currently stands, applicable to all domains, all circumstances, accounting for all flow components (Moneta, 2021).

We conclude this Section with a citation appeared recently in the latest volume of "Advances in flow research" (Peifer & Engeser, 2021a). Moneta (2021, pp. 31–32) precisely captures the existing panorama in flow measurement, recapitulating on the consented and accepted trade-offs currently present in all flow methods:

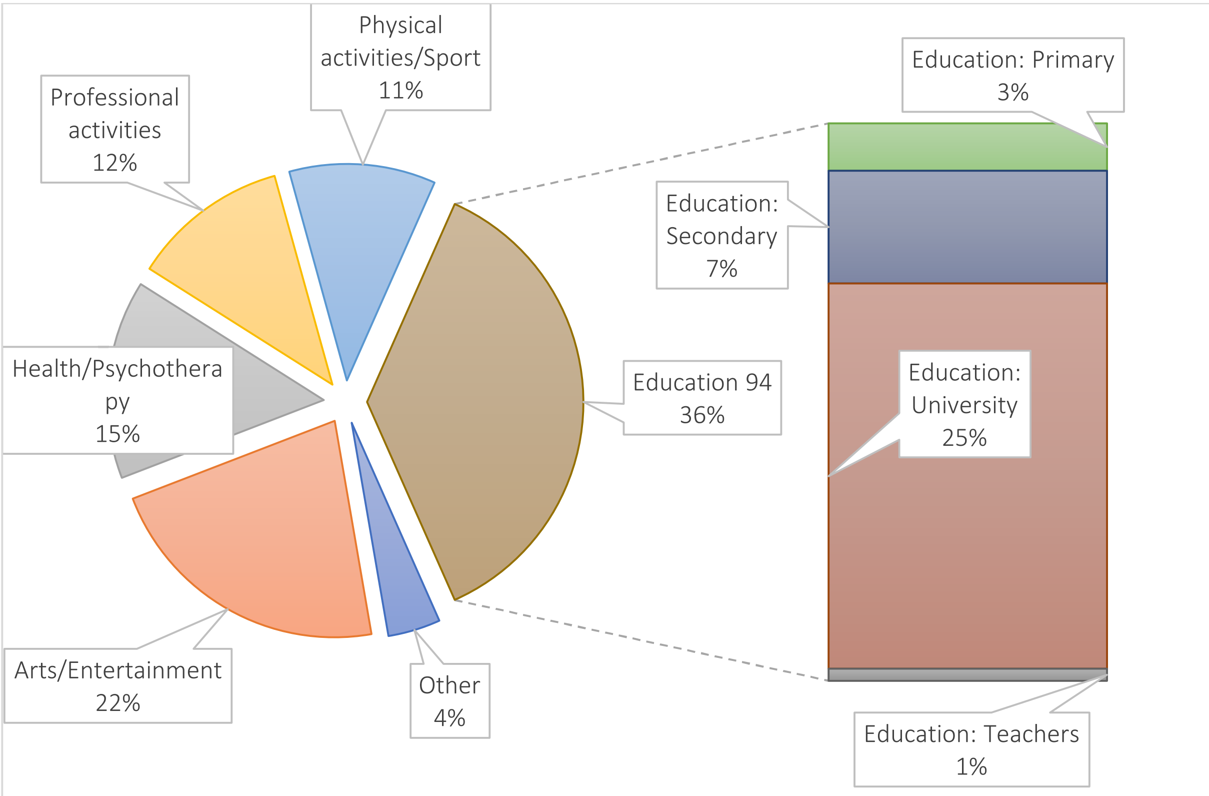
> The key message of this chapter is that no existing measurement method for flow and associated model is watertight, and that a gold standard for the modeling and measurement of flow is not at close reach (2021, pp. 31–32).

Flow in Educational Contexts

Research of flow in education is on the rise. Hinted at in a very recent survey (Peifer et al., 2022) but rather detailed in one of his most recent publications (Heutte, 2021, p. 11), Heutte explains that, out of the 256 reviewed articles, 94 are specifically dedicated to education and or training. That is an astounding number, accounting for 36% of all reviewed articles, putting education as the most researched domain for flow[34], followed by creative endeavors and enjoyment.

Figure 1-9

Distribution by Field of Empirical Studies in Peifer's Scoping Review (Heutte, 2021, p. 11; Peifer et al., 2022)



[34] This contrasts the contents of the Peifer & Engeser (2021a) publication, where the subject is not addressed by any of the contributors.

More precisely speaking, out of these last 94 articles, 69.4% concern studies focused on university contexts, 20.4% on secondary education, 8.2 primary education, and 2% are related to the teachers themselves[35]. This distribution is better appreciated on Figure 1-9[36], where we can clearly appreciate that research of flow in university contexts is the leading subject.

Therefore, this Section gives an overview of flow research and applications in educational contexts. We set off by stating the importance of the psychological state in the learning process. Then, we justify the involvement of flow, a human psychological state in learning, and briefly depict some of its existing applications notably when attempting to measure it in educational contexts. We include a collection of flow measurement instruments historically employed in educational contexts.

Incontestably, the learner's psychological state has a relevant impact in the learning process. Researchers' own studies and literature reviews treating this sentiment (Abyaa et al., 2019; Efklides & Volet, 2005; Medina-Medina & García-Cabrera, 2016) have shown that the learner's psychological state carries a preponderant weight in the learning process. This clear facet of the learning process has been explored in online and distant educational settings as well (Heutte, Kaplan, et al., 2014).

For instance, a comprehensive literature review (Abyaa et al., 2019, p. 1106) on Learner Models[37] stated that, during the construction of an ideal Learner Model "[…] one should identify and select the learner's characteristics that influence his/her learning, then take into consideration the learner's psychological states during his/ her learning […]" (2019, p. 1106). Furthermore, they add that both positive and negative learner's affective characteristics "have a major impact

[35] These percentages are relative to those in education (94) and not to the total (256).
[36] Other 4% might include mixed domain articles, *i.e.*, on something else **and** education, rounding error < 2%
[37] An extended definition on "Learner Model" is proposed in a litterature review by this thesis author (Ramírez Luelmo et al., 2020a, sec. 2.1): Learner Models represent the system's beliefs about the learner's specific characteristics, relevant to the educational practice (Giannandrea & Sansoni, 2013), they are usually enriched by data collection techniques (Nguyen & Do, 2008) and they aim to encode individual learners using a specific set of dimensions (Nakić et al., 2015).

on the learning process, as they can either be the source of the learner's success or failure" (2019, p. 1106).

Also, the findings of Efklides et al. (2005, pp. 426–429; 2005) suggested that positive and negative affects matter in the learning process. More specifically, they found that "mood treatment did have [a direct effect] on learning-related emotions that affect the regulation of learning" (2005, p. 427). This is important to underline because, even if people in a positive mood can "interpret their experiences through the filter of the emotions they happen to be feeling at the time […]" (Hirt et al., 1996, p. 245) it has been shown that "induction of positive affective states […] increase decision-making efficiency and facilitate creative problem-solving (1996, p. 245). Indeed, positive mood induction promotes "prospective interest, willingness to invest effort, and expectation of success in problem solving" (Efklides & Petkaki, 2005, p. 427).

Further on, the Abyaa et al. (2019, p. 1116) literature review confirmed that engagement in learners is composed of psychological aspects such as behavior, emotions, plus cognitive aspects such as effort and strategy "to guide their pedagogical approach for enhanced engagement, motivation and consequently, learning" (Papadopoulos et al., 2013, sec. II).

Not to mention the corroboration by Medina-Medina & García-Cabrera (2016, p. 2), when defining a taxonomy for User Models in adaptive systems, they list the "psychological state" as a subset of data the system must manage. To boot, they state that such model "Must take into account the students psychological aspects: frustration, motivation, satisfaction, disappointment, etc." (2016, p. 11).

Again, Abyaa et al. (2019) revealed that current (2014-2019) renowned research revolves around psychological states such as shame, reproach, distress, joy, pride, and/or admiration[38]. Nevertheless, our research work focuses on the flow state because of its added value promoting learning, as we will shortly show.

[38] Some authors consider motivation and engagement as separate psychological states as well (Abyaa et al., 2019).

Concerning flow and learning, we begin by referring to the works of the Csíkszentmihályi (1988) couple who confirm that at least since Csíkszentmihályi's (1975a) mid-70's publication, work and education were endorsed as the most urgent applications of flow:

> […] the most urgent applications of the flow model were in schools and on the job, where most people spend most of their lives - often in boredom or in states of uneasy anxiety. Therefore educational and occupational uses of the model seem to be the most urgent ones. (Csíkszentmihályi & Csíkszentmihályi, 1988)

Indeed, Csíkszentmihályi & LeFèvre (1989, p. 816) grasped and stated the relationship between flow and learning: "When both challenges and skill are high, the person is not only enjoying the moment, but is also stretching his or her capabilities with the likelihood of learning new skills and of increasing self-esteem and personal complexity" (1989, p. 816).

Shernoff & Csíkszentmihályi (2009, p. 132) further confirmed that "The theory of flow is inherently related to learning" and proceed to describe how learning ultimately implies reaching a challenge-skill balance in a succession of proctored tasks, or in the words of Piniel & Albert (2020, p. 582): "[…] learning […] ultimately involves acquiring new skills by completing tasks that in some way exceed the person's current abilities, that prompt the person's focus, and provide opportunities to meet challenges" (2020, p. 582).

For example, when commenting on the subject of literacy, Csíkszentmihályi (1990b, p. 124) argued that once literacy was firmly established in everyday life, it gave way to more enjoyable activities than "recording pigs and bales of hay forever " such as [written] storytelling (*e.g.*, to keep memory of old distant memories), in the form of literature, an intrinsically motivated activity:

> With time, literature emerged out of literacy, and for many writers and readers it became an end in itself. Writing was no longer motivated only extrinsically, by economic and by political need; it was now possible to enjoy it for its own sake. (Csíkszentmihályi, 1990b, p. 124)

Accordingly, Csíkszentmihályi (1990b, p. 126) insisted that, for the case of literacy, a model of intrinsic motivation to learning was needed to turn the "boring" experiences of reading, writing, or doing sums, into enjoyable intrinsically rewarding activities: "Applying a model of intrinsic motivation to learning may make it possible to advance the cause of literacy beyond the point where technology and a mechanical rationality cease to be useful".

Why a model of intrinsic motivation? Because learning requires investing in a cognitive effort of information processing which reward must not depend solely on external factors to be sustainable. Csíkszentmihályi (1990b, p. 135) reasoned that "Learning involves processing information. Complex information processing requires the allocation of attention to the task. There cannot be any learning unless a person is willing to invest in a symbolic system". Thus, motivation for learners to invest in such a symbolic system is what is needed, and he proposes two main ways (1990b, p. 138), which we re-phrase in a generalized form:

1. Realistic and honest assessment of the extrinsic rewards regarding the learning in question, *i.e.*, real awareness on the expected advantages and disadvantages of acquiring such learning: "[…] involve a much clearer communication of the advantages and disadvantages one might expect as a result of being able to […]" (1990b, p. 138).
2. Render the learning activities more enjoyable, without trivializing them: "[…] make children aware of how much fun […] can be" (1990b, p. 138), which will empower learners in a more efficient and permanent way to re-use the acquired knowledge.

More closely related to flow, in his 2005 book, Csíkszentmihályi et al. (2005) argued that the flow state promotes learning and personal development because deep and total concentration experiences linked to flow are intrinsically rewarding, and in the specific case of learning, those experiences motivate students to repeat any given activity at progressively higher challenging levels.

On that account, Nakamura & Csíkszentmihályi (2009, pp. 199–200) censused multiple studies illustrating the varied benefits of flow in diverse educational contexts, such as extended school commitment, increased motivation, less anxiety, fostering of skills, improved performance, self-esteem and performance prediction, cognitive engagement, across "all school types, pedagogies, and instructional practices." We present two of these illustrations:

Nakamura & Csíkszentmihályi (2009, p. 199) mentioned that during a longitudinal study on talented high-school students (Csíkszentmihályi, 1993), a relationship between quality of experience and persistence emerged:

> Students still committed to had experienced more flow and less anxiety than their peers when engaged in school-related activities; they also were more likely to have identified their talent area as a source of flow. (Csíkszentmihályi, 1993)

Based on the results of a study (Engeser et al., 2005) on two university courses, Nakamura & Csíkszentmihályi (2009, p. 199) pointed out that "flow predicted semester-end performance" despite substantial differences in the study groups (presential/at home, mandatory/optional, and the reviewed subjects). Here follows the original conclusion by the authors:

> *In beide Untersuchungen wurde diese Annahme bestätigt. Auch bei Kontrolle leistungsrelevanter Kompetenzfaktoren sagte Flow-Erleben während der Lernphase spätere Lernleistung vorher. […] Bemerkenswert an dem replizierten Befund ist, dass die Untersuchungen zwar beide im universitären Kontext durchgeführt wurden, aber doch recht unterschiedliche Settings betrafen. [...] Trotz dieser Unterschiedlichkeit zeigten sich in beiden Untersuchungen die erwarteten Beziehungen zwischen Motivation, Flow und Lernleistung*[39]*. (Engeser et al., 2005, p. 17)*

[39] [In both studies this assumption was confirmed. Even when controlling for performance-relevant competence factors, flow experience during the learning phase predicted later learning performance. […] What is remarkable about the replicated finding is that although the studies were both conducted in university contexts, they involved quite different settings. […] Despite this difference, both studies showed the expected relationships between motivation, flow, and learning performance].

Absolutely, many studies (El Mawas & Heutte, 2019, p. 502; Heutte, 2019, pp. 185–188; Heutte et al., 2021, pp. 2–3; Peifer et al., 2022, sec. Motivation) have shown flow to reliably and positively correlate to the ensuing learning-favorable metrics:

- curiosity (Malone, 1981),
- creativity (Culbertson et al., 2015),
- engagement (Mesurado et al., 2016),
- intrinsic motivation (Keller et al., 2011; Rheinberg, 2020; Rheinberg & Engeser, 2011, 2018),
- self-efficacy (Bandura et al., 1997; Bassi et al., 2007; Heutte, Fenouillet, Kaplan, et al., 2016; Mesurado et al., 2016; Rodríguez-Sánchez et al., 2011; Salanova et al., 2006),
- self-determination (Heutte, 2019, pp. 194–195; Schattke, 2011),
- self-regulation (L.-X. Chen & Sun, 2016; Heutte, 2019, p. 197; Rodríguez-Sánchez et al., 2011),
- interest in learning (Bachen et al., 2016),
- achievement motives (Engeser & Rheinberg, 2008),
- goal orientation (Oertig et al., 2014, p. 178; Rheinberg et al., 2003),
- goal attainment (Rheinberg & Engeser, 2011),
- and overall learning and academic performance (Csíkszentmihályi, 1993; Engeser et al., 2005; Engeser & Rheinberg, 2008, pp. 160–161).

Conveniently, this phenomenon is extensive to online, distant settings, as exemplified in the Skadberg & Kimmel (2004, p. 415) journal article: “As expected, flow experience had a positive impact on people’s learning” and “flow experience is positively related to increased learning about the content of a Web site", being the most important factor positively affecting attitude and behavior.

How is this possible? According to the findings of Csíkszentmihályi (1990a, p. 71), flow is more likely to occur within the framework of "a structured activity, or from an individual's ability to make flow occur, or both." One obvious indication would point to the structured nature of an online learning activity, such as those found in a MOOC, but telepresence[40], *a.k.a.* 'remote presence' is a more appropriate contender. Indeed, as Heutte (2021, p. 13) points out, numerous studies identified telepresence, or more precisely 'presence in eLearning' (Jézégou, 2012, sec. 2), as an element susceptible of contributing to the autotelic experience (J. Chen, 2006; Hoffman & Novak, 1996, 2009; Novak et al., 2000; Skadberg & Kimmel, 2004) because of its close proximity with flow elements such as immersion, control, or the loss of self-consciousness (Heutte, 2021, p. 13). Some publications (Redaelli & Riva, 2011; G. Riva et al., 2011) go even further and include flow as a relevant variable of their presence model characterized by mediated flow:

> Mediated flow corresponds to the extent to which (a) the user perceives a sense of control over the computer interaction, (b) the user perceives that his or her attention is focused on the interaction, (c) the user's curiosity is aroused during the interaction, and (d) the user finds the interaction intrinsically interesting. (G. Riva et al., 2011, p. 9; adapted from Trevino & Webster, 1992)

Nevertheless, we remark that many studies (L.-X. Chen & Sun, 2016; Heutte, Fenouillet, Kaplan, et al., 2016; Salanova et al., 2006; Csíkszentmihályi et al., 2005; Salanova et al., 2014) showed that the relation between flow state and learning[41] is a complex one because "the learning process is not simple" (El Mawas & Heutte, 2019, p. 497). For instance, while flow may predict motivational outcomes (such as intrinsic motivation, interest, self-efficacy, self-regulation, persistence, etc.), it does not always predict task performance (El Mawas & Heutte, 2019, p. 497).

---

[40] Presence and Telepresence are "the sense of being in an environment, generated by natural or mediated means, respectively" (Steuer, 1992, p. 3).

[41] Some studies also highlight the importance of collective (or social) flow in this process.

As an example, a study (Durik & Matarazzo, 2009, pp. 158–159) showed that the perceived complexity (*a.k.a.* the flow challenge) is not a significant predictor of performance.

Here we make a pause to promptly inform the reader that this thesis research does not focus on the relationship between flow and (academic) performance (Harris et al., 2021) but rather on the practical detection of flow in learning, online, distant contexts. Moving on.

Also, we take the opportunity to shift focus from the flow correlates relevant in learning to the educational context itself and flow. On that account we put forward the argument of Heutte (2021) establishing a tighter relationship between flow, as an autotelic, gratifying experience, and learning. According to Heutte (2021, p. 11), the autotelic experience allows anybody to feel gratification during an activity in which the individual acknowledges his/her progression beyond what s/he imagined before engaging into it. Thus, Heutte (2019, p. 179) regards flow as the emotion linked to the psychological state characterized by a feeling of mental fluidity and intense concentration on the tasks that mobilize all of the actor's skills.

Accordingly, within an educational context, Heutte (2017a, p. 10, 2017c, p. 206, 2020, sec. 3.2.2) time and again asserts that flow is to be understood via the emotion linked to the fact of realizing that one is progressing, that one understands, that one is understood:

> *Le flow est donc souvent appréhendé au travers de l'émotion liée au fait de s'apercevoir que l'on progresse, que l'on comprend, que l'on est compris. (Heutte, 2017a, sec. Conclusion)*

Because of the intrinsic motivation factor secured through the characterized autotelic experience (Keller & Bless, 2008), researchers (El Mawas & Heutte, 2019, sec. 2.1; Schattke, 2011) also see the flow experience as a state of optimal motivation (Deci & Ryan, 2002; Heutte, 2017c).

Moreover, Heutte (2021, p. 5) winds back and insists on reconsidering the distinction between flow conditions and flow state, alluding to elements[42] from the model presented in the

[42] Rudely oversimplified here, as the Flow experience encompasses proximal conditions and the Flow state. Indeed, Fulfillment of Proximal conditions lead to the Flow state.

Kawabata & Mallett (2011, figs. 2–3) journal article and in doing so, to admit a Cognitive Control[43] dimension as a necessary precondition for flow in education. This precondition, once fulfilled, would then be conducive to the already-known Immersion and Time Transformation, Loss of self-consciousness, and the Autotelic experience dimensions (Heutte, 2017b, p. 88, 2019, p. 181, 2021, p. 18; Heutte et al., 2021, p. 5; Heutte, Fenouillet, Kaplan, et al., 2016; Heutte & Fenouillet, 2010, p. 4).

This Cognitive Control dimension would play the leading role in what Heutte (2014, p. 167, 2019, fig. 3.10) defined as 'cognitive absorption': a state of deep commitment focused on the will to understand with, as without, the use of digital technologies (TEL):

> *[…] état de profond engagement focalisé sur la volonté de comprendre avec, comme sans, l'usage des technologies numériques. (Heutte, 2014, p. 167, 2019, fig. 3.10)*

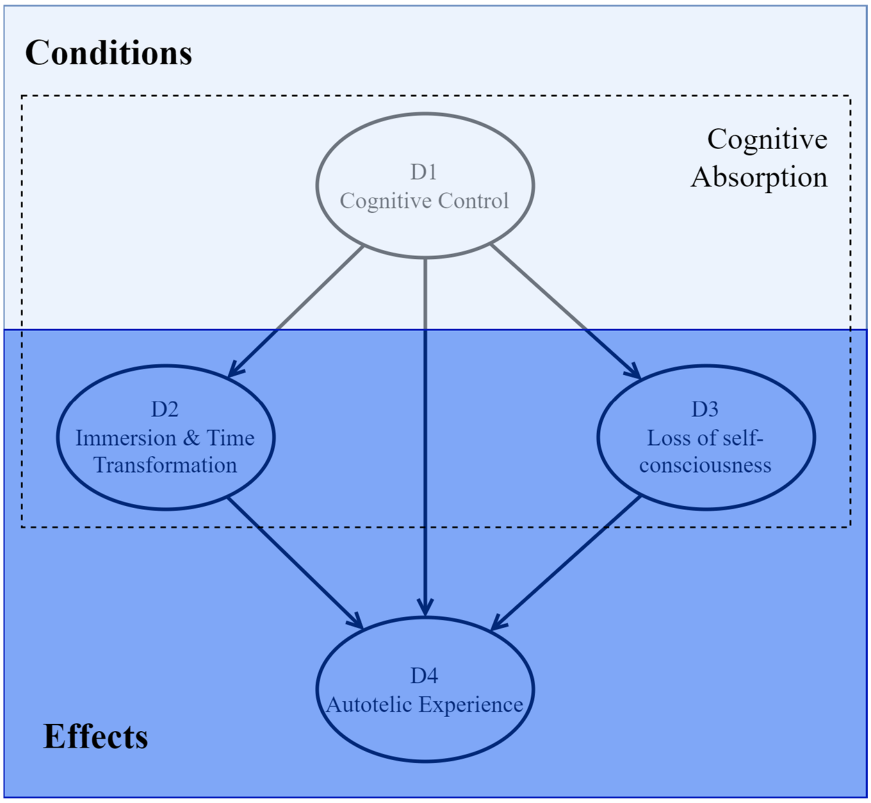
These relationships are better illustrated in Figure 1-10 where the preconditional dimension Cognitive Control can trigger any of the other dimensions (D2-Immersion & Time Transformation, D3-Loss of self-consciousness and D4-Autotelic experience), and where D4 can also be triggered by the then-preconditional dimensions D2 and D3. Cognitive absorption is the ensemble conformed by Cognitive Control (D1), Immersion and Time Transformation (D2), and Loss of self-consciousness (D4).

The crucial-to-flow preconditional Cognitive Control D1 dimension (*cf.* Subsection "Flow in Education (EduFlow & EduFlow-2" below) is described as "a strong feeling of control, specifically over one's actions, characterized by a feeling of ability to deal with the situation and a feeling that the student knows how to deal with whatever comes next" (El Mawas & Heutte, 2019, p. 498).

[43] Linked to Social Cognitive Theory (Bandura, 2001).

Figure 1-10

Relationships Between Dimensions of the Model of Flow in Education (Heutte, Fenouillet, Kaplan, et al., 2016)



Note: This is an simplification (numerical values removed) of the *« Modèle des relations entre les dimensions du flow en éducation »* (Heutte, 2017b, p. 88, 2019, p. 181, 2021, p. 18; Heutte et al., 2021, p. 5; Heutte, Fenouillet, Kaplan, et al., 2016).

Thus, following the previously described train of thought, in an educational context, flow (Heutte, 2019, 2019, p. 179; Heutte et al., 2021, p. 5) is perceived in terms of control over the steps of the task, focused attention (playing a leading role triggering the entire experience), time alteration, immersion in the task, loss of worry for the ego, and it is characterized by an autotelic experience (which would be the motor sustaining the whole experience going) and in a minor measure, in terms of the challenge-skill balance. Under this vision, the learning and understanding are not interrupted neither by any worry related to the next step to take (in the task), nor by what others might think (Heutte, 2019, p. 179).

We have seen (*cf.* Section "Measurement Attempts of Flow" above) how researchers have attempted to measure flow using multiple approaches. Hoffman & Novak (2009) surveyed and analyzed over 20 different conceptual and structural models of flow that span all research fields. In the particular online, distant, educational context that concerns us, recent publications (El Mawas & Heutte, 2019, p. 497; Heutte, 2019, p. 180) have recognized a number of measure instruments known[44] to study flow, such the Experience Sampling Method (Csíkszentmihályi & Larson, 2014 originally published in 1987; Larson & Csíkszentmihályi, 2014 originally published in 1983), Flow Scale (Mayers, 1978), the FlowQuestionnaire (Csíkszentmihályi & Csíkszentmihályi, 1988), the Flow State Scale (FSS) (Jackson & Marsh, 1996), the Flow State Scale 2 (FSS2) (Jackson & Eklund, 2002), the Dispositional Flow Scale 2 (DFS2) (Jackson et al., 2008), the Flow-Kurzskala (FKS) (Rheinberg et al., 2003), Flow in Human-Computer Interaction (Ghani & Deshpande, 1994), Study-Related Flow Inventory (WOLF-S) (Bakker et al., 2017), and the EduFlow and EduFlow-2 scales (Heutte et al., 2021; Heutte, Fenouillet, et al., 2014; Heutte, Fenouillet, Kaplan, et al., 2016; Heutte, Fenouillet, Martin-Krumm, et al., 2016).

Here follows a summary of these flow measurement instruments, starting from the most general to the most educational focused. We precede this summary with the following few notes to keep in mind:

- It is noteworthy to take a detour here and explain that flow measurement instrument design and validation are extensive tasks requiring years of research in psychometrics, terrain experimentation, data collection and processing, *e.g.*, the Mayers (1978) unpublished doctoral thesis found in Subsection "Flow Scale" below. When looking at the steps involved in its cross-cultural validation and internal consistency[45], the demands

[44] Please notice that most of them were not originally designed with that specific context in mind but were nonetheless employed for that purpose.
[45] Cronbach's α coefficient.

of translating a psychometric test or adapting it to a distinct domain or public, are akin to those of its original design (Arafat et al., 2016; Delle Fave et al., 2011), *e.g.*, the *Cuestionario de Experiencia Óptima* (CEO) (Mesurado, 2008b, 2008a) validated a combination of the ESM form (*cf.* Section "Experience Sampling Method (ESM) and the ES Form" below) and the FlowQuestionnaire (*cf.* Subsection "FlowQuestionnaire" below). Delle Fave et al. (2011) deemed the CEO the "Spanish[46] adaptation" (2011, sec. 4.3.5) of both the ESM and the FlowQ but for children and early teenagers only.

- The reader will forgive the content imbalance in the following subsections, but (a) some measurement instruments have historically received more attention than others and more observations and comments are thus available, (b) we present what we consider operational, relevant information to the reader, purposefully avoiding the unaccountable number of study cases, and the very interesting internal consistency and validation metrics publications, and (c) we are limited in the academic content that is made available to us.
- As previously covered in Section "Evolution of the Definition of Flow and its Conceptual Flow Model" above, besides the general flow model, distinct models arose from approaching flow through the scope of different domains, *e.g.*, sports, or work. Therefore, some of the measurement instruments presented here also comprehend (and are thus based on) a conceptual flow model, which we do not detail here. Furthermore, complete measurement protocols exist for some of the measurement instruments presented here but, for the sake of readability, we limit ourselves to presenting the instrument itself, *i.e.*, the scale.
- We do not delve into a debate on the pertinence of other researchers (often, the authors themselves) employing these instruments in the specific context that brings us here, nor

[46] In Spain.

the specific theories (*e.g.*, engagement, motivation, immersion, etc.) authors employed to support their work, nor we discuss these instruments' choices on dimensionality construct nor about their validation, as these matters are considered settled by their own authors, but more importantly, subsequently corroborated by the community employing them.

- And finally, these measure instruments are but a part of the currently existing flow measurement instruments for studying flow in various domains or combinations thereof (*cf.* Section "Measurement Attempts of Flow" above), such as work, play, sports, leisure, creative arts, games, etc. which we do not survey, being out of the educational scope of this thesis. If the reader is interested on a broader spectrum of measurement instruments, a systematic review on validated questionnaires in flow theory has been published very recently (Rosas et al., 2023).

Experience Sampling Method (ESM) and the ES Form

We begin with the ESM (and its accompanying ES Form), the most literature-recurrent flow measurement instrument employed by flow researchers: almost every relevant paper cited in this thesis employed it operationally, or as a frame of reference to validate additional instruments, theories, or models of flow (Csíkszentmihályi & Csíkszentmihályi, 1988, 1988; Csíkszentmihályi & Larson, 2014; S. Han, 1988; Larson & Csíkszentmihályi, 2014; Moneta, 2021; Moneta & Csíkszentmihályi, 1996; Nakamura & Csíkszentmihályi, 2009; Rheinberg et al., 2003).

The ESM is primordially the first data collection method adapted to study how people live quotidianly. Not only it had been employed to study flow, as developed by Csíkszentmihályi (Csíkszentmihályi & Larson, 1984, 1984, 2014 originally published in 1987; Larson & Csíkszentmihályi, 2014 originally published in 1983) based on the results from their first interviews (Csíkszentmihályi, 1975a, 1975b), but also it covers a wide array of topics in humanities and social sciences research, including the phenomenology of everyday life, gender differences, family relationships, professional experiences, intercultural differences and similarities, school experiences, and mental health. Nowadays, the ESM has positioned itself as the-facto reference when attempting to measure flow. It is available (and validated) in Korean, English, Italian, Spanish, Portuguese, and French languages.

The ESM consists in:

1. asking individuals to provide systematic self-reports at random occasions determined by
2. carrying an electronic pager[47], which signals them when to complete a
3. self-report (using the ES form) during most hours of a normal week.

The original ES questionnaire (Csíkszentmihályi & Larson, 2014 originally published in 1987; Larson & Csíkszentmihályi, 2014 originally published in 1983) contained both categorical and scaled items, where the former aimed to "reconstruct the activity and its social context" (Moneta & Csíkszentmihályi, 1996, p. 283), whereas the latter were "designed to measure the intensity of associated subjective feelings" (1996, p. 283). Sets of these self-reports from a sample of individuals create an archival file of daily experience (Larson & Csíkszentmihályi, 2014, p. 40 originally published in 1987).

According to Jackson & Eklund (2002), this measurement instrument fashion relies in the challenge-skill balance dimension, by affirming that "[…] measuring challenges and skills forms the core of the [ESM]".

However, researchers have vocalized their concerns regarding the disrupting nature of the ESM: Jackson & Eklund (2002) recognized that "disrupting performance during the activity is another obstacle to using the ESM approach". Likewise, Engeser & Schiepe-Tiska (2012, p. 15) acknowledged that "the beep signal ruins this state of consciousness". Csíkszentmihályi (2014, p. 258) also admitted this situation: "interrupting deep flow, as the ESM would do […]".

[47] Originally called "Signaling device" (Csíkszentmihályi & Larson, 2014, p. 37 originally published in 1987).

In that regard, Engeser & Schiepe-Tiska (2012) justify the flow interruption through the validation of the flow measure:

> This is clearly the case, but it does not mean that the measure of flow will not be valid. […] It may be compared to a dream; when we wake up, we stop dreaming, but we can still remember the contents and feelings of the dream and could report them (2012, p. 15).

The ES form can be found in Appendix 1. – The Experience Sampling Form.

FlowQuestionnaire

The FlowQuestionnaire (*a.k.a.* Flow-Q or FlowQ or FQ) was developed by the Csíkszentmihályi couple (Csíkszentmihályi, 2014, pp. 216–217; Csíkszentmihályi & Csíkszentmihályi, 1988) when researching life satisfaction in Korean immigrants in the Chicago area (S. Han, 1988). Moneta (2021, p. 32) deems FlowQ the "first measurement method for flow" citing primarily the contribution of the Csíkszentmihályi couple (Csíkszentmihályi & Csíkszentmihályi, 1988). We concur with this assertion as long as one regards the ESM primarily as a data collection method (the initial purpose for which FlowQ was created) based upon interviews (Csíkszentmihályi, 1975a, 1975b).

FlowQ is the revised version of its identically-named predecessor (Asakawa, 2010; Csíkszentmihályi, 1975a; Delle Fave & Massimini, 2003; Jackman et al., 2017; J. A. Johnson et al., 2014). In its current form it comprises only three items, what makes it the shortest measurement instrument listed here, and it is scored in a simple binary (yes/no) scale determining flow or not-flow: "[…] the Flow Questionnaire distinguishes between flow or not-flow" (Peifer & Engeser, 2021b, p. 498).

The FlowQ proposes "definitions of flow and asks respondents to recognize them" but originally it also asked them to "describe the situations and activities in which they experience flow, and rate their subjective experience" (Moneta, 2021, p. 32).

As we will shortly learn, the FlowQ is recognized as a "broad use", effective flow measure/detection instrument by the flow researchers community (Bassi et al., 2014; Bassi & Fave, 2012; Boffi, 2012; Moneta, 2012b; Redaelli & Riva, 2011; Rufi et al., 2014; D. C. Tse et al., 2022), commonly used in tandem with other flow measure instruments, for completeness reasons.

According to the Bassi et al. (2014, p. 832) journal article, in the FlowQuestionnaire, flow is "conceptualized as an all-or-nothing phenomenon, qualitatively distinct from other experiential profiles." Moneta (2012b, p. 494) rises the two main strengths of the FlowQuestionnaire:

> First, it provides a single definition of flow that captures the simultaneous presence of the three key components of flow […]: a) loss of self-consciousness, (b) centering of attention, and (c) merging of action and awareness.
>
> […] Second, unlike any standardized flow scale which provides numerical measures of flow intensity, the Flow Questionnaire does not 'impose' flow on respondents; that is, it does not arbitrarily assume that everybody experiences flow in general or in a specific context. (Moneta, 2012b, p. 494)

Also, Moneta (2021, p. 62) considers the FlowQ as one of the "main measurement methods for flow", the other being the ES Form[48] plus all others "standardized scales of the componential approach." This sentiment was shared by Redaelli & Riva (2011, p. 12), who also see the FlowQ's main application on conventional flow research: "Conventional Flow research has adopted two main methodologies: The flow questionnaire, […] and the [ESM …]". Furthermore, when referring to the Flow-Q and the Dispositional Flow Scale, researchers Rufi et al. praised its versatility: "… both are the most supported models according to cross-cultural studies"(Rufi et al., 2014, pp. 3–4).

[48] Although not specifically designed nor for flow detection nor measurement but historically employed as such.

We don't forget to mention the study that employed the original version of the FlowQuestionnaire (Asakawa, 2010, app. Flow Quotations), which differed with the current version in that (a) the first question "unlocks" the following two questions, and (b) the last question is rated on a 7-point Likert scale, which renders its scoring less straightforward.

The FlowQuestionnaire is available in Appendix 2. – The FlowQuestionnaire in both English and French versions, as well as its predecessor.

Flow-Kurzskala (FKS) / Flow Short Scale (FSS)

The Flow Kurzskala (Engeser & Rheinberg, 2008; Rheinberg et al., 2003) is an attempt to combine the questionnaire technique with the ESM technique, *i.e.*, to have a procedure that captures flow in its various components during ongoing everyday activities, upon receiving a signal.

For this purpose, a short scale was needed that could be answered upon [receiving] a signal with "as little interruption of activity as possible", and capable to "to fit any activity" (2003). So, the FKS is a standardized scale that assesses all components of the flow experience, according to its authors (Rheinberg & Engeser, 2018, p. 612) as a flow short (10 items), 7-points Likert scale (2003). It usually takes 30 to 40 seconds to assess flow and it is suitable for post-event, completed activities but it also has a part for "ESM-based assessments of ongoing activities" (2018, p. 606).

Additionally, the FKS can also measure perceived difficulty in a 9-point Likert scale, and perceived importance/current worries in a 9-point Likert scale (2008, p. 162; 2018, p. 603), by replacing the last three items (perceived difficulty *vs.* perceived importance/worries).

As a side note on the relationship between flow and performance, the results of one of the FKS / FSS studies (2008, p. 169) led authors to conclude that the "The flow state, […] measured by the Flow Short Scale, predicts performance."

The Flow-Kurzskala has been translated into several languages (2018, p. 606) and it is available in both German and English versions, along their difficulty/importance declinations in the Appendix 3. – Die Flow-Kurzskala.

Flow Scale

The Flow Scale (Mayers, 1978) elicits an estimate of the frequency with which a person experiences each of the nine dimensions of the flow experience on a semantic, 12-item Likert, 8-points, differential scale (Csíkszentmihályi, 2014, p. 246; Fave & Massimini, 1988). This instrument has been used as a repeated measure to identify a person's flow activities and his/her ratings thereof to compare them to ratings of known, set of daily activities, *e.g.*, along with the FlowQ (Fave & Massimini, 1988).

More recently, a general flow scale was developed based on existing questionnaires to be applied in specific contexts such as during sport activities (Jackson & Eklund, 2002) and psychotherapy (Parks, 1996). Seemingly, this scale was first utilized in a school classroom context. However, probably due to the unpublished nature of the main research source for this scale (Mayers, 1978 unpublished doctoral thesis) we could not locate it in the sources available to us. Moreover, according to Lonczak (2019), unofficially it is unclear "whether the scale has been validated" (2019).

The Flow State Scale 2 (FSS2) & the Dispositional Flow Scale (DFS-2)

Before presenting the FSS2 & DFS-2, we feel obliged to introduce the original versions of these scales. The FSS and the DFS were designed to examine flow in physical activities, more precisely to understand the relationship of flow to other psychological factors, and to be completed on "events recently experienced" related to a "past flow experience that stood out for them" (Jackson & Eklund, 2002; Jackson & Marsh, 1996). Based on their previous works, Jackson & Marsh (1996) hypothesized that "flow is the psychological process underlying peak performance" and found "correlational support for this idea".

The FSS was designed to assess the state of a situation-specific experience of flow while the DFS meant to assess the dispositional tendency of the frequency with which people experience flow in a physical activity (2002; 1996).

The DFS development as a FSS variation obeyed to the findings by the Csíkszentmihályi couple (1988) that "there were individual differences in the ability to experience flow" (2002) that "are in part inborn, but it certainly can be learned" (1988) and that "the term autotelic personality applies to this propensity to experience flow" (2002). As an interesting side note, the Rufi et al. (2014, pp. 3–4) journal article accorded the DFS the same value to the FlowQ in cross-cultural studies: "[…] both are the most supported models according to cross-cultural studies".

In both the FSS and DFS, special consideration on the multidimensional approach of flow measurement was taken. nine key flow characteristics, or dimensions, are assessed, based on the nine-dimensional conceptual flow model (extensively described in Section "Evolution of the Definition of Flow and its Conceptual Flow Model" above). Items were also developed by employing "qualitative reports by athletes, and other measures of flow or related experiences" (2002). These qualitative reports were particularly "important in developing the wording of items in that it provided actual descriptions of flow states in athlete's own words." (1996).

After the designing of both scales, their authors (2002) relate the results of their subsequent studies with correlations between flow ratings and performance, with "flow state dimensions […] reported as being positively correlated with measures of perceived skill and perceived success, subjective performance ratings, and over-all finishing position".

Item modifications were made to the FSS and DFS to improve the measurement of some of the flow dimensions leading to the FSS-2 & DFS-2 measurement instruments. Two new items replaced "problematic" ones related to loss-of-self-consciousness and unambiguous-feedback, while two new time-transformation items covering time-slowing items as well. Lastly, a new sense-of-control item replaced a previous total-control item in both the FSS and DFS (2002).

As such, the Flow State Scale-2 (FSS-2) and the Dispositional Flow Scale-2 (DFS-2) were presented by Jackson & Eklund (2002) simultaneously as "two self-report instruments designed to assess flow experiences in physical activity", although the Stoll & Ufer (2021, p. 363) publication claimed their uses was not limited to sports. They each feature four items per each of the nine-dimensional conceptualization of flow (2008; 2002).

Thus, the FSS-2 is a 36-item measurement instrument designed to assess the extent to which they experienced the flow characteristics on a 5-point Likert scale (strongly disagree → strongly agree). On its part, the DFS-2 is also 36-item measurement instrument, although designed to assess the frequency with which they experience the flow characteristics within a specified activity in general, *i.e.*, an individual's flow propensity within a given activity (2008, pp. 567–568).

Although both instruments were developed and validated in physical activity settings, they have also been used in other performance-related domains as well (2008). Both instruments have been translated and validated in several languages, with minor cultural variation suggested (Moneta, 2021, p. 60).

Shorter versions (9-items each) of both DFS-2 & FSS-2 instruments feature only item per dimension when compared to their full-size counterparts. They also employ the same response format (5-points Likert scale) (2008, p. 568).

The FSS (English) (1996) and the DFS (Spanish) (García Calvo et al., 2008) can be found in Appendix 4. – The Flow State Scale (FSS). We could not find the DFS-2 nor the FSS-2 within our available academic resources.

Flow in Human-Computer Interaction

This flow measurement instrument was designed within a typical work-related computer session context in mind (Ghani & Deshpande, 1994). The scale is an 15-item, 7 & 10-point Likert instrument measuring enjoyment, concentration, perceived control, exploratory use, and perceived challenge (1994) on a post-event episode.

It aimed to examine the mediating role of flow and the moderating impact of task scope on the factors affecting user behavior, where flow resulted from the "perceived task challenge" and "a sense of being in control" factors (1994).

The entirety of this measurement instrument is available in the Appendix 5. – Flow in Human-Computer Interaction.

The Learning Flow Scale (3 versions)

For the sake of completeness[49], we mention the works of Korean researchers (A.-Y. Kim et al., 2010; Y.-Y. Kim et al., 2017; Suk & Kang, 2007) on their "Learning Flow Scale" for elementary education contexts, for high school contexts, and for adults in learning contexts, although we could not find validated English versions of their respective scales, being mostly employed in Korean environments.

The Learning Flow Scale was firstly developed and validated in 2007 to "measure flow level in real learning situation" (2007) in a Korean, elementary school context. Initially, this scale numbered 51 items, accounting for eight of the flow factors of flow (with the "Time transformation" missing). Subsequently, it was reduced to 30 items while accounting this time for the nine flow dimensions surveyed by Csíkszentmihályi (1975a, 1975b).

Subsequently, the learning flow scale for adults (2010) was developed, accounting 29 items spanning nine subscales, with three items per each flow dimension except for the autotelic

[49] At the time of writing this thesis, the article with the most updated version of this scale (2017, High School students) accounts only one citation in the Google Scholar search engine, which might be an indication it has been abandoned.

dimension, which is measured through five items. Items are rated on a 5-point Likert scale ranging from "strongly disagree" (1) to "strongly agree" (5), with higher scores indicating a stronger learning flow (J.-W. Han et al., 2020, p. 3).

Ten years later, both scales were subsequently employed as a basis to develop the Learning Flow Scale for High School Students (2017), comprising 30 items, measuring the nine factors of flow (1975a, 1975b).

These three scales primarily rely on the flow Immersion dimension. Explicitly mentioned in the case of the latter scale, authors referred to it as "Immersion in learning" and defined it as [experiencing immersion in learning situations] (2017 software translation). Authors of this scale argued this dimension is an important promotor for higher levels of concentration and participation, which they claim to be required for higher level learning based on previous studies (Eppler & Harju, 1997). It also shortens learning time, and enables students to be active in their learning (2017). Indeed, authors claimed the motivation behind the development of such scale was to objectively measure the degree of immersion in learning situations (2017).

As previously mentioned, no version in English[50] was found, validated or not. Again, besides the obvious operational need for adapting (*i.e.*, translating) a scale, such task takes valuable research time and work for its design, internal consistency validation, and final evaluation (*cf.* Notes at the end of Section "Flow in Educational Contexts" above).

The Study-Related Flow Inventory (WOLF-S) & The Work-Related Flow Inventory (WOLFS)

The WOrk-reLated Flow scale (WOLF) was "developed to measure flow in a work setting" (Bakker et al., 2017, p. 147), and to "develop a reliable and valid instrument for the assessment of work-related flow" (Bakker, 2008, p. 401).

[50] Nor Spanish, French, Portuguese, or German; the languages the author of this thesis is most comfortable understanding.

It was conceived upon the premise that flow is mostly characterized by three components: absorption, enjoyment, and intrinsic motivation, and as a short, intensive experience (2008, p. 401). In a work setting, these components are considered as:

- Absorption refers to a state of total concentration, whereby employees are totally immersed in their work. Time passes quickly, and they forget everything around them.

- [On Enjoyment:] Employees who enjoy their work and feel happy make positive judgments about the quality of their working life. This enjoyment or happiness is the outcome of cognitive and affective evaluations of the flow experience.

- Finally, intrinsic motivation refers to performing a certain work-related activity with the aim of experiencing the inherent pleasure and satisfaction in the activity. […] Employees who are motivated by the intrinsic aspects of their work tasks want to continue their work; they are fascinated by the tasks they perform. (Bakker, 2008, p. 401)

According to the Bakker et al. (2017, p. 148) publication, research has shown the "reliability, as well as the factorial, constructive, and predictive validity" of WOLFS. Yet, most recent research (Heutte, 2020, p. 37) points out that the "intrinsic motivation" element considered not only by Bakker (2008, p. 401) but also by other authors (Engeser & Rheinberg, 2008; Lecomte, 2006, 2009) as a component of flow, is instead a to be understood as a consequence of flow, *i.e.*, in work conditions, the very first time one is ordered to perform a [novel] task the intrinsic motivation is nonexistent, as there is no free choice on the matter, thus the motivation cannot be intrinsic[51].

The WOLF-S is born out of the observation of the scarcity of studies concerning the measurement of flow in educational contexts: "[…] little is known about the applicability of these scales to the experience of flow in students' academic work" (2017, p. 148).

[51] That is the "paradox of [flow at] work" (Heutte, 2020), which is resumed as: « *on fait pas ce que l'on veut au travail, pourtant, si l'on réussi, ça deviant plaisant* », [at work, we don't do what we want although if we succeed it becomes enjoyable].

It is an adaptation of the WOLFS inventory "adapted for measuring the flow experience in an educational setting, *i.e.*, study-related flow" (2017, p. 147). Its creators (2017, p. 148) employed the WOLFS as a basis for the WOLF-S because of its strong psychometric properties when employed in work settings. The WOLF-S sees flow as its predecessor, as "a short-term peak experience during study activities that is characterized by absorption, study enjoyment, and intrinsic motivation for these activities" (2017, p. 149).

Accordingly, these components (Bakker, 2005, 2008) are now adapted to reflect the academics settings:

> Absorption refers to a state of total concentration, whereby students are totally immersed in their academic work. Enjoyment refers to a positive judgment about the quality of their study and academic obligations. Intrinsic study motivation indicates the desire to perform a certain study related activity with the aim of experiencing inherent pleasure and satisfaction in the activity. (Bakker et al., 2017, p. 149)

Flow is measured via thirteen items; four items for absorption, four for study enjoyment, and five for study motivation; all items are rated on a 7-point Likert scale (never → always) (2017, p. 150). Participants are to indicate how often they experienced each of the statements during a specific, previous time span (Salanova et al., 2006, p. 9).

The WOLF-S has been reported to correctly correlate to flow in "learning activities during secondary school and university" (2017, p. 149), although the remarks by Heutte (2020, p. 37) on the WOLFS measurement instrument (upon which WOLF-S exists) hint to the uncertainty that the WOLF-S does measure specifically flow in educational contexts.

The Study-Related Flow Inventory (WOLF-S) is validated and available in English and in Croatian in Appendix 7. – The Study-Related Flow Inventory (WOLF-S), and the Work-Related Flow Inventory (WOLFS) is available in Appendix 6. – The Work-Related Flow Inventory (WOLF).

Flow in Education (EduFlow & EduFlow-2)

We conclude this round of flow measure instruments previously employed in educational contexts with the EduFlow-2 (*a.k.a.,* EduFlow-2) measurement instrument.

EduFlow-2 is a measure instrument issued from and considered the successor of the EduFlow theoretical model[52] (Heutte, Fenouillet, et al., 2014; Heutte, Fenouillet, Kaplan, et al., 2016, p. 9). According to Heutte et al. (2021, p. 3), previous to the development of the first Flow in Education Model (EduFlow), no "short, multidimensional scale designed specifically for education" existed.

Although Delle Fave et al. (2011) insisted that flow measurement instruments should "adhere strictly to the nine original dimensions of flow" (2016, p. 130), subsequent research (Heutte, Fenouillet, et al., 2014) observed that this might be difficult to follow in educational context as "[…] barely half of the nine dimensions of FFS2 are actually perceived by learners […]" (2016, p. 130) and explained as plausible that not all nine flow components would be "equally prominent in all contexts" (2016, p. 131).

Consequently, when designing the Flow in Education Scale (EduFlow), its originators (2014; Heutte, Fenouillet, Martin-Krumm, et al., 2016) retained only the following four flow dimensions (2016, p. 131): Cognitive absorption, Time transformation, Loss of self-consciousness, and Autotelic activity. However, EduFlow showed to lack "one of the most important components of the flow experience: intense and focused concentration on the present moment" (2016, p. 17; Nakamura & Csíkszentmihályi, 2009).

This situation led to the EduFlow-2 measurement instrument (2016, p. 24), where Immersion is compounded into the existing Time transformation dimension, and Cognitive absorption is replaced by the Cognitive control dimension (*cf.* Section "Flow in Educational

[52] Based on the Flow in Education theoretical model quickly overviewed at the end of Section "Flow in Educational Contexts".

Contexts" above), preserving without changed the Loss of self-consciousness and Autotelic experience dimensions, totaling four flow components.

Thus, EduFlow[53] and EduFlow-2 are measurement instruments designed specifically for flow measurement in educational contexts (El Mawas & Heutte, 2019; 2021; 2016; Heutte & Fenouillet, 2010). They are both gender neutral, short twelve-item scale differentiating four flow dimensions relevant to cognitive processes, where each dimension is measured by three items (El Mawas & Heutte, 2019; Heutte et al., 2021; Heutte, Fenouillet, Kaplan, et al., 2016; Heutte, Fenouillet, Martin-Krumm, et al., 2016). Furthermore, the EduFlow-2 measurement instrument supports both the flow-continuum, and the binary flow presence/absence conceptual flow model frameworks.

The EduFlow-2 measure instrument has proven to be useful in studies of cognitive activities and be suited to flow measurement in various educational contexts, specifically in MOOC (online, asynchronous, distance learning) and classroom (offline, synchronous, presential learning) situations (Heutte, Fenouillet, et al., 2014; 2016; Heutte, Kaplan, et al., 2014).

The EduFlow[54] and EduFlow-2 measurement instruments have been both tested and validated in English and in French, which are available in Appendix 8. – The EduFlow & EduFlow-2 measure instruments, while other languages (Arab, Croatian, Farsi, Italian, Indonesian, Polish, Portuguese, Romanian, Swedish) are on the works (J. Heutte, personal communication, July 10th, 2023).

In this Subsection we completed an operational overview of flow measurement instruments historically employed in educational contexts, plus the ESM. Whenever possible, we presented the corresponding instrument in the Appendices. We concluded this Subsection with two flow measurement instruments specifically constructed for (EduFlow & EduFlow-2), or adapted to (WOLF-S) educational contexts (Bakker et al., 2017; 2021; 2016). Indeed, before both

[53] EduFlow is known to have been employed as a basis for the creation of a flow scale in physical education in Arabic as well (Abbassi et al., 2021).
[54] More recently, EduFlow is available in Arabic as well (Chalghaf et al., 2019).

the EduFlow and EduFlow-2 measure instruments (2021; 2016), there was no short, multidimensional scale specifically designed for flow research in educational contexts (Heutte, 2017b).

Lastly, in this Section we presented a general overview of flow research and applications in educational contexts. We highlighted the importance of the psychological state in the learning process and justified the involvement of flow in learning, exemplifying its effects in flow measurement in educational contexts. Finally, we included an operational collection of flow measurement instruments historically employed in educational contexts.

In the next Section we census research-supported arguments to acknowledge when developing flow measurement instruments or protocols. Their analysis will bring a small anthology of suggestions to the same end.

Considerations When Measuring Flow

This Section presents research-supported arguments to consider when developing flow measurement instruments or protocols. These motives led to a compilation of research-based but brief suggestions to the same end. We suggest the reader to first refer to Section "Measurement Attempts of Flow" above (if not yet done so) beforehand to better grasp our meaning and intention forward.

First of all, acknowledging the multidimensional nature of flow, Jackson & Marsh (1996) insisted on the importance "to establish (through construct validation approaches) the dimensional nature of flow and to develop instruments designed to measure the dimensions." At the same time, and seemingly in contradiction, Hoffman & Novak (2009, p. 27) recommended flow researchers using more than one type of flow measure instrument when designing new flow measurement protocols, *i.e.*, a well-informed combination of unidimensional and multidimensional instruments.

This is because, on one hand, generally simple, unidimensional measures of flow reduce the data collection burden, both on the participant's side as well as on the data-collecting side, facilitating administration when employed in repeated measures protocols (Hoffman & Novak, 2009, p. 27).

On the other hand, multidimensional flow measures help to identify higher-order factors to provide a more holistic definition of flow, prompt for statistical fit in structural models (2009, pp. 27–28).

A more recent literature review (de Moura Jr & Porto Bellini, 2019) substantiates the trend that research studies are headed in the right direction: all domains and research fields combined, flow is frequently measured in association with other constructs or by means of proxies: "In a large number of cases, flow is studied along with other constructs, while the balance between challenges and skills and enjoyment are particularly popular as proxies for flow" (2019, p. 546).

Likewise, while developing the FSS and the DFS measurement instruments, Jackson & Eklund (2002) remarked the importance of having "self-report measures that […] be used without disrupting performance". They also endorsed that "More generally, a multimethod approach is needed to understand flow, incorporating both qualitative and quantitative research" (Jackson & Marsh, 1996, p. 21). The findings of authors behind the Jackson et al. (2008, p. 562) journal article point out that, among the direct & indirect flow-measurement approaches created and available for researchers (that they are aware of), asynchronous post-event instruments "appear to be the less intrusive" to the flow experience.

In the light of the results obtained by Engeser & Rheinberg (2008, p. 169), it was suggested that flow measurement should ideally be done through its composing dimensions, in an unobtrusive manner (online):

> Future research should probably not only (operationally) define flow with only one component (the skill-challenge balance) and instead measure flow in its

> multidimensionality. Most ideal would be to measure flow 'online' via unobtrusive physiologically based indicators or with some reliable and observable aspects of behavior or expressions. (Engeser & Rheinberg, 2008, p. 169)

This advice remains valid a decade later, when again Engeser & Rheinberg (2018) insisted that flow measurement should "ideally" be applied directly "online" as the activity is performed.

But measuring flow is only half of the story, Engeser & Rheinberg (2008) highlighted the fact that when studying flow, one is to consider the non-presence of flow as well: "When we study flow, we are also studying the absence of flow" (2008, p. 170). That is, to devise instruments that account, at least theoretically, what is being measured if all indicators point to null. Similarly, the previously mentioned survey (de Moura Jr & Porto Bellini, 2019, p. 546) refers the works of Allison & Duncan (1987) concerning the emergence of the "anti-flow" phenomenon, which we are aware of but do not address in this thesis.

Concerning the validity of a post-event recalling and thus validly capturing it in a questionnaire, Jackson & Eklund (2002) acknowledged that a "retrospective approach was a design limitation in that the responses could have been influenced by the passing of time". Despite this possible problem, they suggested that the post-event flow responses could be considered reliable and validated if "provided additional information", implying clear and concise instructions about the period time inquired, *i.e.*, when flow was supposed to have occurred.

Yet, even if the previous considerations are accounted for and a new measurement instrument (or new measurement protocol, comprising more than one instrument) is developed, another unseen factor arises. Engeser & Schiepe-Tiska (2012, p. 4) call our attention to the fact that components of flow have proven to be highly correlated, which would lead to the premature conclusion that "all components of flow could be represented by one dimension only" (2012, p. 4). Instead, their literature review pointed towards the conclusion that "flow cannot be reduced to a single component, and all attempts to take one component of flow as the definitional aspect of flow will consequently disregard essential parts" (2012, p. 4).

Thus, proceeding from the previous research arguments, we can conclude on a compendium of considerations to recognize when developing new flow measurement instruments or protocols (order is unimportant):

- Employ a combination of unidimensional and multidimensional instruments.
- Employ non disruptive, unobtrusive methods.
- Approach the measurement through flow components. This does not mean to approach flow measurement through other theoretical notions, *e.g.,* engagement, immersion, etc.
- Construct and employ reliable, representative flow components indicators (*cf.* Section "Indicators construction" in Chapter 7 Experimentation below).
- Reduce the cognitive load of participants when answering the measurement instrument.
- Be clear, precise, and succinct in the phrasing of, and surrounding the measurement instrument.
- Reduce the time delay between the inquired flow event and data gathering.

Interestingly, not much is found in the literature concerning flow measurement once the measurement instruments have been applied, *i.e.*, multi-instrument scores weighing, ponderation, or confrontation, etc. Still, we discerned two major consideration points:

- Consider conceptually the opposite case of "measuring flow", *i.e.*, no-flow? anti-flow? flow absence?
- Be aware that all flow components will likely be highly correlated among them.

Thus, in this Subsection we presented research-based arguments to consider when developing flow measurement protocols. These arguments led to a compilation of research-based suggestions to the same end. This list of considerations is central to the determination of the flow measurement instruments to employ in this thesis project, explained in Section "Flow measurement instruments identification" in Chapter 7 Experimentation below.

We have reached the end of the first chapter of this thesis. Here, we inspected the concept of flow. We began with the initial research of flow and the context in which it developed. We detailed the changes manifested to the definition of flow, the theoretical flow model, and the components of flow, of which the challenge-skill balance has historically proven of high importance. We quickly overviewed what might constitute general obstacles to flow emergence. We continued with a literature review on how researchers have tackled the measuring of flow, the part of flow in education and learning, along with an operational description of flow measurement instruments employed in such context. Finally, we end this chapter with a compendium of informed considerations to acknowledge when attempting to measure flow.

By now the reader should be aware of what flow encompasses and how it is characterized. We want to reassure the reader that although the previously shown evolution might seem chaotic and somewhat arbitrary, it holds a sound scientific structure that we might have failed to transmit. Furthermore, we want to emphasize that this evolution took place in a relatively short span of time, involving many researchers from different universities all around the world.

On the subject of flow measurement, we want to conclude by reprising (*cf.* Section "Measurement Attempts of Flow" above) the following (extended) quote from a very recent book chapter by Moneta (2021, pp. 31–32), on the suggested directions for conceptual and measurement research on flow:

> This process had some chronological order, but was not always linear or perfectly logical. This pattern is common in science, and in the history of psychology in particular, although researchers may differ in the extent to which they are aware of it. The key message of this chapter is that no existing measurement method for flow and associated

> model is watertight, and that a gold standard for the modeling and measurement of flow is not at close reach. (Moneta, 2021, pp. 31–32)

This thesis is concerned with flow detection in a MOOC environment. In this Chapter 1 we have well delved into the flow notion, with particular attention to flow measurement in educational contexts, with remarkable findings developed in the online and distant contexts as well. Thus, the following Chapter 3 regards the terrain for experimentation for this thesis project, the Massive Open Online Courses, to present their definition, their historical journey, and a few relevant typologies.

Chapter 3. Massive Online Open Course

This Chapter presents the Massive Open Online Course (MOOC), its basic characteristics, its origins, and some basic categorizations relevant to this research work. Parts of this Chapter appear on published articles by this thesis' author (Ramírez Luelmo et al., 2020a, 2020b, 2021a, 2021b).

We begin thus by adopting the ample and complex definition coined by the parents of MOOC:

A Massive Open Online Course (MOOC) is primarily the integration of social networking, an acknowledged expert in the field in question, and a collection of freely accessible online resources (Cormier, 2010; McAuley et al., 2010, p. 4; *What Is a MOOC?*, 2010).

Furthermore, McAuley et al. (2010, p. 4) maintains that a MOOC:

1. Builds on the active engagement of self-organizing participants.
2. May feature a predefined timeline and topics.
3. Generally, it carries no fees and has no prerequisites[55].
4. Generally, it has no predefined expectations and no formal accreditation.
5. Spreads its offer through an online social network, via a public website.

But they also concur that these are just direction pointers and therefore any given MOOC is to diverge from them as much as their participants and organizers need, or pretend to achieve (2010, p. 4). Therefore, the reader should be aware that much of what is treated in this Chapter consists of "commonly accepted generalizations" that research experience has shown so far to be accurate enough for all intents and purposes.

[55] Other than Internet access and the participant's own interest.

Indeed, the need for a basic definition of the meaning behind MOOCs (as in the interpretation of its building blocks), and of their usage perspectives is still an issue waiting to be solved (Stracke et al., 2019, p. 338). Such issue is not within the scope of this thesis, but we care to point out that among the four fundamental meaning-carrying elements, the Online component of MOOCs is the easiest one to achieve and the less divisive within the MOOC community (Stracke et al., 2019, p. 335), which helps to paint a panorama of the MOOC definition issue. Notwithstanding, a MOOC is an example of a Technology Enhanced Learning (TEL)[56] (Yousef, Chatti, Schroeder, & Wosnitza, 2014, p. 1).

In general, MOOCs are known to employ two components for different steps in the course lifecycle: the Content Management System (CMS) and the Learning Management System (LMS). The former is tasked with managing the learners' enrolment, track their performance and/or create and distribute the course's content (El Mawas, Gilliot, et al., 2018). The latter is tasked with course management duties, such as registering users, course tracking, users' data saving, and analytics (El Mawas, Gilliot, et al., 2018).

Currently, MOOCs experience a high dropout rate, a very well-known phenomenon exemplifying the lack of students' engagement (Jordan, 2014; Yuan & Powell, 2013). Studies confirm that social factors, such as the MOOC's learner's community (Dalipi et al., 2018; Fang et al., 2019; Sun et al., 2019), or the inability of the MOOC's design to adapt to the distinct participants' cultural background (Dalipi et al., 2018; Heutte, Kaplan, et al., 2014, p. 22), play a positive part, but they are mostly limited, again, by learners' engagement and self-regulation (Heutte, Kaplan, et al., 2014, p. 22; Kaplan, 2014).

Furthermore, the Breslow et al. (2013) publication pointed at the multifactor complexity of this phenomenon by registering the different reasons participants might exhibit to enroll in a MOOC other than course completion such as course-shopping, dabbling topic courses, auditing

[56] *Environnements Informatiques pour l'Apprentissage Humain* (EIAH) (Balachef, 2018, p. 66).

knowledge on the material and on its difficulty level, just to name a few. For instance, according to the Center for Digital Education (CEDE) & École Polytechnique Fédérale de Lausanne (EPFL) (2016) report, some of the reasons for students to enroll into one of the above-mentioned EPFL MOOCs might be: finding a new job, getting a promotion, meeting family expectations, earning a higher salary, solving a specific problem, and/or helping to pass class.

Clearly, the obvious heterogenous nature of any given global MOOC learner (along with their heterogenous learning needs) is a major factor to consider when delving into this phenomenon (Heutte, Kaplan, et al., 2014; Sein-Echaluce et al., 2016, p. 22). Indeed, studies highlight that the learners' background is quite different. For instance, in her 2019 article, El Mawas et al. (2019, p. 471) remarks that in MOOCs held by the EPFL, prior knowledge surveyed from the MOOC learners is spread among a variety of fields: 34% of learners come with a mathematical, computers or engineering background, 21% with architectural or civil engineering, 12% in education and training, 13% with construction, food, utilities, healthcare, or life-sciences, 4% with arts, design, entertainment, etc., 2% with business, finance, or management, etc., and finally 2% with legal, administration, or social services.

On that matter, a journal article (Kizilcec et al., 2013) stated that:

> Given the heterogeneity of the population, we would be remiss to make *a priori* assumptions about the appropriate characteristics or behaviors around which to categorize learners, or which pathways and outcomes are more or less valuable for their learning. (Kizilcec et al., 2013)

Still, in general, current research directs our attention to the learners making allusions to the fundamental motivations that drive human beings, such as the need for resources, status, and appreciation (El Mawas, Gilliot, et al., 2018).

Indeed, MOOCs remain a complex subject to tackle in research, whether on its definition or its components', their high drop-out rate, or the reasons for enrolling, just to mention a few, relevant to this research work.

A Brief Account

In this Section we overview the history of MOOC; we present their general characteristics, and we introduce their usage and deployment trends, highly accentuated during the recent COVID-19 pandemic.

The MOOC phenomenon started in 2007 when David Wiley created the first MOOC (or Proto MOOC) at the Utah State University (Mota & Scott, 2014; Zhu et al., 2020).

In 2008, Dave Cormier and George Siemens first coined the term to refer to a course entitled "Connectivism and Connectivity Knowledge"[57] (CCK08), developed by Stephen Downes and George Siemens in September of the same year for the University of Manitoba, Canada (Creed-Dikeogu & Clark, 2013; MAUT, 2015; Zhang, 2013; Zhu et al., 2020). Their intention was to exploit the possibility for interactions between a wide variety of participants made possible by online tools to provide a richer learning environment than traditional tools would allow. On the premises of the campus 25 tuition-paying students attended the course, a further 2 300 from around the world participated online (MAUT, 2015).

MOOCs began to gain popularity shortly after the appearance of Siemens & Downes' CritLit "Critical Literacies" and PLENK2010 "Personal Learning Environments, Networks, and Knowledge" open courses during the summer and fall of 2010, each hosting 377 and 1 610 participants, respectively (Creed-Dikeogu & Clark, 2013; Kop, 2011; Zhu et al., 2020). They were based on a Moodle[58] environment, a wiki (for storing the course's resources, information, and recordings), and a daily newsletter (Kop, 2011).

One year later, in the fall of 2011, Stanford offered three courses for free online. Professors Peter Norvig and Sebastien Thrun offered their "Introduction to Artificial Intelligence"[59] to an initial enrollment of over 160 thousand students from around the world.

[57] Originally [http://ltc.umanitoba.ca/connectivism/] but now a defunct link.
[58] https://moodle.org/
[59] https://www.udacity.com/course/cs271

Over 20 thousand students completed the course (Creed-Dikeogu & Clark, 2013). These actions finally booted the modern MOOC movement as we know it (Shah, 2021c). Since, MOOCs popularity have gained traction due to their LifeLong Learning component, *i.e.,* learning anytime, anywhere.

According to a recent systematic review on the topic (Zhu et al., 2020), MOOCs feature characteristics such as free entrance (or at a minimum cost) (Zhang, 2013), they promote discussions involving a large number of students (Kellogg et al., 2014), they provide learning flexibility in terms of time and place (Pérez-Sanagustín et al., 2016), and they allow diverse tasks in one course (Soffer & Cohen, 2015).

As previously mentioned, MOOC main requirement (Zhu et al., 2020) is a working internet connection for anyone to enroll and to gain access to the learning materials (Kop, 2011; Koutropoulos et al., 2012). Researchers (Bali, 2014; Bulfin et al., 2014; Carver & Harrison, 2013; Jacobs, 2013; Liyanagunawardena et al., 2013; Zhu et al., 2020) have long agreed that MOOCs provide increased access to higher education worldwide.

Every year, MOOC statistics[60] surpass those of the previous year (Shah, 2015, 2016b, 2017, 2018, 2019, 2020b, 2021b), with the most current (end of 2022) numbers accounting 220 million learners worldwide, 950 participating universities, almost 20 thousand courses, providing 1 670 microcredentials, and 70 MOOC-based degrees (2021b).

Typically, if learners do not intend to obtain a certification the registration fee is waived (Zhu et al., 2020). As such., the business model of a MOOC is not obvious, but we can shine light on the matter by peeking into the agreement governing Coursera's relationship with the University of Michigan at Ann Arbor, celebrated in 2012, in which eight possible business models were disclosed, including certification, secure assessments, employee recruiting, employee or university screening, human-provided tutoring or manual grading, corporate/university enterprise

[60] Not accounting for China, whose metrics might be unreliable (Shah, 2021c).

model, sponsorships, and tuition fees (Shah, 2021a; University of Michigan & Coursera, Inc., 2012).

Indeed, since 2016, the newest trend in MOOC is MOOC-based credential and degree programs (Zhu et al., 2020), where even newer MOOC providers have joined the large, most known MOOC providers, such as edX, Khan Academy, Coursera, MITx and Udacity (Shah, 2016a). For instance, edX has been running paid courses called "Professional Education" since 2014, Kadenze started their own credential system "Kadenze Programs" in 2016. In the same year Coursera added courses where all materials need to be paid for, as well as "Coursera for Business", aimed at private companies and their employees (Shah, 2021a; Zhu et al., 2020). From an economic point of view this makes sense, as MOOCs boomed (2011~2012) without a business model, being at the time mostly free ("We are committed to making the best education in the world freely available to any person who seeks it." – Coursera's mission statement), with many wondering how free courses would ever make money (Pappano, 2012; Shah, 2021a).

More recently, in mid-March 2020, when quarantine measures linked to the COVID-19 pandemic went into effect in many countries around the world (Onyema et al., 2020), many turned their attention to online courses (Mack et al., 2023, p. 105; Parker et al., 2020) – especially to free courses (Shah, 2020c). As a result, by April 2020, MOOC providers Coursera, edX and FutureLearn saw drastic growth, with their sessions numbers averaging a 78% increase, attracting as many new users in a single month (April 2020) as they did in the entirety of 2019 (Shah, 2020a, 2020c). Coursera received over 10 million course enrollments in a 30-day period, a 644% increase compared to 2019, while edX became one of the world's top 1000 websites[61] (Shah, 2020a). In total, 2020 saw an estimated additional 40 million new MOOC learners; that is one third of all-ever MOOC learners registered (Shah, 2020c).

[61] Based on Alexa Ranking, a now-defunct web service since May 1st, 2022.

It is noteworthy to highlight such previously unseen growth did not come from campuses adapting to the COVID-19 pandemic but from social distancing policies coming into effect worldwide (Shah, 2020a), which clearly points to a personal component at play rather than an institutional one.

Furthermore, the sudden push to distance learning had a never-before-seen impact on all levels of educational systems worldwide (Carroll et al., 2020; Daniel, 2020; Kichu & Bhattacharya, 2021; Onyema et al., 2020; Pokhrel & Chhetri, 2021; Tadesse & Muluye, 2020), with students from public institutions in developing countries being the most disadvantaged in access to digital technology and educational materials (Tadesse & Muluye, 2020, p. 161). Still, this push seemingly fulfilled UNESCO's (UNESCO, 2016b, p. 24) consideration regarding MOOCs as global contributors to "the democratisation [*sic*] of [Higher Education]" (2016b, p. 24) but again, UNESCO remarked that "most MOOC participants today are well educated and have already had access to [Higher Education]" (2016b, p. 25).

Undeniably, MOOCs are not to be considered as "the big idea" itself but instead a tool to "the service of big ideas" (UNESCO, 2016b, p. 26), contributing to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" (2016b, p. 26, 2016a, p. 20).

Yet, fueled by the global pandemic, MOOCs and online Learning Environments (LE), indisputably became indispensable tools to bring the classroom to the learners (Doghonadze et al., 2020, p. 32; Feng et al., 2020, p. 170; Nabukeera, 2020, p. 189; Tadesse & Muluye, 2020, p. 161).

This Section overviewed the history of MOOC, their characteristics and trends in their deployment and usage. In concluded on its highlighted importance during the recent COVID-19 pandemic.

Relevant Categorizations

Since their inception, MOOCs have been distinguished into different typologies, taxonomies, or categorizations according to their "types, pedagogies, orientations, target participants, resources and content" (Pilli & Admiraal, 2016, p. 1), or more specifically, participants' behavior (Cisel, 2017), pedagogical scenarization (Ebner et al., 2020), or MOOC offer (Vrillon, 2017), to name a few examples. We can say that classifying MOOCs is an attempt to "identify, group, properly name and describe the[ir] nature […]" (Pilli & Admiraal, 2016, p. 3) to better fathom their interrelationships.

This Subsection does not intend to cover all distinct MOOC taxonomies[62]. Instead, it overviews first, the main, most known MOOC categorization, and second, it completes it with a few MOOC typologies relevant to this thesis.

First and foremost, MOOCs' most known distinction is between xMOOC and cMOOC (Rosselle et al., 2014, p. 4; Yuan & Powell, 2013, p. 7). Originally, they were named them "transmissive MOOC" (Hollands & Tirthali, 2014, p. 18) and "connectivist MOOC" (Siemens, 2005; Vrillon, 2017, p. 4) respectively; with the "c" conveying the meaning of connectivism, and the "x" meaning "exponential", to carry the idea of a massive participation, and/or also denoting "extension", as in HarvardX (an extension of the Harvard campus), or as in MITx (an extension of MIT[63]) (Hollands & Tirthali, 2014, p. 18).

Their main distinction being that xMOOCs are considered more teacher-led in terms of content, structure, and assignments while cMOOCs are seen as more social and non-hierarchical (Heutte, Kaplan, et al., 2014, p. 14; Hollands & Tirthali, 2014, pp. 18, 34; MAUT, 2015; Ross et al., 2014; Vrillon, 2017, p. 4; Yuan & Powell, 2013, p. 7; Zhu et al., 2020, p. 1689), where "the course relies more on the connections between learners rather than on the content they learn

[62] *Cf.* Relevant recent work on MOOC taxonomies (Blackmon & Major, 2017; Pilli & Admiraal, 2016; Rosselle et al., 2014; Stracke et al., 2019) were consulted while researching for this Chapter.
[63] Massachusetts Institute of Technology; a private research university in Cambridge, U.S.A. (MIT, n.d.)

together" (Rosselle et al., 2014, p. 2). Still, further distinctive features of xMOOC and cMOOC are surveyed by Admiraal, Huisman, & Pilli (2015), including the Student role, the Assessment, and/or the Interaction.

However, in her 2012 post, Lane (2012) plainly rejected the rigidity of this binary categorization and proposed instead a triad; recognizing the two already existing distinctions (xMOOC & cMOOC) but adding a third, intermediate category: Task-based MOOCs (tMOOC). In this new category, skill are emphasized "in the sense that they ask the learner to complete certain types of work" (2012). As such, tMOOCs find themselves composed of "a mix of instructivism and constructivism." (2012). She also privileged the fuller terms of 'Network-based MOOC' for xMOOC and 'Content-based MOOC' for cMOOC (Lane, 2012).

Furthermore, a 2013 study (Gilliot et al., 2013) saw another type of MOOC join the scene; the iMOOC, or 'investigative MOOC'. When considering the participant's point-of-view on the openness or closeness of a pre-defined set of five pedagogical dimensions (Jézégou, 2010) in a MOOC (*cf.* Table 3-1), they also noticed the need for an intermediary categorization between xMOOC (most dimensions closed) and cMOOC (most dimensions open), *i.e.*, the need for MOOCs mostly « […] *basés sur la résolution de problèmes ou sur la démarche par investigation »* (Gilliot et al., 2013, p. 1) [based on problem-solving or on the investigative approach].

Table 3-1

MOOC Categorization (Gilliot et al., 2013)

Pedagogical dimension	**cMOOC**	**iMOOC**	**xMOOC**
Learning goals	O	C	C
Choice of resources	O	O	C
Organization of the learning activities	O	C	C
Organization of the group's work	O	O	C/O
Collaborative co-production	O	C/O	C

Both these categorizations confirm the xMOOC – cMOOC continuum and the impending need for categorization flexibility and therefore extensibility, at least between these two major distinctions. Additional categorizations address equally important axes, such as the Pilli

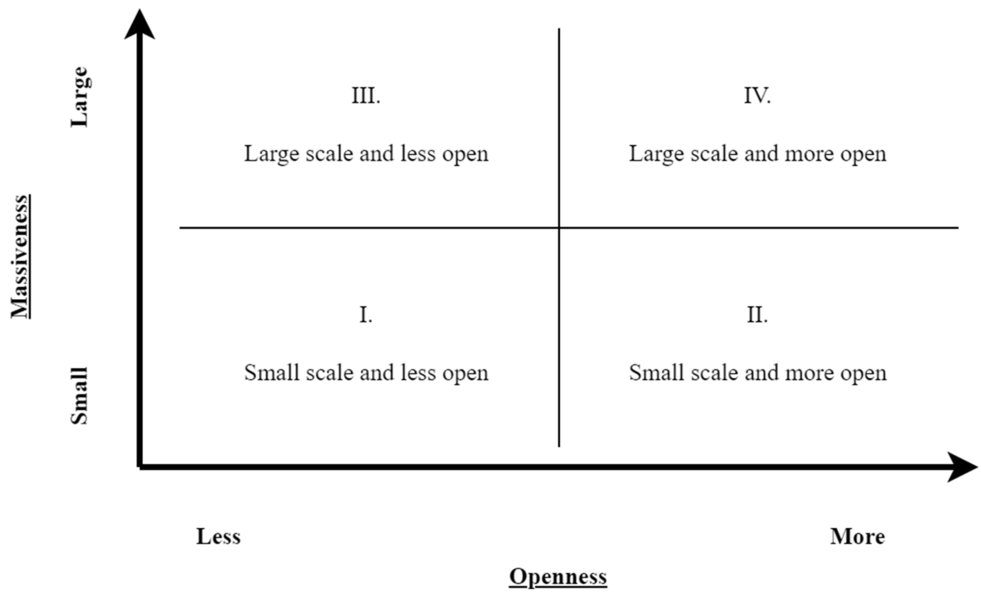
& Admiraal (2016) taxonomy, which is also devoted to the openness of a MOOC, *i.e.*, the possible "barriers to participation with regards to time, place, pace, adaptivity, accessibility and costs" (2016, p. 225), while also contemplating its massiveness, *i.e.*, its scale, size, or number of participants.

This publication (2016) conceived a two-dimensional Model for MOOC Taxonomy comprising four quadrants, determined by the Massiveness and Openness axes, ranging from Small to Large and from Less to More, respectively. Quadrants are numbered I, II, III, and IV, starting from the bottom left towards the top right. This model is illustrated in Figure 3-1 with the explanation of the quadrants' taxonomy following:

- I. Small scale and less open: The number of participants vary between 200-500, and likely a fee is required to access some parts of the course: a "typical, traditional online course" (2016, p. 227), *e.g.*, tMOOCs.
- II. Small scale and more open: Similar number of participants as in (1) but "course materials and/or exams are free to all participants" (2016, p. 226), *e.g.*, cMOOCs.
- III. Large scale and less open: Limitless participation but content is fee restricted, *e.g.*, traditional distance higher education, *e.g.*, Flipped-MOOCs.
- IV. Large scale and more open: Most-known type of MOOCs where there are no limits to content access nor to participants, *e.g.*, xMOOCs, iMOOCs.

Figure 3-1

Two-Dimensional Model for MOOC Taxonomy (Pilli & Admiraal, 2016)



Their study revealed the at-the-time trend where newer MOOCs mostly fell under the Small scale and more open quadrant (II), while noticeable for-profit MOOCs (opposite to pedagogical-oriented) rather fell into the Large scale and less open quadrant (III) (2016, p. 236).

We move on to the cartographical study performed by Vrillon (2017, pp. 11–16) on 195 French MOOCs from the *France Université Numérique* (FUN) platform during a three-year time span. This French study revealed eight[64] typical forms of MOOC, according to their covered discipline domain, pre-requisites, target public, estimated weekly effort, length, and certification, *i.e.*, the MOOC offer. Here we present a simplified view of the her analysis (Vrillon, 2017, pp. 11–16) where we highlight the main components[65] (as Vrillon herself highlighted) determining each form:

1. MOOC *spécialiste vs.* MOOC *profane*: First considered dimension, primarily axed on the prerequisites and target public variables, with the specialist MOOC being characterized by

[64] Four pairs of complementary forms determined by their positive *vs.* negative values linked on the corresponding dimension.
[65] Authors clarify that variables length and certification do not primarily contribute to their proposal.

requiring Superior education and targeted to Professional students while the layman MOOC is free of pre-requisites and is open to All public.

Table 3-2

First dimension of the Vrillon typology (Vrillon, 2017)

specialist MOOC	**layman MOOC**
pre-requisites: Superior education target public: Professional students	pre-requisites: None target public: All public

2. MOOC *généraliste d'approfondissement vs.* MOOC *spécialisé introductif*: Second considered dimension, axed on the discipline domain and the pre-requisites variables, with the in-depth, generalist MOOC being characterized by mostly targeting an All public with Secondary education whereas the specialized MOOC is aimed at Information technologies majors while not requiring any (besides the implied) pre-requisites.

Table 3-3

Second dimension of the Vrillon typology (Vrillon, 2017)

in-depth, generalist MOOC	**specialized MOOC**
discipline domain: Sciences pre-requisites: Secondary education	discipline domain: Information technologies pre-requisites: None

3. MOOC *spécialisé intermédiaire vs.* MOOC *spécialisé exploratoire*: Third considered dimension, axed on the weekly effort and discipline domain variables, with the intermediate, specialized MOOC defined by a two-to-four hours of estimated weekly effort, centered on Law, Economics and Management fields whereas the exploratory, specialized MOOC needs only less than two hours of estimated weekly effort, focused on the Health domain.

Table 3-4

Third dimension of the of the Vrillon typology (Vrillon, 2017)

intermediate, specialized MOOC	**exploratory, specialized MOOC**
weekly effort: 2-4 hours discipline domain: Law, Economics, Management	weekly effort: 0-2 hours discipline domain: Health

4. MOOC *spécialisé intensif vs.* MOOC *spécialisé intermédiaire spécialiste*: Fourth considered dimension, closely resemblant to the third dimension and axed on the same variables. It characterizes an intensive, specialized MOOC mostly by its estimated weekly effort of over four hours, focused on the Information technologies domain. On the opposite side of the spectrum, it also characterizes an intermediate, specialist, specialized MOOC requiring a medium (2-4 hours) amount of effort in the Health domain.

Table 3-5

Fourth dimension of the Vrillon typology (Vrillon, 2017)

intensive, specialized MOOC	**intermediate, specialist, specialized MOOC**
weekly effort: over 4 hours discipline domain: Information technologies	weekly effort: 2-4 hours discipline domain: Health

They concluded that in general, a MOOC from the « *plateforme nationale française* » FUN primarily targets a vast public without pre-requisites, it lasts from six to eight weeks, demands between two to four hours of weekly estimated effort, and mostly addresses "hard science" subjects (Vrillon, 2017, p. 19).

Besides, Ebner et al. (2020, p. 11) distinguish in their study of the "only MOOC platform in Austria" (2020, p. 4) seven types of MOOC, differentiating them on the learning/teaching scenarios happening behind the curtains. In their typology, a "face-to-face learning event" is primarily an offline, in-person, organized learning activity, although they admit the use of "online webinar tools to participate in these 'offline' sessions" (2020, p. 9).

1. The conventional MOOC is "a pure online course for many users" (2020, p. 11) and it is "characterized by thousands of learners worldwide" (2020, p. 6).

2. The Pre-MOOC is an online course serving as a preparatory activity for a subsequential face-to-face learning event.

3. A Blended-MOOC starts and closes with face-to-face learning events, with the online part being integrated to happen between other several face-to-face learning events, *i.e.*, there is an alternating exclusivity between the MOOC and the face-to-face learning events.
4. The In-Between-MOOC can be understood as a special case of the Blended-MOOC in which the MOOC happens only once between two face-to-face learning events.
5. The Inverse-Blended-MOOC is programmatically enriched by regular face-to-face meetings and learning events, not necessarily happening within a classroom setting, *i.e.*, the MOOC offers regular small groups, in-person reunions.
6. The Flipped-MOOC is an attempt to utilize the premise of a flipped classroom with a MOOC. It usually focuses on video watching and online material reading in parallel preparation to the face-to-face activities (like the 3. Blended-MOOC) which focus instead on discussing, training or knowledge application.
7. Finally, a Lecture-MOOC behaves like a Blended-MOOC but enriched by an accompanying Learning Management System within an educational organization. This organizational coupling allows for "additional non-public discussions and tests" (2020, p. 11).

This study showed that the reviewed Austrian MOOCs are usually divided into weekly sections (with each section featuring an assessment), they last in total from six to ten weeks, they are mostly based on video-watching but also offer additional learning content, with a discussion forum available for exchanges between lecturers and students, or students-only (2020, pp. 4–5).

The surveys behind these typologies reveal common MOOCs' real-world practical characteristics, in line with the main guideline presented at the beginning of this Chapter, such as their length (about eight weeks), weekly effort (about three hours), focus material (videos), subjects (usually "hard science"), exchange spaces (online forums), fees (generally open and free access to content but certifications cost), and access (generally addressed to all public but university-based MOOCs do enforce requirements, which in turn curtain access).

In this Chapter we introduced the MOOC along its basic, common characteristics, backed up by recent surveys on the subject. We also saw the MOOC history, from its origins to its current re-emerging role during the COVID-19 pandemic. We also stated the effort behind categorizing MOOCs, illustrated in a few typologies relevant to this research work.

This thesis uses a MOOC as a terrain for its experiments, which data, after being treated, will require the application of Machine Learning techniques to make sense thereof. The following Chapter takes a dive into the complex notion of Machine Learning, mainly its purposes and the vast amount of terminology employed therein, necessary to understand the experiments carried out in Chapter 7 below.

Chapter 4. Machine Learning

In this Chapter we cover the basics of Machine Learning, its definition, purpose, and scope in this research work. We quickly overview the various paradigms in which Machine Learning is classified and describe commonly used metrics for the tasks performed and described later in this thesis (*cf.* Section "Experiments"). This Chapter contains extracts from published articles by this thesis' author (Ramírez Luelmo, 2020, 2022; Ramírez Luelmo, El Mawas, & Heutte, 2022; Ramírez Luelmo et al., 2021c).

What is Machine Learning? According to Géron (2019) Machine Learning is "the science (and art) of programming computers so they can *learn from data*" (2019, p. 2).

But how can computers learn? A good parallel to recognizing this process is to make an analogy of a police dog tasked to find a missing person. Experienced Machine Learning readers might skip this oversimplified explanation of the complex workings of a supervised Machine Learning model training with noisy and unprocessed training data, deployed on an environment with also missing data, and directly continue to Section "Definitions and Purposes" below.

On with the oversimplified analogy: if we are to train a police dog for this task, we are to present the dog with a few examples of the smell of the person in question such as clothes or objects known to have been recently handled by the individual. After being presented with a few samples, the dog might create a profile on what the characteristic smell of the person is.

Once the dog has been trained on creating this smell profile for this specific person, other samples, objects, or locations can be presented to the canine and based on the dog's reaction, we would be able to notice if the missing person passed about, handled, or wore the presented sample. There it is: the police dog has been trained to successfully detect a person's scent on objects.

When described in such bold strokes, the procedure seems quite straightforward and problem free. However, we will see what this simple narrative entails and which few sections could derail it all, and while doing so, illustrate the analogy to situations analog to computers learning.

First of all, it is important to remark that this hypothetical smell profile is known to this specific one and only dog[66], based on the smell characteristics the dog alone chose to consider pertinent and relevant for humans[67], from the sample of humans the dog has encountered (and smelled) ever before[68], and it is entirely unbeknownst to us[69], who cannot enter the dog's head and see if this smell profile is composed only of traces of chemical compounds but also of the dog's own personal experiences, which might not make sense to us, humans, but somehow do to the dog, for instance.

For the dog's benefit, we can intuitively affirm that the more scent samples we present the dog, the most they will contribute to create an accurate profile for this specific person in the dog's brain, up to a point where the profile is sufficiently consolidated, and more scent samples would be redundant to the dog. Inversely, we can see that a profile composed of only one item presented to the dog is a poor start for the missing person's search because it obviously contains insufficient information, or more accurately said, contains too much general information to discern one and only one person who handled the item.

Additionally, it is plain to see that if we were to try to fool the dog and present items that were never handled by the missing person the dog might either disregard the memorization of the entirety of the item because of its strangeness to the array of already-presented items, or keep it and add disturbances to the smell profile (how can this person clearly smell like a child, an old

[66] Meaning that a hypothetical second dog, next to the first, equally competent, and subject to the same sampling, might not create the same identical smell profile.
[67] Male or female, old or young, healthy, or unhealthy, happy, sour, acrid, sweaty, or rancid, etc. for a lack of a better way to convey to the reader the specter of options available to the canine to detect and choose from.
[68] The dog has not met all the possible types of humans on earth as to know what characteristics define a human smell profile in its entirety. However, the dog knows what characteristics define a human smell for **most** of the humans the dog has smelled before.
[69] Or the police, in this analogy.

person, and a milking cow at the same time?), or keep it and wait for other items to come to clarify the intent behind this last one (is the missing person close/far when this other scent profile is detected?). Now, we can see the conundrum we (and the dog) face if we were to make an honest mistake and present the canine items that we were assured were trustworthy, but were later found out not to be: how do we convey to the dog the instruction to disregard what we previously presented? What if the dog cannot understand the instruction to forget? Do we replace the dog for this task?

Plus, as any police TV-show has taught us, presenting the dog with items that are temporarily distant from the present moment when the dog is being trained carries the risk of such items being deprived of information the dog can utilize effectively. Insisting on employing them would probably expose the dog's smell profile to information about the storage medium employed to preserve the item rather than the missing person's scent.

As man's best friend lacks complex verbal skills to express interrogatives and Boolean queries on the nature and purpose of the items presented, we realize that, assuming the dog's nose is sound and healthy, the combination of items used to identify the missing person during the dog training constitutes the area where most of the challenging decisions are found:

- How is the dog to understand what the task at hand is if we were to present 50 items related to the missing person and unbeknown to us, 50 items related to another specific person, and we insist that they all belong to the same person?
- What about if we were to present 50 items related to the missing person and unbeknown to us, 50 items related to many other people, and we insist that it is the same person?
- What about 50% missing person + 25% another missing person + 25% random people?
- What about 25% missing person + …? The list of similar interrogatives is never ending.
- How do we explain the canine to disregard the old box smell and focus on the faint traces still lingering (if any) from a year ago?

- How do we explain the dog to consider some evidence, disregard some other and to lightly consider other, if certain conditions are met?
- How do we explain the dog that although similar, the smell of siblings, ascendance, descendance, relatives in general, do not constitute the smell of the missing person yet deformed? Up to what point is the dog able to identify individuals from among a group of extremely close people's scents?
- The list of exceptions and considerations goes on.

Once the dog's training is complete, for the sake of thoroughness, we would test the police dog by presenting a known item own by the missing person, previously set aside for this purpose, as well as items known to be unrelated, and observe the dog's tail. Again, the validity of this test heavily depends on how sure we can be that both sets of items for testing are accurately what they claim to be: 100% related and 100% unrelated to the missing person. All prior assumptions validated, our detecting canine should be able to properly discern related and unrelated items and to fail, up to an acceptable degree, to no fault of their own, and still be considered a "good boy".

On the same subject, to insist on searching for a missing person in a place that it would eventually prove the person never was (and therefore failing to detect the person's scent) is a fact that not even the best-trained police dog can ever change. Likewise, presenting our best, high-quality trained dog with a field search on a $2000m^2$ selected area pointed out by trustworthy witnesses with all probability the person was last seen but since being entirely ravaged by a flood beyond recognition seems an unfair and impossible task for the canine. That is, the chances the dog detecting anything relevant at all are slim, independently whether smell proof ever was, passed by, still is, or traces were left there, as now in the best of cases, such scent proof is still there but covered and surrounded by other extremely intense odors, whereas in the worst scenario any remaining scent proof has been washed away.

Evidently, in both cases, we cannot accuse the dog of anything when failing to detect, wrongly or rightly so, any relevant scent. In the second, most complex case, we cannot come to any conclusion on the failure of our trained police dog to detect the missing person's scent: nor this is proof the person never passed by, nor this is proof the person passed by, got out and the scent is now destroyed, nor proof that nobody at all was there before the flood. That is, the operational conditions, whether wind, rain, snow, or more destructively fire, greatly affect the performance of any trained police dog on the task.

Yet, even if any clue relevant to the missing person was found by any other means, unrelated to the dog, still our trained canine companion cannot be put at fault for failing to detect it on that specific terrain. Most likely, we can then hypothesize the items presented to the dog did not contribute to painting an accurate smell profile of the missing person.

Finally, let us paint two slightly different scenarios to underline the role of our sniffing police dog:

First, what if we were to employ a flipping coin to determine if a recovered item belongs or not to the missing person, instead of our trained dog? By flipping a coin, we see that any correct guesses would happen simply by chance and not because there is a real relationship between the item and the missing person. Obviously, such a method (through random determination) would be entirely useless, and any police department would be out of its mind to try to implement it on the field. However, it can be used as a baseline for comparing the performance of our trained sniffing dog, *i.e.*, can our sniffing dog detect a smell profile on an item better than luck?

And second, what if we were to use a chimpanzee instead of a dog, as a smelling animal? As we all know, dogs are commonly selected for police service because of their innate smelling talents, their service attitude, and their communicational skills, all of which might not be found in a chimp, mainly the necessary sniffing talent. In this new scenario, if we were to try to train the newly acquired chimp on this same specific task (presenting accurate sample items related to a missing person's scent), we all should unequivocally conclude that, once on the field, any correct guess by the chimpanzee would be the result of either pure chance or owing to a fairly intense smell, simply because of its lacking skills (which make of the chimp not the best animal candidate for this task, unless the smell profile is evident even to the human police).

This blatantly obvious inefficiency should not be accredited to an incorrectly trained chimp, nor to a lack of work investment from the chimp's trainer, nor to the lacking funds of the police department. The chimp does not achieve an equivalent level of performance when trying to detect smell profiles simply because this is not a task a chimpanzee is best at.

Well, in this dog analogy, for some authors Machine Learning is the equivalent of the entire process just described; from picking up the most appropriate individual dog for the task, up to finding the missing person in the terrain. For other authors, Machine Learning is limited to the phase of presenting items to the police dog. Following the analogy, the items presented to the dog are the training data and thus, this phase is likewise called the training phase.

The police dog inherently performs better than a human at discerning scents. This inherent set of skills is the equivalent of the mathematical model behind the Machine Learning model: it knows how to do this one thing exceptionally well. The scent profile stored in the dog's mind, unique for that specific task, for that specific human, ready to perform, is the Machine Learning model trained: it is the general mathematical model although specialized for one task. The scent profile is formed by many components, which quantify or qualify (to the dog) in different scales the various aspects of the smell in question. Those elements are the features (or dimensions) of the Machine Learning model, and their specific quantifications for solving the task are also considered as a Machine Learning model.

Thus, when we talk of a Machine Learning model, without further clarification or context, literature might refer to (a) the mind of the untrained dog (the math behind), (b) the entirety of the trained dog (the portable, re-useable and useful component), or (c) the smell profile the trained dog created for that specific human (the weights and bias for the math to work in that fashion), *i.e.*, the list of quantities of each scent component the dog considers relevant.

The number of items presented to the dog during the training phase are analog to the training data sample size: usually the more the better, up to a saturation (or hyper definition) point. The various inconsistencies present in the training data sample refer to the quality of the training data sample: they can reflect the data's age and thus relevance to the phenomenon to be detected, its trustworthiness, but more damaging, the noise still lingering, unknown to us and which will play a role in the model's training and prediction phases. Just like in the police dog's analogy, the quality of the training data sample determines the success or failure of the entire process: it is entirely possible to create a Machine Learning model that successfully detects things that have no logical sense in real-world situations.

The acceptable degree our preciously trained dog can fail at detecting the missing person is analog to the agreed-upon values on the various metrics employed to test our Machine Learning model. Also, some metrics might carry a heightened importance depending on the nature of the task the model has to accomplish, and their gauge values vary if our trained model (the dog) has seen the items beforehand, during the training or testing phases, or never.

Our two final scenarios highlight the role of the sniffing dog (the trained model).

First, the flipping coin determination method is what is known in Machine Learning jargon as a random classifier, and it has a success rate of 50% (for binary classifications) on the metric AUC ROC (*cf.* Subsection "Receiver Operating Characteristic (ROC) Curve" 0 below). A random classifier is used as baseline to compare against Machine Learning models. If our trained dog does no better than the pure luck of wildly lucky guesses, we might have to start the whole process again, under different conditions.

Second, the choice of a sniffing dog for this sniffing task is determinant to its success: employing a Machine Learning algorithm (the math behind) that does not cover this specific kind of task would yield poor results, not matter the quantity, nor the quality of the training data. It is not that the chimpanzee would not be able to correctly determine some items, but a sniffing dog would certainly perform better at this task most of the time.

We conclude this long dog training analogy by listing in Table 4-1 the given names employed and their real corresponding terminology equivalents in Machine Learning.

Table 4-1

Police Dog Training Analogy to Machine Learning

Police dog training analogy	**Machine Learning notion**
The entirety of the process described above: from picking up the most appropriate individual dog for the task, up to finding the missing person in the terrain.	Machine Learning (for some authors)
The dog training phase.	Machine Learning (for other authors)
The presented items during dog training.	Training data
The presented items to the dog for testing the training.	Testing data
The never-before-seen items presented to the dog after testing.	Evaluation data
The number of items presented to the police dog	The training data sample size
The mind of the untrained dog.	Machine Learning model (the math behind)
The entirety of the trained dog.	Machine Learning model (the reusable and useful component)
The smell profile the trained dog created for that specific human.	Machine Learning model (the weights and biases for the math to work in that fashion, for that purpose)
The components of the smell profile (amount and names).	The features (or dimensions) of the Machine Learning model
The inconsistencies of the items presented to the police dog	The training data sample quality, age, trustworthiness, noise, etc.
The acceptable tolerance for failure of the trained dog	Various Machine Learning metrics
The flipping coin	A random classifier
The success rate of the coin	AUC ROC of 0.5 of a random classifier
The chimpanzee's mind	A Machine Learning model not suited to the required task

We hope this analogy helps non-initiated readers to understand the complexities incurred when dealing with Machine Learning models. It is intended to instill a general knowledge of what Machine Learning models can do and why they could fail or succeed at their designed task.

Then, we proceed to cover definitions and purposes of Machine Learning. We caution the reader that the following Sections are, despite our best efforts, heavily technical in nature. They cover many details of Machine Learning, *viz.* its different paradigms, its main terminology (of which we quickly covered the main notions in this oversimplified analogy with police dogs), its metrics, and many definitions and/or precisions on the related technical tasks and choices employed in this thesis. Feel free to skip to Subsection "Scope and importance" below, where we underline and justify the use of Machine Learning in diverse applications. Likewise, proceed to Chapter 5 below if the lecture of Sections "Machine Learning workflow" and below feels like a dictionary reading.

Definitions and Purposes

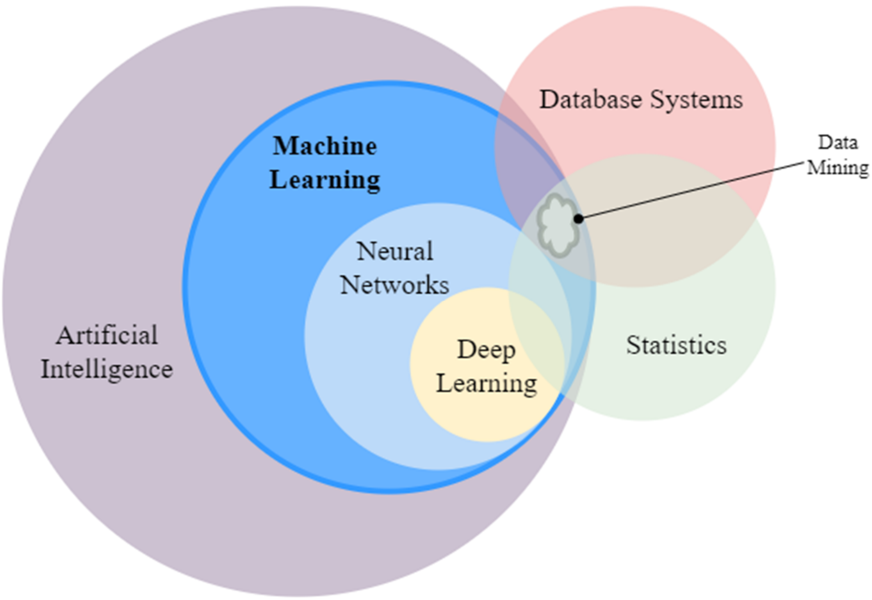
In this Section we present Machine Learning-related notions, starting with its own placement among other commonly used terms. We first dedicate a Subsection to cover its definition, surveyed among many others found in the literature, before settling on the one from Géron (2019, p. 2). In the ensuing Subsection we quickly overview the many terms employed in this thesis. We proceed by listing a few common tasks prevalent in Machine Learning relevant to this research work, and we conclude on the scope and relevance of its utilization.

We introduce this Section by placing the field of Machine Learning among other commonly used terms. Indeed, Machine Learning is often confused with other notions surrounding it such as Data Mining, and/or Artificial Intelligence. A very recent literature review by Mehta et al. (2019, pp. 115–116) acknowledges this confusion in a dedicated section, and authors attribute it mostly to the hype surrounding the term.

Thus, based on the many characterizations found in the reviewed publications (Chakrabarti et al., 2006; Dangeti, 2017; Das & Behera, 2017; IBM, 2020; Mehta et al., 2019, p. 4; Schmidhuber, 2015; Subasi, 2020, pp. 93–94; Tran, 2019), the literature review by this thesis' author (Ramírez Luelmo et al., 2021c) accurately places the field of Machine Learning at the crossroads of the fields of Database Systems and Statistics, while admitting the fields of Neural Networks and Deep Learning within itself. We present the corresponding Venn Diagram issued from that publication (Ramírez Luelmo et al., 2021c) in Figure 4-1 to illustrate the positioning of the field of Machine Learning (blue circle) against other commonly used terms.

Figure 4-1

Situational context of Machine Learning (Ramírez Luelmo et al., 2021c)



Main Definitions

The term Machine Learning has its origins in the works of Arthur Samuel (1959, 1967, 1969) by developing a computer program[70] which would learn to play checkers "better […] than can be played by the person who wrote the program" (1959, p. 211) when given only a limited set of information on the game rules and goals. In the original 1959 paper abstract, one can notice Arthur Samuel's content with the results, seemingly even surprising to him, the developer of the program:

> […] Furthermore, [the computer] can learn to do this in a remarkably short period of time (8 or 10 hours of machine-playing time) when given only the rules of the game, a sense of direction, and a redundant and incomplete list of parameters which are thought

[70] More specifically, a mini-Neural Network ("highly organized network"), limited at the time per the "data-handling ability and […]computational speed" (Samuel, 1959, p. 211).

> to have something to do with the game, but whose correct signs and relative weights are unknown and unspecified (1959, p. 211).

It is also to Samuel we own the first hint to a definition of Machine Learning: "[…] programming of a digital computer to behave in a way which, if done by human beings or animals, would be described as involving the process of learning […]", or further down the text: "Programming to learn from experience […]" (Samuel, 1959, p. 211), paving up the way for the automation of the task.

This singular hint is later reprised and understood by uncountable researchers[71] as the basis for the first formal definition of Machine Learning: "[Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed" (Géron, 2019, p. 2; supposedly quoting Samuel, 1959).

Fellow engineer readers might appreciate the following definition that reached us from Tom Mitchell (1997) (*italic* & **bold** emphasis are from the original text):

> ***Definition***: A computer program is said to **learn** from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*. (T. M. Mitchell, 1997, p. 2)

Furthermore, this same book also outlined the main concern of Machine Learning as the question of "how to construct computers programs that automatically improve with experience" (1997, p. xv).

Tran (2019), just like in the above mentioned book (1997), considers the past experience an important component in the definition of Machine Learning and, although he also views Machine Learning as a paradigm, him along Mehta et al. (2019, p. 4) effectively succeed to narrow its field of action: "The sole focus of this field is automatic learning methods" (2019, sec. 1), and

[71] This definition is widely attributed (over a thousand results on a strict text query in Google Scholar) to Arthur Lee Samuel as part of his 1959 paper (Samuel, 1959). However, we could not find this quote nor in the original 1959 paper (Samuel, 1959), nor in its official 1963 reprint (Samuel, 1963), nor in its second part ("II-Recent Progress") published a decade later (Samuel, 1969).

"[…] with the goal of developing algorithms capable of learning from data automatically" (2019, p. 4). Nevertheless, Tran cares to clarify that learning refers to the "modification or improvement of algorithm based on past *experiences* automatically without any external assistance from human" (2019, sec. 1).

On his part, Subasi (2020, p. 92) sees an "intelligence" component to Machine Learning: "Machine learning is the study of computer algorithms to help formulate accurate predictions and reactions in certain circumstances, or to act intelligently" (2020, p. 92). Accordingly, the International Business Machines Corporation frames the definition of Machine Learning within the field of Artificial Intelligence, considering Machine Learning as a subset or a branch "of Artificial Intelligence focused on building applications that learn from data and improve their accuracy over time without being programmed to do so" (IBM, 2020). These definitions, although vast and ambitious, in our opinion, leave aside the contextualization and intent behind employing Machine Learning.

In that matter, for authors Raschka & Mirjalili (2019, p. 1) Machine Learning is the "application and science of algorithms that make sense of data […]" (2019, p. 1). The clearly outlined intention of stating Machine Learning's purpose ("making sense of data") is appreciated and very much aligned with the point-of-view of researchers Deisenroth, Faisal, & Ong (2020) who, besides bringing up to the table the importance of the phase of algorithm design, declare that Machine Learning is "about designing algorithms that automatically extract valuable information from data" (2020, p. 11). Bisong (2019b, p. 169) shares a similar angle, incorporating the mathematical notions of methods and algorithms as well: "The set of methods and algorithms for discovering patterns in data is what is known as machine learning" (2019b, p. 169). Indeed, the importance of "making sense of data", "discovering patterns in data", or "extract[ing] valuable information from data" is undeniable and consistently hoisted as the main reason for employing Machine Learning (*cf.* "Scope and importance" below).

We open a parenthesis to remark that in the Tran (2019) survey, not including the purpose in their Machine Learning definition may indicate it was to be carried on (and linked to) elsewhere, such as in their own Data Mining concept; highlighting ‘knowledge’ as the result of this process: “Data mining can be defined as a method of extracting or ‘mining’ knowledge from large amount of data” (2019, sec. 1) and subsequently “[Data Mining] lays the foundation for machine learning” (2019, sec. 1). However, for Géron (2019, p. 5) this task is reversed, with Data Mining being the application of “Machine Learning techniques to dig into large amounts of data [to] discover patterns that were not immediately apparent” (2019, p. 5), while we insist that patterns *per se* do not unequivocally constitute knowledge. Closing parenthesis.

Still, Géron (2019, p. 2) goes a bit further and consider Machine Learning as “the science (and art) of programming computers so they can *learn from data*” (2019, p. 2) (parenthesis and *italic* emphasis are from the original quote). One cannot but to agree with this vision of Machine Learning and to welcome the hint of the many subtleties in decision-making involved when designing Machine Learning models: just like art, technique is necessary but there are times when one is to squint at the rulebook.

Google Developers (2022f) are less poetic and consider Machine Learning but a part of what previous authors saw as an ensemble of design phases. Still, they pragmatically agree on the importance of the intent behind Machine Learning’s utilization, joining the ranks of the authors behind (Deisenroth et al., 2020; Raschka & Mirjalili, 2019): “[Machine Learning] is the process of training a piece of software, called a model, to make useful predictions from data” (Google Developers, 2022f). Suitably, they also provide us with an clarification of what this model is: a representation of “the mathematical relationship between the elements of data that an [Machine Learning] system uses to make predictions” (Google Developers, 2022f).

Why not directly prompt an entity entirely concerned by Machine Learning? In January 2023, we asked the OpenAI's ChatGPT model[72] (2021) to define Machine Learning:

> Machine learning is a subfield of artificial intelligence (AI) that involves the development of algorithms that allow computer systems to learn from data, identify patterns and make predictions or decisions without being explicitly programmed to do so. The aim of machine learning is to enable computers to improve their performance on a specific task over time, by using algorithms that can automatically learn and adapt to new data (ChatGPT, 2021).

Which is a great example of how Machine Learning relies on the quality of training data to exhibit adequate performance, or in this case, a relevant answer properly assembled from previously shown text definitions. Indeed, Data is the next major component of Machine Learning.

(i) Data. As shown, elements of the definition consistently reoccur: "automatic", "learning", "data", "algorithm", "model". For authors Deisenroth et al. (2020), a Machine Learning system encompasses three major components or concepts: data, models, and learning (2020, p. 11), of which data is at the core, because "[…] Machine Learning is data driven", a resolute statement shared in some form or another by many other researchers (Amershi et al., 2019; Bisong, 2019a, p. 170; 2020, p. 11; Shirmohammadi & Al Osman, 2021, p. 85; Sugimura & Hartl, 2018, p. 2).

On our part, we could not agree more on the importance attributed to the notion of data. Ultimately, the adage GIGO (Churchill, 1979, p. 64; Jacoby, 1978, p. 90) ("Garbage In, Garbage Out"[73]) predominant in Computer Science, plays an even more crucial role in Machine Learning, where inherently (because of its automatic nature) no systematic user input verification can be performed.

[72] https://chat.openai.com/chat ; no citation knowingly exists at the time of querying the model.
[73] Flawed input data will produce flawed or nonsensical output.

Naturally, other contemporary authors (Kulkarni et al., 2020, p. 1) also concur on the data relevance when they point out that data quality is of the uttermost importance: "questionable quality [data] can introduce different types of biases in various stages of the data science lifecycle" (2020, p. 1). They further add that a widespread factor introducing bias is data imbalance, which simply means that "one of the classes has a higher percentage compared to the percentage of another class" (2020, p. 1). Imbalanced class distribution is a known concern when dealing with self-reports (Hussain et al., 2012, p. 81). Still, this complex issue can be tackled by using different sampling methods[74], among which SMOTE (Chawla et al., 2002) is a popular effective choice which do not incur in data loss, and of which ADASYN (He et al., 2008) is one of its many extensions (Kulkarni et al., 2020, p. 15).

Yet, despite its center role, data cannot bear the burden of what Machine Learning is. For instance, Géron (2019, p. 2) wonderfully explains the major, yet incomplete role of data in Machine Learning:

> If you just download a copy of Wikipedia, your computer has a lot more data, but it is not suddenly better at any task. Thus, downloading a copy of Wikipedia is not Machine Learning. (Géron, 2019, p. 2)

(ii) Model. In turn, we see the definition of a Machine Learning model as three-fold, mostly because, as authors Deisenroth et al. (2020, p. 251) point out, "the word 'model' has many subtleties" (2020, p. 251). Indeed, all authors cited in this Chapter 4 (but one) refer indistinctly as a "Machine Learning model" to any of the following first two similar, yet distinct notions. Also, we noticed this referral often varies according not only to the development stage (*cf.* Section

[74] Undersampling, Oversampling, and Hybrid methods (Kulkarni et al., 2020, p. 8).

"Machine Learning workflow" below) where the referral occurs but also to the complexity of the model being created. The third notion is less recurrent but existing nonetheless:

1. "A Machine Learning model" corresponds to the mathematical inner workings employed, *i.e.*, the algorithm(s) or method(s) (*cf.* Section "Machine Learning methods" below) applied. This meaning is often conveyed during the decision-making stage of the development workflow, *e.g.*, "Tree-Based Machine Learning Models", "*k*-nearest neighbors is a non-parametric machine learning model", "[…] to implement a linear regression machine learning model", etc.

2. "A Machine Learning model" is the instantiated algorithm or method *a.k.a.*, the trained model, *i.e.*, the actual numerical values contained in vectors, matrices, weights, etc. determining what and how the algorithm performs for the specific task for which it was trained. Accordingly, this meaning is usually carried out during the training stage.

3. We also noticed that such denomination equally befalls on the grouping, coupling, chaining, or linking of at least two of the previous notions (1 & 2) to other related processes, yielding a complex model in which the output of one model is the input of the next, yet confined within the same indissociable structure, *a.k.a.* a Machine Learning pipeline. This extension of the definition sometimes encompasses elements external to any Machine Learning processes, *e.g.*, the corresponding service running on a server, "[Machine Learning] trained models either as an online or batch prediction service".

Still, among all cited works in the present Chapter, only researchers Amershi et al. (2019, p. 9) clearly discerned notions 1 & 2: in one hand, a Machine Learning model is "the algorithm that powers the particular machine learning technique being used" (2019, p. 9), *e.g.*, Neural Networks, Support Vector Machines, Bayesian Networks, etc. In the other hand, it is "the set of parameters that controls the function" such as the weights or support vectors, which "are learned during training" (2019, p. 9).

We hope this much needed precision will help distinguish the various meanings the word 'model' carries in the field of Machine Learning and thus, Artificial Intelligence, and prevent even more confusion when used in other, unrelated fields.

(iii) Learning. For Bisong (2019a, p. 175) learning is "the ability to generalize to previously unseen samples" (2019a, p. 175). Authors Deisenroth et al. (2020, p. 11) boldly claim that the learning component of Machine Learning is "the crux of the matter" (2020, p. 12), and that it is to be understood as "a way to automatically find patterns and structure in data by optimizing the parameters of the model" (2020, p. 11). They further clarify: "The goal of learning is to find a model and its corresponding parameters such that the resulting predictor will perform well on unseen data" (2020, p. 257).

Both precisions point to the importance of [hyper-]parameter-finding and performance on yet-unseen data, which find their corollary on the subsequent assertion, drafted as one of the guiding principles of Machine Learning: "good models should perform well on unseen data" (2020, p. 251). This principle inevitably begs the question of what a good model is and how to objectively quantify such performance. Opportunely, authors also provide an exemplary explanation of what a good model is and the tool to use to calculate such performance:

> […] good models can also be thought of as simplified versions of the real (unknown) data-generating process […] A good model can then be used to predict what would happen in the real world without performing real-world experiments. (Deisenroth et al., 2020, p. 12)

And,

> It is not enough for the model to only fit the training data well, the predictor needs to perform well on unseen data. We simulate the behavior of our predictor on future unseen data using cross-validation. (Deisenroth et al., 2020, p. 257)

We conclude on a major distinction in the learning process: the design choice for data used in the modeling pipeline. When data is constantly being (produced and) fed in streams to the learning algorithm the process is called Online Learning (Bisong, 2019a, Chapter 15). Conversely, when data is "at rest" (*i.e.*, considered a "dataset") and available "at a certain point in time" (2019a, Chapter 15), the process is called Offline Learning, *a.k.a.* Batch Learning (2019a, Chapter 15).

A major consequence of this design choice befalls on the model updating: if new data is available (and the need to update the model), during Offline Learning the entirety of the already-trained model is to be trained all over again using the original (old) examples plus the new examples. Contrarily, this situation is already accounted for during Online Learning, where data streams are continuously being generated and sent into the training process (Bisong, 2019a, Chapter 15).

Terminology

Authors Raschka & Mirjalili (2019, p. 11) list in their book other useful terms (and their synonyms) commonly used in Machine Learning while highlighting that "many terms and concepts have been rediscovered or redefined […]" (2019, p. 11). We quote and reprise these definitions literally for their straight-to-the-point simplicity and comprehensiveness, particularly in the synonyms area, albeit sorted differently from their source (2019, p. 11), beginning with those most frequently employed in this thesis (*italic* emphasis from the original text).

- Feature, abbrev. *x*: A column in a data table or data (design) matrix. Synonymous with predictor, variable, input, attribute, or covariate.
- Target, abbrev. *y*. Synonymous with outcome, output, response variable, dependent variable, (class) label, and ground truth.
- Training: Model fitting, for parametric models similar to parameter estimation.

- Training example: A row in a table representing the dataset and synonymous with an observation, record, instance, or sample (in most contexts, sample refers to a collection of training examples).
- Loss function: Often used synonymously with a *cost* function. Sometimes the loss function is also called an *error* function. In some literature, the term “loss” refers to the loss measured for a single data point, and the cost is a measurement that computes the loss (averaged or summed) over the entire dataset. (Raschka & Mirjalili, 2019, p. 11)
- Overfitting means that the model captures the patterns in the training data well but fails to generalize well to unseen data (test data) (2019, pp. 58, 75). If a model suffers from overfitting, […] the model has a high variance, which can be caused by having too many parameters, leading to a model that is too complex, given the underlying data. Similarly, […] underfitting […] means that our model is not complex enough to capture the pattern in the training data well and therefore also suffers from low performance on unseen data. (Raschka & Mirjalili, 2019, p. 75)

Concerning the very important notion of overfitting, the previously cited publication points to the bias-variance trade-off as the link between overfitting and underfitting: “In general, we might say that ‘high variance’ is proportional to overfitting and ‘high bias’ is proportional to underfitting” (2019, p. 76). On the bias *vs.* variance trade-off Dangetti (2017) concludes: “The ideal model will have both low bias and low variance” (2017).

On his part, Ratner (2017, p. 501) defines an overfitted model as “one that approaches reproducing the training data on which the model is built—by capitalizing on the idiosyncrasies of the training data” (2017, p. 501), *i.e.*, the model has memorized the training data instead of capturing the underlining pattern in the data. For Brownlee (2020c), if “the performance of the model on the training dataset is significantly better than the performance on the test dataset, then the model may have overfit the training dataset” (2020c) and he suggests to watch out for a specific pattern when training a model: "[if] model performance on the training dataset continues to improve […] and performance on the test or validation set improves to a point and then begins to get worse […] then training should stop at that point, where performance gets worse on the test or validation set for algorithms that learn incrementally" (2020c). This point of view is also shared by Muhamedyev (2015, p. 23), of which the corollary is of our choosing:

> The correct balance between underfit and overfit means the search of an algorithm and its parameters, which would be able to show consistent results for a testing set (or a cross validation set). An underfit algorithm will show equally inconsistent results both for test and train sets, while an overfit algorithm will demonstrate a high result for a train set and a low one for a testing set. (Muhamedyev, 2015, p. 23)

We complement this list with a simplified account of very much needed definitions from a different source (Mohri et al., 2018, sec. 1.4). On a side note, we direct the reader's attention to the definitions of 'feature' (a column) and 'features' (a set) to point out their complementary nature, and that the plural *vs.* singular choice befalls on their authors' preference:

- Examples: Items or instances of data used for learning or evaluation. […]
- Features: The set of attributes, often represented as a vector, associated to an example. […]
- Labels: Values or categories assigned to examples. In classification problems, examples are assigned specific categories [… while] in regression, items are assigned real-valued labels.
- Hyperparameters: Free parameters that are not determined by the learning algorithm, but rather specified as inputs to the learning algorithm.
- Training sample: Examples used to train a learning algorithm. […]

- Validation sample: Examples used to tune the parameters of a learning algorithm when working with labeled data. The validation sample is used to select appropriate values for the learning algorithm's free parameters (hyperparameters).
- Test sample: Examples used to evaluate the performance of a learning algorithm. The test sample is separate from the training and validation data and is not made available in the learning stage. […]. (Mohri et al., 2018, sec. 1.4)

The above definitions of Train, Validation, and Test Datasets are in line with those portrayed in previously mentioned publications (Bisong, 2019a, p. 176; Brownlee, 2020b; Dangeti, 2017) although they differ from another set of articles (Deisenroth et al., 2020, p. 262; Mohri et al., 2018, sec. 4.5; Subasi, 2020, p. 105), where only training and test sets are contemplated. Still, all publications agree that one is to reserve a minor set of data to test or evaluate the model that has not been used before hand during training nor hyperparameter tuning, if this phase exists.

For the sake of completeness, we also quote the various "fundamentals" highlighted by Dangeti (2017):

- Population: This is the totality, the complete list of observations, or all the data points about the subject under study.
- Sample: A sample is a subset of a population, usually a small portion of the population that is being analyzed.
- Mean: This is a simple arithmetic average, which is computed by taking the aggregated sum of values divided by a count of those values. The mean is sensitive to outliers in the data. An outlier is the value of a set or column that is highly deviant from the many other values in the same data; it usually has very high or low values.
- Median: This is the midpoint of the data, and is calculated by either arranging it in ascending or descending order. If there are N observations.

- Mode: This is the most repetitive data point in the data.
- Range: This is the difference between the maximum and minimum of the value.
- Variance: This is the mean of squared deviations from the mean (x_i = data points, μ = mean of the data, N = number of data points). The dimension of variance is the square of the actual values.
- Standard deviation: This is the square root of variance.
- ANOVA: Analyzing variance tests the hypothesis that the means of two or more populations are equal. ANOVAs assess the importance of one or more factors by comparing the response variable means at the different factor levels. The null hypothesis states that all population means are equal while the alternative hypothesis states that at least one is different. (Dangeti, 2017)
- Imbalanced dataset: [It is defined as] a dataset with unequal class distribution (Fernández et al., 2018, p. 19; Chawla et al., 2002, p. 1), [*i.e.*,] the class which has majority instances is considered as a majority class or a negative class, and the underrepresented class is viewed as a minority class or a positive class (Chawla et al., 2002). Or in simpler terms “a difference in the numbers of positive and negative instances, usually with the negatives outnumbering the positives” (Saito & Rehmsmeier, 2015, p. 2).

Related tasks

Furthermore, authors Mohri et al. (2018, sec. 1.3) list “some standard” Machine Learning tasks, which exist to solve a specific problem, which in turn defines them. Bear in mind this list is not exhaustive[75], and it is presented to help contextualize and review the undertaking performed in this thesis:

[75] Other standard tasks include anomaly detection, structured annotation, translation, density estimation, etc.

- Classification: Its purpose is to assign a category to each proposed item, *e.g.*, a document classification tasks consists of assigning a category such as politics, business, sports, or weather to each document (2018, sec. 1.3). Usually, the number of categories prevalent is "less than a few hundreds, but it can be much larger in some difficult tasks and even unbounded, as in [Optical Character Recognition], text classification, or speech recognition" (2018, sec. 1.3).
- Regression: Its purpose is to predict a real value for each proposed item, *e.g.*, stock values prediction. In regression, the "penalty for an incorrect prediction depends on the magnitude of the difference between the true and predicted values, in contrast with the classification problem, where there is typically no notion of closeness between various categories" (2018, sec. 1.3).
- Ranking: Its purpose is to order (sort) items according to some criterion, *e.g.*, during a web search, sorting web pages results by their relevance to a search query.
- Clustering: It aims to "partition a set of items into homogeneous subsets" (2018, sec. 1.3). It is often employed to analyze very large datasets, *e.g.*, during social network analysis, clustering algorithms "attempt to identify natural communities within large groups of people" (2018, sec. 1.3).
- Dimensionality reduction or manifold learning: It consists of "transforming an initial representation of items into a lower-dimensional representation while preserving some properties of the initial representation" (2018, sec. 1.3).
- Variable and Feature Selection, considered as "constructing and selecting subsets of features that are useful to build a good predictor" (Guyon & Elisseeff, 2003, p. 1158) with the potential benefits of "facilitating data visualization and data understanding, reducing the measurement and storage requirements, reducing training and utilization

times, defying the curse of dimensionality to improve prediction performance" (2003, p. 1158).

Finally, we present the peripheral notions of GridSearch and the importance of estimating and reducing both the energy used and the emissions produced by training Machine Learning models.

Although not a Machine Learning task *per se*, GridSearch is a very important process in Machine Learning, closely linked to the notion of Model Tuning explained by Bisong (2019a, p. 302) as:

> Each machine learning model has a set of options or configurations that can be tuned to optimize the model when fitting to data. These configurations are called **hyperparameters**. Hence, for each hyper-parameter, there exist a range of values that can be chosen. (Bisong, 2019a, p. 302)

Now, GridSearch is an approach for parameter search, more specifically the hyperparameters of an estimator (Dangeti, 2017; Pedregosa et al., 2011, sec. 3.2). For a set of given values, GridSearch "exhaustively considers all parameter combinations" (2011, sec. 3.2), "explores all the specified hyper-parameter values for an estimator" (Bisong, 2019a, p. 302) to "find the best combination for determining the best fit" (Dangeti, 2017).

According to the corresponding documentation of the scikit-learn project (Pedregosa et al., 2011, sec. 3.2) such a search consists of "an estimator (regressor or classifier), a parameter space, a method for searching or sampling candidates, a cross-validation scheme, and a score function" (2011, sec. 3.2). Examples of a classifier are shown in Chapter 4(i) below, instances of typical Cross-validation methods are to be found in Subsection "Cross-validation & data splits" below, and finally, relevant scoring metrics are described in Subsection "Classification metrics" below.

In a seemingly unrelated but of utmost importance note, we approach the notion of green information technology, which at the very least is "an effort to reduce overall waste or limit consumption" (Hignite, 2009). Indeed, since almost two decades, the Hignite (2009) article argued that "computing, data processing, and electronic file storage collectively account for a significant and growing share of energy consumption […]" (2009). This concern has only intensified, considering the costs of electricity required to train AI algorithms (Kirkpatrick, 2023, p. 17). In the advent of the imperious "Carbon Bombs" (Kühne et al., 2022) and their massive effect on Global Warming (Houghton, 2005; Kerr, 2007), greater emphasis is placed on tracking and reducing (Anderson et al., 2009) "emissions to the atmosphere of large amounts of 'greenhouse gases', of which the most important is carbon dioxide" (2005).

More specific to the task at hand, authors of a recent preprint article (Lottick et al., 2019, p. 1) insist that the "carbon footprint of algorithms must be measured and transparently reported" (2019, p. 1) and invite the scientific community to "take an honest and active role in environmental sustainability" (2019, p. 1). Otherwise, and for cases not considered in this research project on flow, excessive electricity consumption linked to AI activities in datacenters[76] might jeopardize the reliability of any power grid (Chien, 2023, p. 5). Mostly based on Green AI and Deep Learning studies respectively (Schwartz et al., 2020; Strubell et al., 2019), authors of the preprint article (Lottick et al., 2019, p. 2) acknowledge the reporting of energy usage and CO_2 emissions to be the "gold standard", which we obviously reprise in this thesis.

Under that light, carbon dioxide (CO_2) emissions tracking is a method to estimate the carbon footprint of Machine Learning models when drawing computing power, *viz.* during their training. This estimation is measured as kilograms of CO_2-equivalents (KgCO_2eq or Kg. eq. CO_2), a standardized measure used to express the global warming potential (as electricity is generated by different means, one of which is by combusting fossil fuels) of various greenhouse gases, *i.e.*, the amount of CO_2 that would have the equivalent global warming impact (Schmidt et al., 2021).

[76] Authors specifically point out "Generative AI capabilities and applications" (Chien, 2023, p. 5).

Software like energy-usage[77] (Lottick et al., 2019), eco2AI[78] (Budennyy et al., 2023), or Tracarbon[79] (Valeye, 2022) monitor CPU & GPU usage during Machine Learning tasks to estimate the kilograms of CO_2-equivalents attributed to that task, adjusted by world region[80]. More specific to our study, project energy-usage emerged and evolved into the initiative CodeCarbon[81] (Schmidt et al., 2021) which goal is to approximate the CO_2 produced while running computer code by estimating the "hardware electricity power consumption (GPU + CPU + RAM)" and by transposing this information to the "carbon intensity of the region where the computing is done" (2021).

Indeed, these reporting endeavors are themselves aligned to the broader effort of holding AI research to a deeper academic scrutiny by calculating and reporting "the number of floating point operations (FLOP) performed by computer chips when training a Machine Learning model" (Sevilla et al., 2023, p. 30). Such major step would support experiments' reproducibility in research contexts, facilitate comparison between models, contextualize computational power between individuals, institutions, and eventually global entities, facilitate comprehensions between scale and performance, regulate the deployment of AI systems, and finally extrapolate results to make educated guesses on new AI capabilities (Sevilla et al., 2023, p. 31).

Nevertheless, it is currently acknowledged that there are difficulties when estimating FLOP numbers among which the absence of standards for their reporting and publication (Sevilla et al., 2023, p. 31) is at the top of the list.

Scope and importance

The relevance of Machine Learning (and thus, Artificial Intelligence) is undeniable in the 21st century, with broad applications in day-to-day life situations, ranging from groceries'

---

77 https://github.com/responsibleproblemsolving/energy-usage
78 https://github.com/sb-ai-lab/Eco2AI
79 https://github.com/fvaleye/tracarbon
80 Different world regions generate electricity in different enegy mixes, with wildly different proportion of fossil-fuels-based primary sources.
81 https://github.com/mlco2/codecarbon

stocking in your closest supermarket to disease detection assistance. In that context, and facing more traditional methods, Géron (2019, p. 5) summarizes the cases where Machine Learning constitutes an approach worth reviewing:

- Problem of which the solutions currently require excessive manual fine-tuning or long lists of rules.
- Problems of such complexity[82] that no traditional approach can yet yield a good-enough solution.
- Everchanging environments that require constant adaptation of the program in charge.
- "Getting insights about complex problems and large amounts of data" (2019, p. 5), *i.e.*, understand the problem better.

Subasi (2020, p. 92) argues that real-world problems benefit from training machines because machines can be "more efficient at saving energy, time, and resources […]" (2020, p. 93), more specifically in cases such as:

1. Human expertise on a given domain is unavailable or lacking.
2. If human expertise is not lacking then such expertise might be difficult to explain, or convey into computational tasks, *e.g.*, speech or image recognition, translation, cognitive tasks, among many others.
3. Addressing issues on a massive scale with data comprising dynamic requirements and restrictions.
4. Scenarios and behavior continuously change.

Typical real-world applications for this include healthcare data analysis, product recommendation (ecommerce), speech, object and image recognition, sentiment and emotion

[82] "[…] ML systems are often used for the most complex and ill-defined tasks—if the tasks were easy, we would not need an ML solution" (Isbell et al., 2023, p. 37).

analysis, content recommendation, click-through predictions, among many others (2020, pp. 94–95). This multivalence can mostly be explained by the flexibility offered by Machine Learning, by not assuming the underlying shape or model behind the data provided and therefore, learning these complex patterns automatically (Dangeti, 2017, p. 10).

For instance, the applications of Machine Learning in health care are expandingly rapidly, improving diagnostic accuracy, patient safety, monitor disease progression and many more, according to the latest literature review on the topic (Thomas et al., 2021, p. 1). Furthermore, an extensive study (Kumar & Chong, 2018) evaluates various Machine Learning techniques at predicting depressive disorders and other emotion states via atmospheric features, such as temperature, wind speed, ozone levels, or visibility, alongside assessment questionnaires (2018).

Nevertheless, just as its parent Artificial Intelligence, Machine Learning might suffer from an explainability issue (The Royal Society, 2019), producing highly accurate yet extremely complex models promptly called "black box" models (Rudin, 2019), which are so "complicated for even expert users to fully understand" (The Royal Society, 2019, p. 8). Otherwise plainly put by IBM: "[…] not even the engineers or data scientists who create the algorithm can understand or explain what exactly is happening inside them or how the AI algorithm arrived at a specific result" (IBM, 2022).

Fortunately, public policy calls for "some form of AI explainability […] into the design and deployment of AI-enabled systems" (The Royal Society, 2019, p. 8), and public opinion and mass media begin acknowledging the issue (Perota et al., 2023). Indeed, Machine Learning models' explainability is part the broader Explainable Artificial Intelligence (XAI) (Sanneman, 2023, p. 52) challenge: "AI systems that can explain their rationale to a human user, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future" (Sanneman, 2023, p. 53). One possible solution to this challenge crystallized in the Situation Awareness Framework for Explainable AI (SAFE-AI), comprising three levels of AI explainability: XAI for perception, XAI for comprehension, and XAI for projection (Sanneman, 2023, p. 54).

Thus. in this thesis project, we fully endorse Machine Learning models' explainability through a white box approach (Kompare et al., 1994), *a.k.a.* transparent, glass, or white-box.

Undeniably, to avoid having "people become the objects of prediction in a pipeline" (Chancellor, 2023, p. 78) we also intend to develop this research project within the framework of human-centered machine learning (HCML), with a heightened focus on "fair and transparent algorithm design, and human-in-the-loop decision-making […]" (2023, p. 78), without treating Machine Learning as a silver bullet *a.k.a.* "technological solutionism" (2023; Morozov, 2013).

More specific to our educational context, Machine Learning techniques help in modelling students' states and abilities, as shown by the studies on self-efficacy using Bayesian Networks (Mavrikis, 2010), on engagement detection in image recognition (Monkaresi et al., 2016), or those on students' similar interaction behavior clustering (Fratamico et al., 2017).

Thus, the role of Machine Learning is only but beginning, with ample opportunity for solving difficult-to-tackle problems, especially in areas where human expertise might be difficult to transpose into source code (Dangeti, 2017, p. 10), or where human expertise cannot be easily propagated over regional frontiers, as previously pointed out by Subasi (2020, p. 93). In an era where data is plentiful (Mehta et al., 2019, sec. 1.2), Machine Learning might make all the difference in assisting humans with otherwise difficult decision-making, if enough importance is given to the relevant data processes.

We conclude this Subsection with what we consider an effective synopsis on the importance of Machine Learning obtained from the Conati et al. (2018, p. 22) article when addressing the issue of learning from data:

> Machine Learning techniques are instrumental […] because they can help learn from data the knowledge and models that might be challenging to obtain from human experts and

compute predictions of students' cognitive and mental states in highly dimensional and ill-defined spaces of human behaviors. (Conati et al., 2018, p. 22)

In this Section we presented many Machine Learning-related notions, *viz.* its own definition (Géron, 2019, p. 2) (a) heavily dependent on the notions of data, model, and learning, but also many other terms[83] (b) employed in this thesis. We listed common tasks prevalent in Machine Learning relevant to this research work (c), and we concluded on the general scope and relevance of its utilization (d).

Machine Learning workflow

In this Section we present the Machine Learning workflow selected (Amershi et al., 2019) for this thesis project, its design stages, and details on how to approach them. Most importantly, we gathered a comprehensive and updated list of best practices issued from the conclusions of two literature reviews on the subject, which we quote as-is, and which we believe to be of great interest for Machine Learning beginners and experts.

An extensive literature review (Lorenzoni et al., 2021, sec. IV) presents two finalists works (Amershi et al., 2019; de Souza Nascimento et al., 2019) proposing relevant Machine Learning workflows. Among these two, the Amershi et al. (2019) works represent the "most comprehensive and accepted Machine Learning workflow" found among the 563 papers reviewed. Authors behind the same literature review (Lorenzoni et al., 2021, sec. IV) group the nine stages of the comprehensive Amershi et al. (2019, p. 2) Machine Learning Model Development into six stages:

- A Model requirements stage which is related to the agreement between stakeholders and the way the model should work.

[83] *Cf.* https://scikit-learn.org/stable/glossary.html for a comprehensive glossary of terms and API elements.

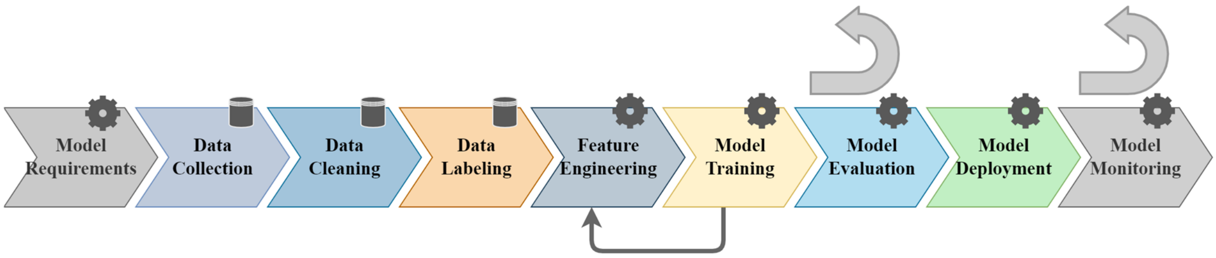
- Data processing stage which involves data collection, cleaning and labelling (in case of supervised learning).
- Feature engineering stage which involves the modification of the selected data.
- Model training stage which is related to the way the selected model is trained and tuned on the (labeled) data.
- Model evaluation stage which regards to the measurements used in order to evaluate the model.
- Model deployment stage which includes deploying, monitoring and maintaining the model. (Lorenzoni et al., 2021, sec. IV)

The original nine stages (Amershi et al., 2019, p. 2) are detailed in Table 4-2 and a linear representation of their Machine Learning workflow is shown in Figure 4-2 illustrating the successive nine stages and their feedback loops: the two larger U-turn feedback arrows mean the loop can go back to any of the previous stages, while the small feedback arrow only loops to the previous stage, *e.g.*, Model Training may loop back only to the Feature Engineering stage.

Furthermore, these comprehensive works (2019, pp. 2–3) make the distinction between data-oriented workflow stages; focused on the collection, cleaning and labeling of data (represented by a database icon hovering the stage in Figure 4-2), and model-oriented workflow stages; concentrating on the model requirements, the feature engineering, and the model training, evaluation, deployment, and monitoring (represented by a gear icon hovering the stage in Figure 4-2).

Figure 4-2

The Nine Stages of the Machine Learning Workflow (Amershi et al., 2019, p. 2)



It is important to underline that the individual components of these nine stages (2019, p. 2) are still somewhat represented in other workflows considered, such as the four stages in de Souza Nascimento et al. publication (2019, sec. IV), the six steps in Dangeti (2017, p. 11), and even the previously shown above six grouped stages shown in the Lorenzoni et al. (2021, sec. IV) publication, *i.e.*, choosing one workflow over another makes little difference as they all comprehend the same phases.

Table 4-2

The Nine Stages of the Machine Learning Workflow (Amershi et al., 2019, pp. 2–3)

Stage	Description
Model Requirements	Designers decide which features are feasible to implement with machine learning and which can be useful for a given existing product or for a new one. Most importantly, in this stage, they also decide what types of models are most appropriate for the given problem.
Data Collection	Teams look for and integrate available datasets, (*e.g.*, internal or open source) or collect their own.
Data Cleaning	Involves removing inaccurate or noisy records from the dataset, a common activity to all forms of data science.
Data Labeling	[…] ground truth labels to each record [is assigned …]. Most of the supervised learning techniques require labels to be able to induce a model. Other techniques (*e.g.*, reinforcement learning) use demonstration data or environment rewards to adjust their policies. Labels can be provided either by engineers themselves, domain experts, or by crowd workers in online crowd-sourcing platforms.
Feature Engineering	Refers to all activities that are performed to extract and select informative features for machine learning models.
Model Training	During model training, the chosen models (using the selected features) are trained and tuned on the clean, collected data and their respective labels.
Model Evaluation	Engineers evaluate the output model on tested or safeguard datasets using pre-defined metrics. For critical domains, this stage might also involve extensive human evaluation.
Model Deployment	The inference code of the model is […] deployed on the targeted device(s) […]

Model Monitoring	The inference code of the model is […] continuously monitored for possible errors during real-world execution.

Surveys carried out (Amershi et al., 2019) to Microsoft® experts reveal many relevant points (best practices) to keep in mind while designing and implementing a Machine Learning workflow. It would be a pity not to list a few of the most relevant to our research work:

- The stages of Data Availability, Collection, Cleaning, and Management are consistently ranked among the top challenges to face in Machine Learning.
- Often, the success of Machine Learning projects depends on Data Availability, Quality, and Management (Polyzotis et al., 2017).
- The panorama on platforms and techniques is constantly changing, which is a challenge when building new applications.
- Visual tools may be helpful with beginning data scientist but once they become proficient, these tools might "get in the way".
- Machine Learning tasks involve frequent iterations over the selected model, its hyper-parameters, and recurrent dataset refinement.
- Labeling datasets is costly and time consuming. Also, once labeled, it is key to try and reuse the data as much as possible to reduce duplicated effort.
- Datasets rarely have explicit schema definitions describing the columns and/or characterizing their statistical distributions. A given data set may contain data from several different schema regimes.
- Revisiting the modeling choices (stage one) is possible if the problem itself evolves or if a more appropriate algorithms emerges.
- Any workflow may become more complex if the system contains multiple Machine Learning components interacting together in likely complex and unexpected ways.

- It is important to develop a "rock solid, data pipeline, capable of continuously loading and massaging data, enabling engineers to try out many permutations of AI algorithms with different hyper-parameters without hassle" (Amershi et al., 2019, p. 5).
- Developing and providing "rich dashboards" to users showcase the usefulness and the added value of Machine Learning.
- If there is a large distribution shift between the training data and the real-world data, it might be prudent to go back and collect more representative data before running workflow again.
- Rigorous data versioning and sharing techniques are a must: resulting models are to be detailed with what dataset it has been trained with and with what untrained version model.
- Automating testing is as important in Machine Learning as it is in software engineering, without leaving the human outside of the loop.
- Compliance with corresponding principles on fairness, accountability, transparency, and ethics is non optional.
- Learned models' modularity is difficult for two reasons. First, models are not obviously extensible, *i.e.*, one cannot expect to couple a classifying model for email spam detection to a separate classifying model for virus detection and obtain a working result: both models would have to be developed and trained together. Second, models interact in non-obvious ways, *i.e.*, changes to one model will certainly affect the other's training and tuning processes, a phenomenon known as component entanglement.
- Significant changes to the resulting Machine Learning model might be needed when it is to run on a different domain (than the one employed during its training), or when using a slightly different input format. Any of these circumstances will likely require retraining the

model with additional, yet-to-be discovered, collected, and cleaned training data. Such retraining may take as much work and expertise as the original model initially took[84].

Moreover, Subasi (2020, pp. 95–96) reviewed a set of emerging challenges in Machine Learning, of which we quote a few:

- Problems with data quality lead to problems with data processing and extraction of features.
- Data acquisition, processing, and retrieval are procedures that are very tedious and time consuming.
- There is a lack of high-quality and sufficient training data in many scenarios.
- Feature extraction, particularly hand-crafting features, is one of the most difficult tasks in machine learning. Recently, deep learning seems to have gained some value in this area.
- […]
- Overfitting and underfitting models may lead to poor quality of the model learning configurations and relationships from the training data, leading to detrimental performance.
- The curse of dimensionality can be a real challenge, that is, too many features.
- It is not easy to implement complex models in the real world .(Subasi, 2020, pp. 95–96)

In this Section we covered in detail the design stages of the Machine Learning workflow (Amershi et al., 2019) to be employed in this thesis project. We also presented a comprehensive, useful, and updated list of best practices quoted from the state of the art on the subject, which we believe to be of great interest for Machine Learning beginners and experts.

---

[84] Very recent research suggest the creation of "tools and research advances that will allow pretrained models to be built in the same way that we build open source software" to approach this phenomenon (Raffel, 2023, p. 38).

Machine Learning methods

In this Section we quickly overview a primary classification for Machine Learning methods. Among these, we focus on the specific case of Supervised learning and that of Logistic Regression, both relevant to this thesis project.

Although some authors (Brownlee, 2020a; Das & Behera, 2017; Mohri et al., 2018, sec. 1.5; Sah, 2020, secs. 2, 3, 4) admit or survey several Machine Learning methods (*a.k.a.* styles or paradigms or scenarios or types), we retain the following more straightforward primary categorization, as more or less shared by many experts on the field (Bisong, 2019a, p. 171; Dangeti, 2017, p. 9; Géron, 2019, pp. 7–23; IBM, 2020; Muhamedyev, 2015, p. 15; Raschka & Mirjalili, 2019, p. 2):

- Supervised Machine Learning.
- Semi Supervised Machine Learning.
- Unsupervised Machine Learning.
- Reinforcement Learning.
- Deep Learning.

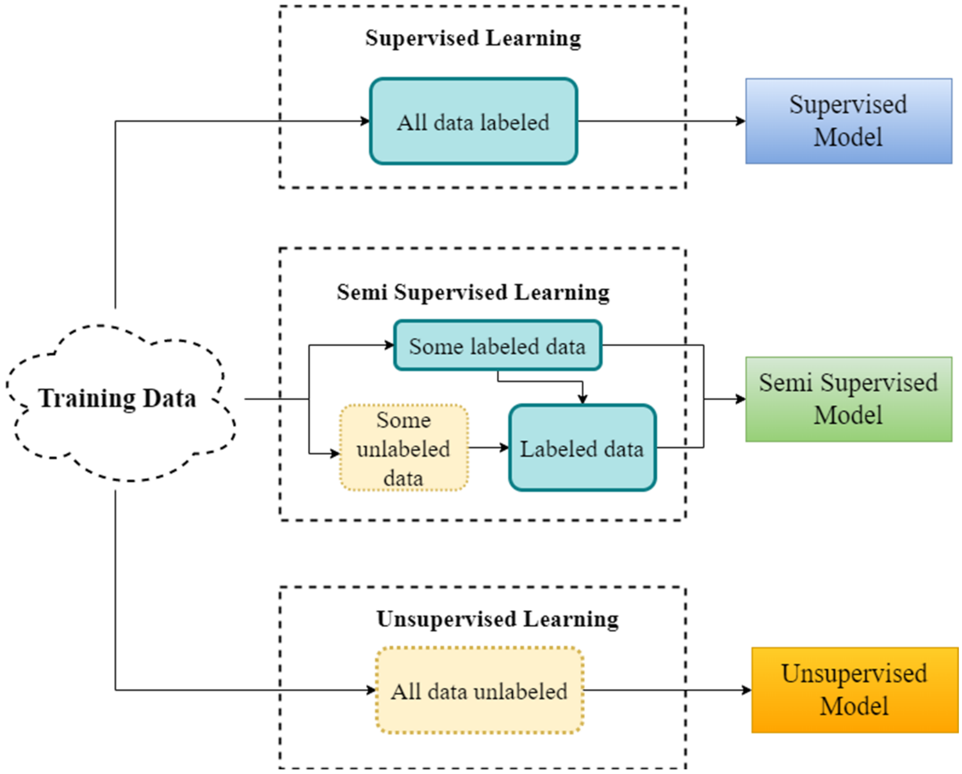
Succinctly, the first three differentiate each other on the labelling of the input training data during the model's creation while the latter two constitute special cases altogether (Brownlee, 2020a; IBM, 2020) that we do not address in this thesis. We illustrate this primary categorization in Figure 4-3 along with an over-simplification of the data flow (*cf.* Section "Machine Learning workflow" above) usually followed when generating a Machine Learning model. Training data of any form (represented by a cloud on the left of the illustration) flows into any of the three main Machine Learning scenarios depending on the labeled state its input data:

- If all input data is labeled, a Supervised learning model is to be used to yield a trained Supervised model.

- If some input data is labeled and some other is not, the labeled data will be used to deduce the labels of the missing input sample using a Semi Supervised Learning model, yielding in turn a trained Semi Supervised model.
- If all data is unlabeled, an Unsupervised Learning model is to be used to generate a trained Unsupervised model.

Figure 4-3

Machine Learning Primary Categorization (by Input Data Labeling)



Supervised learning

First, in the case of Supervised Learning, labels are provided, *i.e.*, metadata containing information that the model can use to determine how to classify it. However, properly labelled data is expensive[85] to prepare, and there is a risk of creating a model so tied to its training data that it cannot handle variations in new input data accurately, a phenomenon already visited in

85 Mostly in terms of computational resource allocation.

Subsection "Terminology" above called "overfitting" (Brownlee, 2021; Gottgtroy et al., 1970, p. 177; Ruck et al., 1993, p. 369).

Among the many models performing Supervised learning, here we introduce only Logistic Regression because of its relevance to this thesis.

(i) Logistic Regression. Logistic Regression (*a.k.a.* Logit Regression, Maximum-Entropy classification, Log-linear classifier) is a Machine Learning linear model for binary classification (Raschka & Mirjalili, 2019). Despite the term 'Regression' (Pedregosa et al., 2011, sec. 1.1.11; Raschka & Mirjalili, 2019), it has been pointed out that Logistic Regression is "a model for classification and not for regression" (2019, p. 60).

A Logistic Regression model estimates the probability that an instance belongs to any given class[86] (called the *positive class*, usually labelled "1"), and otherwise it predicts that it does not (*i.e.*, it belongs to the *negative class*, usually labelled "0") (Dangeti, 2017; Géron, 2019, p. 8; IBM, 2016). The two classes are designed to be mutually exclusive and exhaustive categories (Ratner, 2017, p. 105). When Logistic Regression has more than one input variable, it is called Multi-variate Logistic Regression. Similarly, when Logistic Regression can output more than one class, it is called Multinomial Logistic Regression (IBM, 2016).

Logistic Regression has its bases on odds: the odds in favor of a particular event (Raschka & Mirjalili, 2019). The odds can be written as:

$$\frac{p}{(1-p)} \tag{1}$$

where p stands for the probability of the positive event.

Now, let it be *logit* a function for the logarithm of the odds (log-odds):

$$logit(p) = \log\left(\frac{p}{(1-p)}\right) \tag{2}$$

[86] *i.e.*, the probability of an event occurring.

, where *log* is the natural logarithm. The *logit* function takes input values in the range 0 to 1 and transforms them to values over the entire real-number range. This leads to a linear relationship between feature values and the log-odds of the following form:

$$logit\big(p(y=1|\boldsymbol{x})\big) = w_0x_0 + w_1x_1 + \cdots + w_mx_m \tag{3}$$

$$= \sum_{i=0}^{m} w_ix_i$$

Here, ($p(y=1\,|\,\boldsymbol{x})$) is the conditional probability that any given instance belongs to class 1, given its features $\boldsymbol{x}$. The inverse form of the *logit* function is the probability that any given instance belongs to a particular class. This inverse form is also called the logistic sigmoid function ϕ, often abbreviated simply as the sigmoid function, due to its characteristic S-shape:

$$\phi(z) = \frac{1}{1+e^{-z}} \tag{4}$$

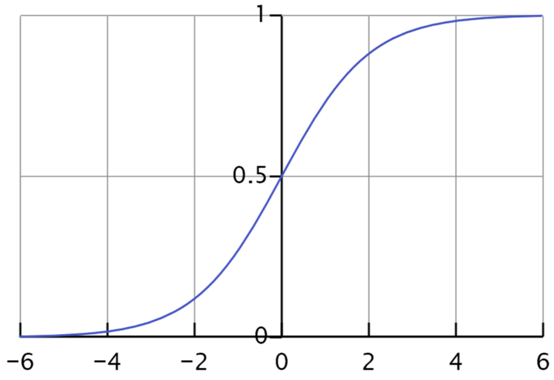
Here, z is the net input, the linear combination of weights $\boldsymbol{w}$, and the inputs $\boldsymbol{x}$ (the features associated to the training data):

$$z = w_0x_0 + \ w_1x_1 + \cdots + w_mx_m \tag{5}$$

As such, the sigmoid function $\phi(z)$, of which an example is plotted in Figure 4-4, takes real-number values as input, and transforms them into values in the range [0,1], with an intercept at $\phi(z) = 0.5$. More precisely, just before this last transformation, Logistic Regression does return a probability, *e.g.*, the probability of the Positive case, which can be used "as is" or be converted to a binary value, depending on a cutting point called the classification threshold (*a.k.a.* decision threshold) (Google Developers, 2022d).

Figure 4-4

A typical S-shaped curve (sigmoid curve), with an intercept at $\boldsymbol{\phi(z) = 0.5}$



Thus, Logistic Regression determines the best weights (*a.k.a.* estimators or coefficients) w_m such that the output of the function *p*(*x*) (the predicted probability that the output for a given $\boldsymbol{x}$ equals 1) is as close as possible to all real responses. The process of calculating the best weights w_m using available data is called model training or fitting (Raschka & Mirjalili, 2019).

We conclude this overview of Logistic Regression with a couple of remarks on its nature. Logistic Regression is a special case of "Generalized Linear Models with a Binomial / Bernoulli conditional distribution and a Logit link" (Pedregosa et al., 2011, sec. 1.1.11) employing a default threshold of classification of 0.50 . Also, Logistic Regression is considered a discriminative model, meaning that while it attempts to discriminate between classes or categories, it cannot "generate information, such as an image, of the class that it is trying to predict" (IBM, 2016). Finally, Logistic Regression is the most frequently used technique for MOOC learners dropout prediction showing satisfying results (Dalipi et al., 2018, fig. 2).

Unsupervised learning

Second, Unsupervised Learning must use algorithms to extract meaningful features to label, sort and classify its training data (which is unlabeled) without human intervention. As such, it is usually used to identify patterns and relationships (that a human can miss) than to automate decisions and predictions. Because of this, Unsupervised Learning requires huge amounts of training data to create a useful model (Brownlee, 2020a).

Semi-supervised learning

Third, Semi Supervised Learning is at the middle point of the two previous methods: it uses a smaller labelled dataset to extract features and guide the classification of a larger, unlabeled dataset. It is usually used when not enough labelled data is made available (or it is too expensive) to train a preferred, Supervised Model (van Engelen & Hoos, 2020).

Reinforcement learning

Fourth, Reinforcement Learning is a behavioral machine learning model akin to Supervised Learning, but the algorithm is not trained using sample data but by using trial and error. A sequence of successful outcomes will be reinforced to develop the best recommendation or policy for a given problem. Reinforcement Learning models can also be deep learning models (IBM, 2020).

Deep learning

Lastly, Deep Learning is a subset of Machine Learning (all Deep Learning is Machine Learning, but not all Machine Learning is Deep Learning). Deep Learning algorithms define an artificial Neural Network[87] that is designed to learn the way the human brain learns. Deep Learning models require a large amount of data to pass through multiple layers of calculations, applying weights and biases in each successive layer to continually adjust and improve the outcomes. Deep Learning models are typically unsupervised or semi-supervised (IBM, 2020).

For clarity reasons, the figure illustrating this Machine Learning categorization is available in Appendix 15. – Machine Learning algorithms.

---

[87] A quite complete and updated chart of existing Neural Networks was made available by van Veen & Leijnen (2019).

In this Section we overviewed a primary classification for Machine Learning methods, centering the reader's attention to the specific case of Supervised learning and the LogisticRegression model, both relevant to this thesis project.

Common metrics & measurement tools

In this Section we give an overview of metrics employed in Machine Learning classification tasks as well as other useful measurement tools. We begin by listing common terms usually employed in a Confusion Matrix. We then introduce the Receiver Operating Characteristic Curve, followed by a short description on commonly employed Classification metrics. To this, we add a small text on the Precision and Recall trade-off. We conclude with a comprehensive explanation on the Cross-validation notion, its variants, and the pitfalls if incorrectly performed.

Bisong (2019a, p. 180) suggests employing the Confusion Matrix (*cf.* Subsection "Confusion Matrix" below) and the ROC AUC (*cf.* Subsection "Receiver Operating Characteristic (ROC) Curve" below) for classifications tasks, whereas the Root mean squared error (RMSE) and R-squared (R^2) evaluation metrics are to be applied to regression tasks[88]. Besides, he insists on the importance of evaluating the model on previously unseen examples:

> The model's performance on the training data is evaluated to get the training set accuracy, while its performance on the test data is evaluated to get the test data accuracy when the model predicts the targets of previously unseen examples. Evaluation on test data helps us to know the true performance measure of our model. (Bisong, 2019a, p. 180)

Confusion Matrix

A Confusion Matrix represents counts from predicted *vs.* actual values (Bisong, 2019a, pp. 180–181; Dangeti, 2017; Géron, 2019, p. 90; Kulkarni et al., 2020, sec. 3.1). They can be

[88] Neither of which we cover in this thesis.

employed both for binary classification problem as well as for multi-class classification problems (Kulkarni et al., 2020, sec. 3.1). In a 2x2 square grid, each row represents "an *actual* class, while each column represents a *predicted* class" (2019, p. 91), and their intersections are defined as the True Positives (TP) count, the False Negatives (FN) count, the False Positive (FP) count, and the True Negatives (TN) count, as shown in Table 4-3

An Empty Confusion Matrix:

Table 4-3

An Empty Confusion Matrix

	Predicted: yes	**Predicted: No**
Actual: Yes	TP	FN
Actual: No	FP	TN

These terms are then defined as follows:

- True Positives (TP): indicates the number of positive examples classified accurately.
- False Negatives (FN): the number of actual positive examples classified as negative.
- False Positive (FP): the number of actual negative examples classified as positive.
- True Negatives (TN): shows the number of negative examples classified accurately. (Kulkarni et al., 2020, sec. 3.1)

Or even more precisely, accounting for correctly/incorrect predictions (Bisong, 2019a, p. 181; Google Developers, 2022e) one can infer that:

- A **true positive** is an outcome where the model *correctly* predicts the positive class. Similarly, a **true negative** is an outcome where the model *correctly* predicts the negative class.
- A **false positive** is an outcome where the model *incorrectly* predicts the positive class. And a **false negative** is an outcome where the model *incorrectly* predicts the negative class. (Google Developers, 2022e)

Classification metrics

Many publications (Bisong, 2019a, pp. 181–182; Géron, 2019, pp. 88–89; Google Developers, 2022b, 2022a; Kulkarni et al., 2020) list the Accuracy, Precision, Recall and F1 metrics as follows:

- Accuracy: the proportion of the total numbers of predictions that are correct:

$$\frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

- Precision: the ratio between the total of correctly classified positives and the total of correctly and incorrectly classified positives (and the inverse):

$$\frac{TP}{TP + FP} \; or \; \frac{TN}{TN + FN} \tag{2}$$

- Recall (*a.k.a.* Sensitivity, True Positive rate): measure of positives correctly classified as positives (and the inverse):

$$\frac{TP}{TP + FN} \; or \; \frac{TN}{TN + FP} \tag{3}$$

- F1-Score: it is the weighted harmonic mean between Recall and Precision. It is a special case of the F-measure where $\beta = 1$ in $F_\beta = (1 + \beta^2) \times \frac{Precision \times Recall}{(\beta^2 \times Precision) + Recall}$ (2020, p. 88):

$$2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

A useful tool for reviewing the previous metrics is via a Classification report, which simply is "a text report showing the main classification metrics" (Pedregosa et al., 2011, sec. 3.3.2.7) and their supporting counts.

Precision/Recall Trade-off

The effectiveness of a model is to be evaluated by examining both the Precision and Recall (Google Developers, 2022b), often in tension. Indeed, the trade-off between these two metrics hinges on the agreed threshold when assigning a given instance to the positive or to the negative class (Géron, 2019, p. 93). Depending on the value given to this threshold, a given instance may switch from being a False Positive to a True Negative, thereby increasing the Precision. Conversely, if one True Positive becomes a False Negative, the Recall decreases, *i.e.*, lowering the threshold for class assignation "increases Recall and reduces Precision", and raising the threshold "in general" also raises the Precision (2019, p. 93).

Receiver Operating Characteristic (ROC) Curve

The Receiver Operating Characteristic curve is a plot of the True Positive rate (*a.k.a.* Recall) against the False Positive rate (Centor, 1991; Géron, 2019; Google Developers, 2022c; Kulkarni et al., 2020, sec. 3.4).

The concept first appeared in the scientific literature in the works of Lusted (1971). It has its origins in military radar operations in World War II as a "means to characterize the operator's ability too correctly identify friendly or hostile aircraft based on radar signal" (Brown & Davis, 2006). It was originally meant to "explore the trade-offs between these competing losses at various decisions thresholds when a particular quantitative variable y, is used to guide the decision" (2006, p. 27). Nowadays the ROC Curve is employed to showcase the "performance of a classification model at all classification thresholds" (Google Developers, 2022c).

One way to compare Machine Learning binary classifiers is to measure the area under the curve (AUC) of the ROC curve, or more literally speaking the "Area Under the ROC Curve" (Centor, 1991). The AUC measures "the entire two-dimensional area underneath the entire ROC curve […] from (0,0) to (1,1)" (Google Developers, 2022c). It is to be interpreted as "the probability that the model ranks a random positive example more highly than a random negative example" (Google Developers, 2022c). Thus, a perfect classifier[89] will have a ROC AUC = 1, a good classifier would approach the top left corner (Fernández et al., 2018, p. 54), a random

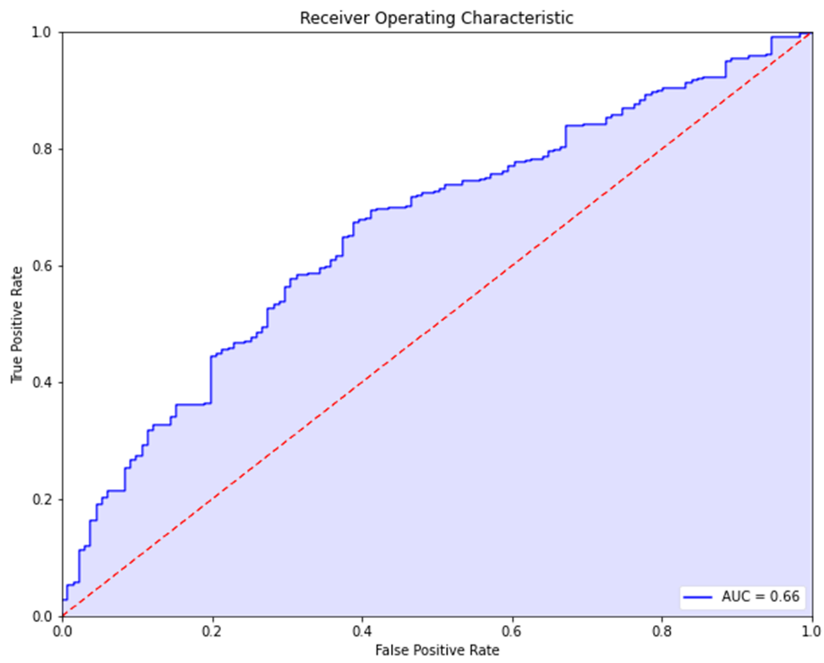
[89] A perfect classifier is considered a theoretical construct and thus, its appearance in real-life scenarios is likely an indication of a modeling issue.

classifier will feature a ROC AUC = 0.5 (Géron, 2019), and a classifier performing poorer than random guessing will feature a ROC AUC < 0.5, which "defeats the purpose" (Kulkarni et al., 2020, sec. 3.4).

An example of such a plot is shown in Figure 4-5, where the red dotted line represents a random classifier while the blue continuous line is the ROC Curve of the classifier being evaluated. The ROC AUC = 0.66 is the (digitally added) light blue area under the blue continuous ROC Curve.

Figure 4-5

Example of a ROC Curve and AUC



Thus, a ROC curve (shown in blue in Figure 4-5) "should always be in the upper diagonal" (Kulkarni et al., 2020, sec. 3.4), or in other words it is "to be in the upper-left-hand corner" (Davis & Goadrich, 2006, p. 1).

Computing the ROC AUC is interesting because of the following two[90] reasons:

[90] These could be alternatively reconsidered as caveats in specific model cases where invariance is necessary.

1. The AUC is scale-invariant, *i.e.*, it measures how well predictions are ranked instead of their absolute values.

2. The AUC is also classification-threshold-invariant, *i.e.*, it measures the quality of the model's predictions irrespective of the chosen classification threshold (Google Developers, 2022c).

Precision/Recall (PR) Curve (PRC)

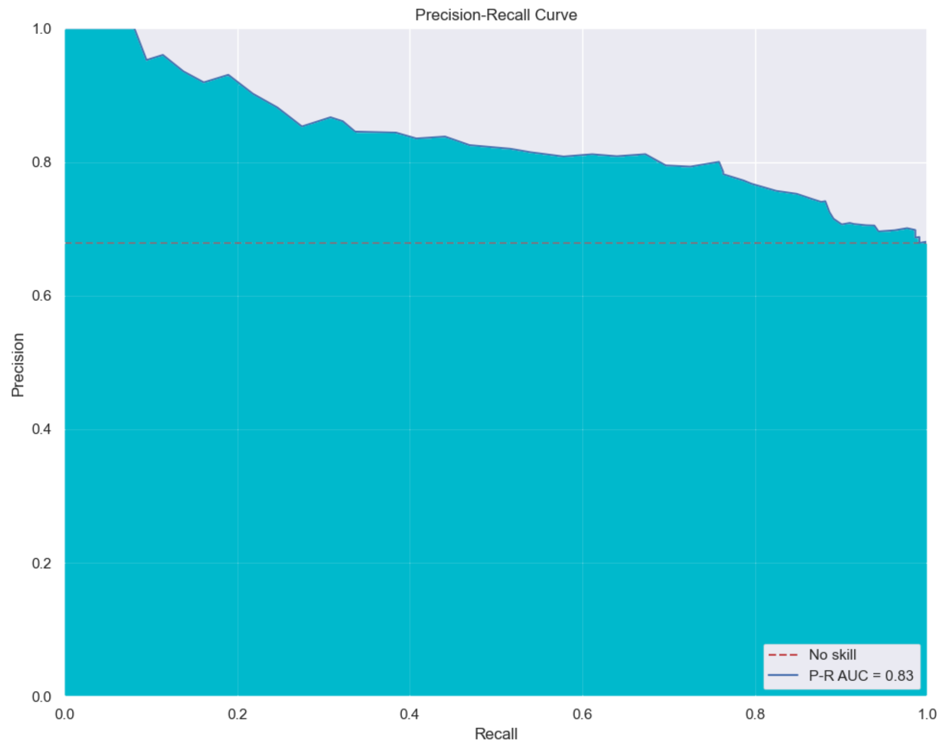
A Precision/Recall Curve is a plot of the Precision *vs.* the True Positive Rate (*a.k.a.* Recall, or Sensitivity) (Saito & Rehmsmeier, 2015, p. 7), alongside an horizontal baseline defined by the ratio of positive *vs.* negative cases.

We present an example of such a plot in Figure 4-6, where the red dotted line represents a no-skill classifier positioned at the ratio of positive *vs.* negative cases (~0.68, in this plot) while the blue continuous line is the PR Curve of the classifier being evaluated. In this example, the PR AUC = 0.83 and it corresponds to the (digitally added) blueish area under the blue continuous PR Curve. When reading a PR Curve, "[…] the goal is to be in the upper-right-hand corner"[91] (Davis & Goadrich, 2006, p. 1).

[91] Please notice that the ROC AUC is to be in the upper-LEFT-hand whereas the PR AUC is to be in the upper-RIGHT-hand corner.

Figure 4-6

Example of a PR AUC Curve



There are a couple of known cases when the ROC (*cf.* Subsection "Receiver Operating Characteristic (ROC) Curve" above) might not provide sufficient information to evaluate the performance of models. One of them is the case of imbalanced, binary sets, *i.e.*, those differing "in the number of positive and negative instances" (Saito & Rehmsmeier, 2015, p. 2), for which, once the model has been frozen, the PR Curve "allows for a quick and intuitive judgment of classifier performance" (2015, p. 18). Another case is that of "highly skewed datasets", for which PR Curves "give a more informative picture of an algorithm's performance" (Davis & Goadrich, 2006).

This is because, first, in an negative, imbalanced set, a significant change in the number of False Positives can lead to a small change in the False Positive Rate used in the ROC, while Precision, employed in the PR, instead "captures the effect of the large number of negative examples on the algorithm's performance" (Davis & Goadrich, 2006, p. 1).

And second, contrary to the ROC, the PR accounts for a moving baseline determined by the ratio of positive *vs.* negative cases, which affects the PR Area Under the Curve (Saito & Rehmsmeier, 2015, p. 7), *i.e.*, the plot accounts for the imbalance of the dataset and the resulting predictions of the model.

Yet, some authors provide with useful advice and recommendations; according to Géron (2019), a rule of thumb is that a PR Curve should be preferred over a ROC Curve "whenever the positive class is rare or when you care more about the false positives than the false negatives" (2019, p. 98). Also, Davis & Goadrich (2006) strongly suggested not to "linearly interpolate between points" in a Precision/Recall plot, as it would be "insufficient". They also insisted to keep in mind that optimizing an algorithm using the AUC ROC does not guarantee to also optimize the PR Curve.

Cross-validation & data splits

Cross-validation is a research subject on its own, with newer mathematical methods being developed as we speak (Bates et al., 2022) aiming to improve prediction error detection in the most diverse of scenarios of Machine Learning models and their training sets. As such, we employ it in its most general form and purpose, commonly considered a technique for estimating prediction error (Bates et al., 2022), clearly explained in the following entry in the specialized Encyclopedia of Database Systems (Liu & Özsu, 2020):

> Cross-validation is a statistical method of evaluating and comparing learning algorithms by dividing data into two segments: one used to learn or train a model and the other used to validate the model. In typical cross-validation, the training and validation sets must cross over in successive rounds such that each data point has a chance of being validated against. (Liu & Özsu, 2020, sec. Cross-validation)

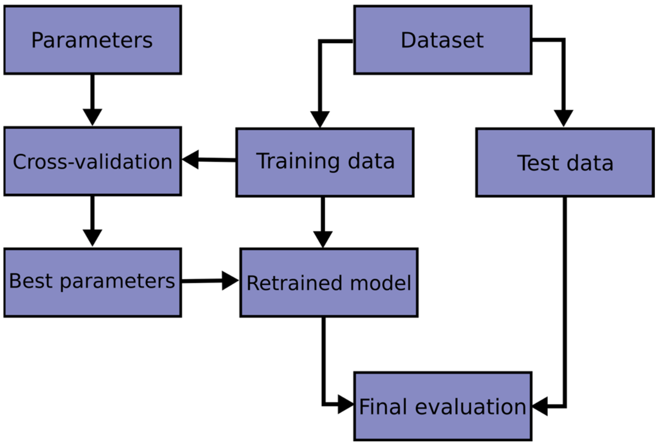
For Dangeti (2017), Cross-validation is yet another way to ensure "robustness in the model at the expense of computation" (2017), because more often than not, the train and test datasets "might not have been homogenously selected and some unseen extreme cases might appear in the test data, which will drag down the performance of the model" (2017). This means that by shuffling and splitting the train and test datasets for each of the several training runs, we ensure that outliers are accounted for on several occasions, either decreasing or increasing their effective participation in the model.

A corollary of the previous statement is that Cross-validation might be less effective if it is employed a single time. This is clearer on the conclusion reached on a survey (Neunhoeffer & Sternberg, 2019, p. 102) on this notion: "A problematic use of cross-validation occurs when a single cross-validation procedure is used for model tuning and to estimate true error at the same time" (2019, p. 102). This conclusion is also reached and described in the corresponding documentation of the scikit-learn project (Pedregosa et al., 2011): "Learning the parameters of a prediction function and testing it on the same data is a methodological mistake" (2011). Indeed, in a 2008 study, researchers (Rao et al., 2008, p. 8) suggest employing whenever possible "a sequestered test set that is only used when the classifier has been frozen" (2008, p. 8). This recommendation is further confirmed by many more researchers (Berrar, 2019, p. 6; Bisong, 2019a, p. 176; Isbell et al., 2023, p. 36; Pedregosa et al., 2011, sec. 3.1; Rao et al., 2008, p. 8; Simon, 2007, p. 175), extending it not only to the Cross-validation process but also to the general workflow of Machine Learning.

The sequestering or reservation of a test data set is illustrated (*cf.* Figure 4-7) in the corresponding documentation section of the scikit-learn project (Pedregosa et al., 2011, sec. 3.1) which shows the location of the Cross-validation process (left, center) during the determination of the best parameters of the model in the training phase, and also the prevalent isolation of the Test data (right, center), which is to be employed only for the final evaluation of the model (bottom).

Figure 4-7

Flowchart of Typical Cross-Validation Workflow in Model Training (Pedregosa et al., 2011)

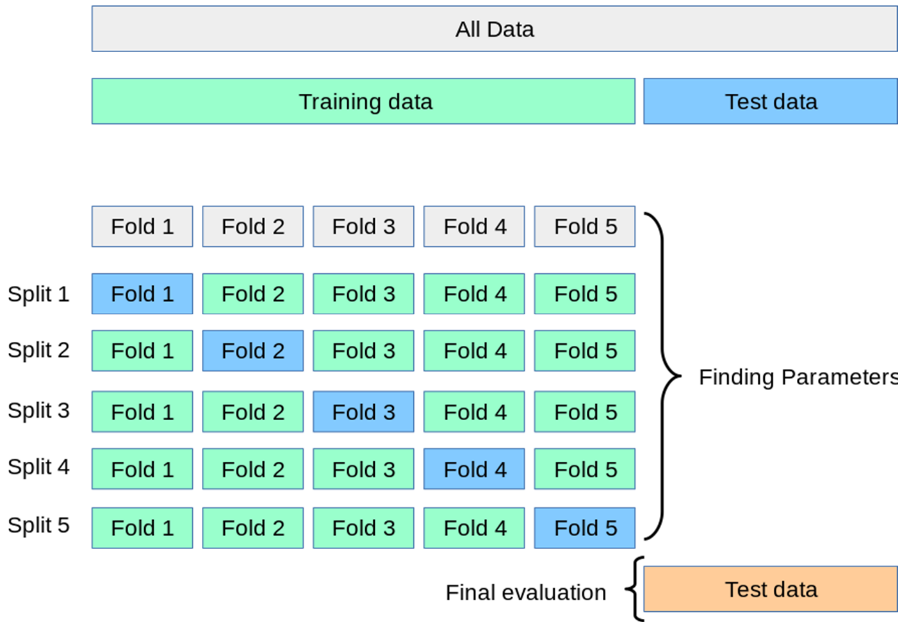


k-fold Cross-validation is a specific case of Cross-validation. During *k*-fold Cross-validation, data is divided into *k* near-equal parts and model training is performed using all but one of the parts of the data which in turn is used for model testing (Dangeti, 2017). Each of this distinct parts or subsets is called a *fold* (Géron, 2019, p. 73). The entire divide, train and evaluate process is repeated *k* times (Dangeti, 2017), *i.e.*, as many times as *k* parts the data has been split into, leading to training and validating the model *k* times. This process is called a *k*-fold Cross-validation, which is the basic form of Cross-validation (Géron, 2019, p. 73; Liu & Özsu, 2020).

We present Figure 4-8 from the corresponding documentation section of the scikit-learn project (Pedregosa et al., 2011, sec. 3.1) to illustrate a 5-fold Cross-validation data split: All Data is first separated into Training data and Test data, of which the latter it to be left untouched until the very end of the Machine Learning workflow (*cf.* Section "Machine Learning workflow" above). Then, a 5-fold Cross-validation is performed on the Training data, to figure out the best parameters for the model. During the 5-fold Cross-validation, the Training data is split into five Folds, employing only four folds (in green) to train the model, and one single fold (in blue) to evaluate the model. This is repeated for the five splits.

Figure 4-8

Example of a 5-fold Cross-validation data split (Pedregosa et al., 2011)



Concerning the number of folds, research (Cawley & Talbot, 2010, p. 2101) suggested that a small number of folds "five or less" would lead to poor model performance, due to overfitting whereas increasing the number of folds stabilized the performance across all reviewed algorithms. Opportunely, the results in a 90's article (Breiman & Spector, 1992, p. 17) seem to point to a specific number of folds: "The major surprise here is that ten-fold cross-validation is uniformly better in selection/evaluation than complete cross-validation (1992, p. 17)" while Berrar (2019, p. 6) similarly concludes on the results of a comparative study (Simon, 2007, p. 185) that "[a] sensible choice is probably k = 10, as the estimate of prediction error is almost unbiased in 10-fold cross-validation" (2019, p. 6) for small and large datasets (Simon, 2007, p. 185).

Yet another specific case of Cross-validation (among many others[92]) is the Stratified *k*-fold Cross-validation. In this variation, the fold splitting is performed in such a way that each fold preserves "the percentage of samples for each target class as the complete set" (Pedregosa et al.,

[92] Repeated *k*-fold, Leave One Out, Leave P Out, Shuffle & Split, Group *k*-fold, Leave One Group Out, etc. (Pedregosa et al., 2011, sec. 3.1).

2011, sec. 3.1), *e.g.*, each fold contains a representative distribution of the target classes as in the original Training data. This is extremely useful in the case of imbalanced sets (*cf.* Subsection "Main Definitions" above), where a different strategy for data splitting, *e.g.*, purely random data splitting, may present the model a data sample even more imbalanced as the original dataset, leading to easily avoidable bias. Still, when dealing with imbalanced sets, Neunhoeffer & Sternberg (2019, p. 103) further clarify on the importance in keeping the test fold imbalanced during Cross-validation:

> The right way of combining down-sampling of imbalanced data with cross-validation would be to first split the entire data set into the *k* folds and then only down-sample the folds that are used for training. The test fold should remain imbalanced to reflect the imbalance in unseen data. (Neunhoeffer & Sternberg, 2019, p. 103)

Whether for Cross-validation or for training a simpler model, the proportions of the data split has been subject of debate since the advent of the technique, for instance Simon (2007, p. 175) suggested a 50/50 ratio and up to a 66/33 ratio:

> There are no well established guidelines for what proportion of the data to use for the learning set and what proportion for the test set, or whether the split should be made randomly or in some systematic manner. Often half to two-thirds of the cases are used for the learning set. The split is often made randomly although in multi-center studies a closer emulation of external validation is obtained if one uses samples from some centers for learning and samples from other centers for testing. (Simon, 2007, p. 175)

Commonly, data is split into train and test data in a 70/30 ratio according to Dangeti (2017), although the same publication further on (2017, p. 10) also cites a 50/25/25 split when considering train, test, and evaluation sets. Considering the same tri-partite split, Bisong (2019a, p. 176) advocates instead for a 60/20/20 rule.

We take with a grain of salt these wildly different values and instead settle to employ as a starting point a 70/30 ratio, splitting then the 70% into an 80/20 ratio for a tri-partite split, and subsequently adjust both splits according to the nature of our input data.

Lastly, we conclude on the notion of Cross-validation with an important quote from the team behind Rao (2008, p. 8) highlighting its importance:

> A final note of warning: experienced machine learning researchers know not to tune a classifier by continuously observing the classifier performance on the test data until a desirable performance is achieved. When a classifier is tuned according to its performance on the test data, then the test results lose all their credibility since the classifier may no longer simulate real-world settings. More importantly, such a classifier loses its ability to generalize on new data, which is the key reasons to use cross validation in the first place. (Rao et al., 2008, p. 8)

In this Section we reviewed the most common metrics and measurement tools employed in Machine Learning, namely the notions of Confusion Matrix, the ROC AUC Curve, typical metrics employed during classification tasks, and we concluded on the importance of performing a Cross-validation correctly.

The first Part of this thesis presented the state-of-the-art involved in this research project. We presented flow, the human psychological state, generally dependent on an individually determined challenge-skill balance. In the framework of our specific research context, flow is to be understood via the emotion linked to the fact of realizing that one is progressing, that one understands, that one is understood.

We also described what MOOCs are, their role in the current historical educational context, and how they are usually categorized, according to their appropriate features.

We also presented a general overview of Machine Learning techniques, of which we highlighted Logistic Regression, a binary classifier, pertinent to our research context and constraints. Relevant terminology was also extensively presented.

Once all relevant elements, *i.e.*, flow, MOOCs, & Machine Learning, concerning this thesis are clearly detailed, we can present the proposed method to pool them for the research goal. Thus, the following Part is dedicated to the Research method approached in this thesis. It spans macro tasks and within, activities grouped in tasks.

Part II

Research method

*Research method

Part II concerns the research method employed in this thesis. It has been split into the following macro tasks: flow measurement literature review, flow measurement hypothesis, tools, & indicators determination, flow measurement protocol design, experimentation, evaluation, and finally conclusions & future perspectives. Each of these macro tasks contains detailed activities, which can be organizationally speaking grouped into smaller tasks.

The sequencing between the previously mentioned macro tasks is coarsely illustrated in Figure 5-1. Large blocks describe macro tasks englobing other, smaller tasks associated. We reassure the reader that activities developed within each task did take place even if they are not depicted in Figure 5-1. Macro task blocks connect to each other by arrows: the first executed tasks are at the top and last tasks are at the bottom. Concurrent arrows converging into a single task indicate both previous tasks need to be completed before proceeding to the converging new task, *i.e.*, conditional tasks.

Here follows a brief recap of what each of these macro task block entails, followed by Chapter 5 below where they are detailed.

- Flow measurement literature review.

During the literature review, we survey the existing efforts in flow measurement, more specifically in the context that concerns us: online, distant, educational settings. Furthermore, we position our research within the attempts already effectuated and refine our contribution to the fields our research touches.

- Hypothesis determination; tools and indicators determination.

The previous task's conclusions are determinant to establish our initial hypothesis: can the human psychological state of flow be automatically detected solely by seemingly disconnected usage traces in a MOOC? Such hypothesis verification will require decisions on the ample choice of tools (hardware, software, conceptual, terrain, participants, etc.), and the adequacy of indicators to correctly gauge its reach. Candidates for both categories will be roughly proposed and passed on to the next step, where they will be crystallized in a flow measurement protocol.

- Flow measurement protocol design.

This macro task concretizes the choices into a flow measurement protocol. This specifies the scientific context and the requirements, the research terrain and participants, various criteria for data collection, data processing, indicators, metrics, and results, all of which are closely interconnected to each other.

- Experiments.

The experiment macro task comprises two data collection tasks spanning two years. In the Proof-of-Concept experiment, we verify that the two flow measurement instruments reliably measure flow when paired together. This Proof-of-Concept task results in a Machine Learning model that successfully detects flow when using only one of the two flow measurement instruments. In the flow-detecting model experiment, we additionally include the user logs from the MOOC into the Machine Learning model training. This task results in a Machine Learning model that successfully detects flow automatically, transparently, in MOOC learners.

- Evaluation.

Primarily, we use relevant metrics for Machine Learning classification tasks to measure the performance of the resulting models. Once the Machine Learning flow detecting model has been successfully obtained, we compare its final performance to the pre-defined indicators on unseen data to conclude the hypothesis.

- Conclusion & perspectives.

Finally, we reach and present the general conclusions on the Machine Learning flow-detecting model, its limits, and the future perspectives to undertake.

The following Chapter 5 below goes into detail for the above-mentioned macro tasks.

Chapter 5. Method

This Chapter details the macro-tasks described in the introductory part above, and illustrated in Figure 5-1. Also, this Chapter features extracts from a publication (Heutte et al., 2022) with participation of this thesis' author.

First of all, we must indicate that this doctoral thesis project follows the THEDRE[93] method (Mandran & Dupuy-Chessa, 2017) for Research in Human-Centered Computer Science (RHCCS[94]). The THEDRE method insists on a continuous quality improvement process, with the ultimate goal of creating/obtaining a "traceable research process" (2017, p. 5). It is axed on a process structure that follows the PLAN, DO, CHECK, and ACT subprocesses, focusing on research construction, experiments development, experiments assessment, and decision-taking and results communication, respectively.

Specific to the content of this Chapter 5, the THEDRE method distinguishes and proposes three types of research-related indicators: 1) steering indicators, 2) activity indicators, and 3) results indicators[95].

1. Steering indicators can be broadly described as "expectations before publishing" (2017, p. 7), and they are defined to follow the research work evolution.

2. Activity indicators are evolving, Boolean or numerical values related to tools (or their components) employed during tasks in the research subprocesses, *e.g.*, number of publications read, number of interviews performed, etc.

[93] https://thedre.imag.fr/
[94] RHCCS aims to consider people and their IT environment, whether it is for work (in the information system domain), in a learning environment (in the learning domain), or simply to account for interactions with a given machines (in a human-computer interaction context) (Mandran & Dupuy-Chessa, 2017, p. 1).
[95] Please notice that the term "indicator" within the THEDRE method is vastly different to the meaning employed in Chapter 6, "Trace analysis in a MOOC".

3. Lastly, results indicators correspond to productions issued from tasks in the research subprocesses, *e.g.*, plans, data files, documents, protocols, etc.

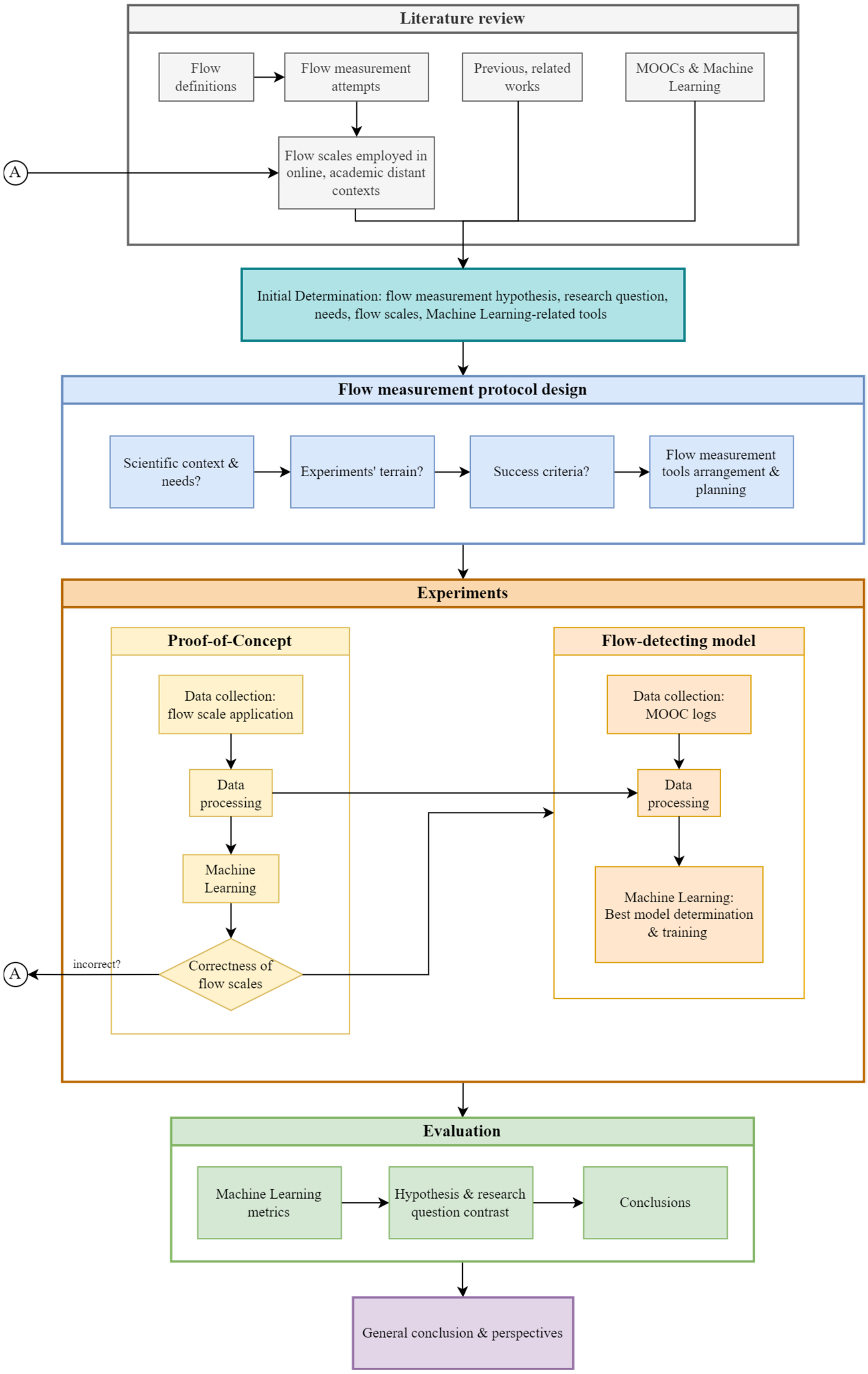
It is important to highlight that results indicators constitute a checklist of productions (2017, p. 8). For the sake of readability, we define and detail such indicators (steering, activity, or results) only during this Chapter.

Also, indicators are SMART: Specific, Measurable, Acceptable, Realistic, Time-bound, within the realms of reality and feasibility (Doran, 1981).

Likewise, this research project led us to conduct literature reviews on a variety of connected subjects (*cf.* Chapter "Published articles" on page 357), for which we employed both the "PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions" (Emily Jones, 2020; Liberati, Altman, Tetzlaff, Mulrow, Gotzsche, et al., 2009; Liberati, Altman, Tetzlaff, Mulrow, Gøtzsche, et al., 2009; Moher et al., 2009, 2009), and the "Guidelines for performing Systematic Literature Reviews in Software Engineering" (Kitchenham & Charters, 2007) which share important parts, such as describing a Rationale, Objectives & Research questions, Eligibility criteria, Information sources & Search strategy, Screening process & Study selection, and Data collection & Features.

Figure 5-1

Method: Macro Task Sequencing



Flow measurement literature review

We started by surveying the existing flow measurements approaches and attempts, careful not to repeat existing work or miss out relevant tools adapted to our goal. This phase is firmly founded and described in Section "Measurement Attempts of Flow" above.

Such measurement approaches depended heavily on the theoretical flow model chosen by their authors, if any (*cf.* Section "Evolution of the Definition of Flow and its Conceptual Flow Model" above), which were often conducted on a specific research field, and thus, *a priori*, incompatible between each other. This led us to focus on the much smaller and manageable number of flow measurement tools employed in educational settings, specifically in online, distant training and education (*cf.* Section "Flow in Educational Contexts" above), which constitutes the research context of this thesis.

Nevertheless, we took notice and benefitted of the suggestions of experienced researchers from across different conceptual flow models and flow measurement tools attempts, explained in Section "Considerations When Measuring Flow" above.

For the sake of completeness, in Figure 5-1, we provide a return loop (circle A) to allow for process' correction, once the results of the Proof-of-Concept (*a.k.a.* Experiment 2, *cf.* Section "Experiment 2 – Prototype" below) come in.

Indicators

For the literature review tasks, and given the limitations in research time, software availability and capabilities, scientific databases access restriction and security limitations[96], we settled into the following indicators.

(i) Steering indicators.

- Determining a method for performing literature reviews (*cf.* Chapter 5 above).

[96] The University's academic access performs a security ban on accounts attempting to mass download articles within a short period of time. A few hours' wait time to lift the temporary ban ensues.

- Employing an academic citation retriever and analyzer software.
- Employing the Google Scholar, Scopus, and Web of Science scientific databases. As Google Scholar is a web search engine indexing other scientific databases with acceptable performance (Beel & Gipp, 2010), we paid added attention to results from Taylor & Francis Online, Science Direct, Sage Publications, Springer, and IEEE Explore.
- Identifying flow measurement instruments and measurement considerations from previous and concurrent attempts (*cf.* Section "Measurement Attempts of Flow" above).
- Identifying a recognized flow-in-learning-context theoretical model (*cf.* Section "Flow in Educational Contexts" above).

(ii) Activity indicators.

- At the very most, 1 000 results yielded from a research query (Publish & Perish ® own hard limitation but also an exhaustive limiter).
- At most, 500 abstracts to read, once citations, misses, and duplicates are removed.
- At most, 100 articles to fully read, once misses are removed.
- At most, 50 articles to consider relevant for actual research.
- Depending on the review, focus on the 1975-2022 period (flow and flow measurement instruments), or the 2019-2022 period (similar works, *cf.* "Statement of Contribution" on p. 4).
- At most, two months of time allocated for each of the literature review's all sources reading subprocess, without accounting for the writing process.

These indicators carry the associated risk of missing any relevant literature, particularly when looking at the existing similar works. However, further revisions showed no breaking-through contributions happened between the determined periods.

(iii) Results indicators.

- A literature review on flow (*cf.* Chapter 1 above).
- A compendium of possible psychometric flow measurement tools (*cf.* Appendices).

(iv) Additional results.

- Source code for web scraping literature metadata.

Flow measurement hypothesis, tools, & indicators determination

We posit that the human psychological state of flow can be automatic and transparently detected via the digital traces left by MOOC participants.

We intend to address this issue by employing Machine Learning techniques to pair traditional flow measurement tools to the digital traces left by MOOC participants. We postulate that Machine Learning techniques will help unearth the subjacent relationship between the learners' flow state and their corresponding digital traces, up to some reasonable degree.

This led us to mobilize diverse means and methods available to us, ranging from psychometric questionnaires (*cf.* Section "Flow in Educational Contexts" above), mass survey application (*cf.* "Terrain: the MOOC "*Gestion de Projet*"" & "Data collection" below), trace analysis life cycle and methods (*cf.* Chapter 6 below), Machine Learning methods (*cf.* Section "Machine Learning methods" above) and workflows (*cf.* Section "Machine Learning workflow" above), multi-metric performance measurement (*cf.* Section "Common metrics & measurement tools" above), big data collection (*cf.* Subsection "Traces: log data" below), and finally security and ethical issues (*cf.* Section "Flow measurement protocol design" below & footnote [110] below).

Indicators

This mobilization led us to propose (and determine candidates) tools of the most diverse nature and their corresponding indicators to validate them, most notably what constitutes the "reasonable degree" of certitude mentioned just above, and its threshold.

(i) Steering indicators.

- Determining a research question.
- Researching a research terrain.
- Determining mass survey tools.
- Researching data analysis tools, software, and general metrics.

(ii) Activity indicators.

- At most six months of time allocated for determining candidates.
- Machine Learning metrics candidate threshold > 0.5 minimum.

(iii) Results indicators.

- A research question: Can the human psychological state of flow be automatic and transparently detected via the digital traces left by MOOC participants?
- Access to a free and open-source on-line survey web-app: LimeSurvey[97].
- Arrangements for access to research terrain candidate: the MOOC « Gestion de Projet » (*cf.* Section "Terrain: the MOOC "*Gestion de Projet*"" below).
- Data analysis software candidates: IRaMuTeQ[98] (textual statistical analysis software), R[99] (programming language), Python[100] (programming language), IBM® SPSS[101] (statistical analysis software), Tableau Desktop[102] (data analysis and visualization), and/or ELAT (Open edX log analysis tool) (Torre et al., 2020).

---

97 https://www.limesurvey.org/
98 http://iramuteq.org/
99 https://www.r-project.org/
100 https://www.python.org/
101 https://www.ibm.com/products/spss-statistics
102 https://www.tableau.com/products/desktop

- Machine Learning methods candidates: supervised classifiers featuring numerical multivariable input, yielding performance metrics between 0 and 1, where higher is better. Also, we are aware that many Machine Learning methods fulfill these initial, broad requirements.

(iv) Additional results.

- A method for supporting Scoping Review (Heutte et al., 2022).

- Determination of needed additional software: Publish or Perish[103] (academic citation retrieving software), Libre Office Calc[104], or Apple Numbers[105], or Microsoft Excel[106] (spreadsheet software), Mendeley Desktop[107], Zotero[108] (reference managers), and Microsoft Word[109] (word processing software).

- Arrangement for admin access to a calculation PC, plus a virtual calculation server, for data storage, database storage and access, and data analysis capabilities.

Flow measurement protocol design

Flow measurement protocol design is a well-researched topic within the flow community (*cf.* Section "Measurement Attempts of Flow" above). Alas, it has been a central part of flow research since the inception of the first flow theory (*cf.* Section "Evolution of the Definition of Flow and its Conceptual Flow Model" above). Yet, flow measurement in educational contexts is still in its infancy (*cf.* Section "Flow in Educational Contexts" above), with research specialized in online and distant education can be hand-picked as only recently it surfaces world-wide.

---

103 https://harzing.com/resources/publish-or-perish/
104 https://www.libreoffice.org/discover/calc/
105 https://www.apple.com/numbers/
106 https://www.microsoft.com/en-us/microsoft-365/excel
107 https://www.mendeley.com/download-reference-manager
108 https://www.zotero.org/
109 https://www.microsoft.com/en-us/microsoft-365/word

Indeed, it is the research field that interests us what constitutes the main constraint when designing a flow measurement protocol: educational context, yes, but online and distant?

That is why, besides considering the suggestions of experienced researchers from across different conceptual flow models and flow measurement tools attempts (*cf.* Section "Considerations When Measuring Flow" above), we need to primary account for the online, distant, educational settings. Under the risk of seeming repetitive but indeed precise, the designed flow measurement protocol thus must:

Be adapted to online, distant, educational settings: the chosen instruments, tools and methods involved in the protocol must have previously been shown to work and yield reasonable results when applied in this specific context. For learners, they must allow for controlled, transparent, online, distant access, and participation, have a low-pronounced learning curve, *i.e.*, of easy learning and adoption, demanding little to no additional cognitive effort.

Besides the user-centered constraints above, researchers face additional considerations:

- legal and ethical, according to the EU General Data Protection Regulation[110] (EU GDPR), the tenets of Human-Centered Machine Learning[111] (HCML), and the upcoming Regulatory framework proposal on artificial intelligence[112] (Artificial Intelligence Act).

- reliable, *i.e.*, adapted to an online, distant controlled access, producing the same quality of service and performance independently of the system load, for all users involved, and consistently produce the same results (*e.g.*, scores, predictions, etc.) when given the same input.

[110] *Cf.* The EU GDPR Practical Guide (Voigt & Von dem Bussche, 2017).
[111] *Cf.* Practices of Human-Centered Machine Learning (Chancellor, 2023, p. 81).
[112] *Cf.* The Regulatory framework proposal on artificial intelligence (Artificial Intelligence Act, 2021; European Union, 2022).

- cost-effective, favoring free or open-access solutions (given similar or better performance), especially when considering their deployment to different geographical regions.
- complete, avoiding artifacts necessitating specialized, separated modules to work as supposed to, *i.e.*, wholesome tools and methods, *e.g.*, traces tracking working out-of-the-box, scoring calculation unveiled, relevant file formats compatibility, data import/export and backup capabilities bundled, libraries included, batteries included, etc.

Similar constraints apply to the research terrain, a TEL platform. Close collaboration with a TEL platform is imperative, preferably one able to recognize the benefits of the results of applied research in the learning process, the learners' and thus the platform itself. Aware of these advantages, the TEL platform's managers are also familiar following the EU GDPR, and it is likely they have already participated in research experiences in the past. If it is so, the TEL platform managers already have in place research protocols, accounting for legal and ethical data management, security, results publication, and ulterior data exploitation. If not, they are willing to go out of their way and invest themselves to implement such measures and protocols as well as to adapt their existing workflow for scientific research.

Obviously, such TEL platform, already experienced in delivering educational content in online and distant settings, should count with a steady stream of eager learners, from diverse demographical, professional, and/or educational backgrounds, likely open to participation in scientific endeavors, able to constitute a legitime research body of participants.

Moreover, the TEL platform should already have in place methods for tracking learners' actions while on the platform, up to a certain degree of granularity. Again, if not, the platform's managers are willing to implant and exploit such tracking devices, under the corresponding normativity, while informing their participants and accounting for the due diligence applicable to such cases. These tracking devices should yield, and store standardized log data while TEL platform's managers should provide us with access to this learners' log data in a timely fashion, allowing for their copy, storage, treatment, and exploitation for research purposes.

If deemed necessary, eventual contracts should be drafted ahead, accounting at least for each party's obligations and responsibilities.

Lastly, success criteria are dependent on the chosen tools, instruments, and methods. Generally, wholesome solutions provide instructions for scores, metrics, and performance indicators, some being as simple and basic as "it works" to complex plot-reading and graph-interpreting. If an adopted solution was not to provide indicators for success and an alternative is impossible, diverse gauge suggestions can be found in the literature, on a per-case basis. Specific to research contexts, acceptance, and publication of proposals of scientific production involving this research also constitute a valid indicator of success criteria (*cf.* "Communications in International Conferences" & "Communications in National Conferences" on page 352).

It is noteworthy to mention that constructing *ad-hoc* solutions can prove to be extremely well adapted to the specific research context, but it is a task that requires additional, time-consuming, run-in-parallel, confirmation and validation processes. Furthermore, such solutions can become a speedbump for replicability and sharing, and thus affect their scientific/commercial adoption. It is our view that adopting such solutions should be kept to a minimum, after all: *nanos gigantum humeris insidentes*.

Indicators

Given the previous points, we concluded on the following choices, on the distinct areas concerned:

(i) Steering indicators.

- Employing the FlowQ and the EduFlow-2 flow measurement instruments, as detailed in Section "Flow measurement instruments identification" below, selected after accounting for the Considerations When Measuring Flow described above.

- Consequently, adopting the EduFlow-2 theoretical model for flow in educational contexts (*cf.* Section "Flow in Education (EduFlow & EduFlow-2)" above).

- Employing the MOOC « Gestion de Projet » as terrain for questionnaire application, data collection, and log tracking (*cf.* Section "Terrain: the MOOC "*Gestion de Projet*"" below).

- Employing an explainable (no black box) Machine Learning method, *i.e.*, dispelling all Neural Networks and Deep Learning possible choices.

- Employing the survey web-app tool LimeSurvey.

(ii) Activity indicators.

- Collecting at least two Sessions (one year) worth of data (*cf.* Section "Data collection" below).

- Collecting at least the psychometric questionnaires' data for the above-mentioned timespan (log data access was hindered by the effects of the preventive measures of the COVID-19 pandemic).

- Collecting at least 400 MOOC participants' worth of data collection (both for the psychometric questionnaires and their MOOC log data).

These indicators carry the associated risk of not having enough data samples at the end of the data collection process, both in terms of number of participants (*n*) and in terms of their MOOC log data, because at this point it is not possible to assess if log granularity in the selected MOOC platform will be enough to our research ends. Both situations might be conducive to a Machine Learning model unable to detect flow.

(iii) Result indicators.

- A flow measurement protocol design (*cf.* Subsection "Flow measurement protocol" below).

(iv) Additional results.

- None

Flow measurement protocol

Choose a research terrain with at least 15x ratio of historically-assured number of completing participants *vs.* total number of registered participants (*cf.* Background on page 1 on MOOC completion rates and the Activity indicators just above).

Once the terrain has been established, determine the platform software engine behind the MOOC platform terrain and replicate its database setup[113] in an accessible and secure server.

1. If permitted and if it is not collected elsewhere, design and create a separate demographics survey to be applied before the flow measurement tool survey. Possible variables include gender, birth year, socioeconomic status, maximum level of studies, etc. but it mostly depends on the application context and research needs. Avoid open-answer fields and privilege drop-lists or tick boxes. Code-identify (internally) these items separately from the flow measurement tools (*e.g.*, demo_year, demo_country).

2. Create the flow measurement tools survey in the selected survey platform (items' codes are visibly only to the researcher):

    a. Code-identify (internally) the flow measurement items per instrument name (*e.g.*, wolf) and per item number (*e.g.*, wolf1, wolf2, wolf3, etc.), or dimension (*e.g.*, eduflow_d1_1, eduflow_d1_2), if the instrument does not provide any.
    b. To discern and later discard random-answering participants, design verification items unrelated to the flow measurement tool, of similar text length, even if slightly unnatural, with possible answers in the available scale (Likert 1-7, or Likert 1-5, etc.), avoiding 0 or null as possible valid answers, *e.g.*, "Please choose among the proposed numerical answers the one corresponding to the solution to the 2+3 operation", or simply "Please select the option 5 for this question and continue".

[113] https://edx.readthedocs.io/projects/edx-installing-configuring-and-running/en/latest/installation/index.html

c. Code-identify (internally) the verification items including their answer in the code, along with the instrument name code, *e.g.*, wolf_test5, eduflow_test5, for the above sentence example.
d. Do not use the same text for all verification items. Do not give all verification items the same answer.
e. Include and insert the verification items strategically; one per every 30 items in a single page, 0 if the page displays less than 7 items at once.
f. Proof-test for grammar and typos with a selected 3rd party, while still editing the survey.
g. Play an access-controlled survey run test with a selected 3rd party, plus a data-collection test, *e.g.*, verify data comes in, in the right format, and it is stored as intended.

3. Restrict survey access to MOOC participants only, making sure a stored and tracked common key uniquely and reliably identifies users between the MOOC platform and the survey platform (private access).

4. Inform MOOC learners (if the system allows it and access to personal data is permitted by applicable normativity, using personalized communication, *e.g.,* "Good morning Mrs. Smith") of their valued research contribution by participating in the survey, reminding and reassuring them no repercussions would arise (grades, fees, etc.) from declining involvement: participation is voluntary.

5. Collect allowed demographics, if possible, via the same survey tool.

6. Inform MOOC learners the availability of the survey, providing an opt-out link in the personalized communication (an example is shown in Table 5-1 featuring automatically personalized fields, *e.g.*, {SURVEYURL}). Honor those who choose to opt out.

Table 5-1

Example of an Invitation Email in French for a Second Survey Participation

Survey invitation email
Bonjour {USER}, Vous avez répondu au premier questionnaire recherche et nous vous en remercions ! Le MOOC est maintenant à mi-chemin et nous souhaitons à présent mieux

comprendre la manière dont vous organisez votre formation. Ce deuxième questionnaire est un peu plus long mais facile à remplir.

Pour vous remercier de votre implication, un tirage au sort aura lieu parmi les répondants aux questionnaires recherche et permettra à l'un ou l'une d'entre vous de gagner un chèque cadeau Amazon d'une valeur de 25 euros.

Cliquez sur le lien suivant : {SURVEYURL}

Merci de votre participation !

L'équipe "recherche" du MOOC GdP

Si vous ne souhaitez pas participer à ce questionnaire et ne souhaitez plus recevoir aucune invitation, veuillez cliquer sur le lien suivant :
{OPTOUTURL}

7. Apply the flow measurement instruments as often as possible, at least twice; in the middle and at the end of the entire MOOC.
   a. During the same Session alternate between the a-dimensional instruments and the dimensional instruments, starting with an a-dimensional one, to alleviate cognitive load.
   b. Apply to all participants (no participant filtering).
   c. Apply as independently as possible from other known cognitive-charged tasks, such as exam dates, local elections, platform's evaluations (where it applies, if possible).
8. Remind the MOOC learners to start the survey, if they have not opened it, or to complete the survey, if they abandoned it half-way. One reminder per application moment is the maximum. Provide an opt-out link in the reminder and honor those who choose to opt out.
9. At the end of the Session:
   a. Collect and save incrementally demographic data into a relational database, if possible.
   b. Collect and save incrementally the questionnaires' data into a relational database.
   c. Calculate flow scores, per measurement instrument.
   d. Collect and save incrementally log data into a database (if the terrain log structure allows it, privilege a relational database,). Keep in a different table/collection the most recent, concatenated version of all log data.

e. Perform backups of all data.
f. Verify compliance with the EU GDPR and the tenets of HCML.
g. Start exploratory data analysis (*cf.* Subsection "Data filtering and cleaning" below).

10. Once all Sessions contemplated for data collection have elapsed:

a. Concatenate demographic and questionnaire data into a single data source, database, or file, respecting appropriate fields (columns, corresponding to the flow measurement and verification items, demographics, and scores) and users (rows, corresponding to the uniquely identifying common keys) distribution.
b. Clean up questionnaire data: filter out participants who did not finish the questionnaires (unfinished), who left empty answers (unanswered), who answered validation items incorrectly (invalid), and who answered highly unlikely demographic data (*e.g.*, birth year=1906, ticked all gender boxes), where it applies.
c. Concatenate log data into a single table/collection, append a Session identifier per source (*e.g.*, 2019s1, 2020s2, etc.) Make a backup of this monolithic source.
d. Clean up log data: filter out participants who have invalid platform data (invalid sessions, corrupt logs, timestamped data unrelated to the study, etc.), and who did not participate in the questionnaires. Verify collected data compliance with the EU GDPR and with the HCML tenets.

For Experiment 1 (Proof-of-Concept), using the selected Machine Learning method, pair the cleaned-up questionnaire data containing the flow measurement instruments scores. Relevant Machine Learning metrics (*cf.* Subsection "Results and metrics" below) should point to the adequacy of the selected flow measures. Consider reviewing prior choices if the metrics show poor results.

For Experiment 2 (Prototype) design indicators (*cf.* Subsection "Indicators construction" below), based on the selected flow theoretical models:

- Consider the issue of overfitting and therefore strike a balance between general and terrain-specific indicators, according to the theoretical flow models, *e.g.*, is the specific title of the MOOC relevant for the Machine Learning model? Is the grammar/vocabulary displayed on forum posts relevant for flow?

- Construct indicators based on proxies likely available in the digital traces, *e.g.*, intense eye fixation on text (a possible immersion indicator) is highly unlikely to be found in typical MOOC digital traces.

Experiments

Given the novelty of our proposed approach for flow measurement (relying only on digital traces), we cautiously split this macro task into a Proof-of-Concept experiment first (Experiment 1), and a subsequent Prototype flow-detecting experiment (Experiment 2).

The Proof-of-Concept aims to verify that the chosen flow measurement tools correctly behave as intended on their own, but most particularly when paired together, alongside the selected Machine Learning technique. Discrepancies in results at this stage would put into question the validity of the entire process.

During this macro task, we filter, clean, and standardize data issued from the flow measurement tools' results. Ensuing statistical exploratory analysis confirms data contain enough information (*cf.* Subsection "Data filtering and cleaning" below) to put in motion Machine Learning techniques. The resulting Proof-of-Concept Machine Learning model is itself a flow-detecting tool, ready to be implemented into a MOOC dashboard, if it is accompanied by the EduFlow-2 measurement instrument application.

Thereafter, the Prototype flow-detecting Experiment 2 capitalizes on the results of the Proof-of-Concept Experiment 1 and steps forward involving the learners' traces. We gather log data (*cf.* Subsection "Flow measurement protocol" above) from the same users participating in the Experiment 1, *i.e.*, behavioral data generated when interacting with the MOOC platform. We filter it, clean it, transform, and aggregate it to generate flow-related indicators (*cf.* Subsection "Logs data filtering, cleaning, and aggregating" below) according to the chosen flow theoretical model (*cf.* Subsection "Flow in Education (EduFlow & EduFlow-2)" above).

These indicators are paired with the same data serving as input data to Experiment 1. Together, they train a Machine Learning model to learn to identify flow in participants in a MOOC.

Indicators

(i) Steering indicators.

- For Experiment 1; defining relevant metrics and obtaining satisfactory results with the selected flow measurement instruments and the Machine Learning method.
- For Experiment 2; simulate deployment by further splitting data (*cf.* Subsection "Model training" below).
- For Experiment 2; design and construct at least one indicator per flow component (*cf.* Subsection "Logs data filtering, cleaning, and aggregating" below).
- For Experiment 2; obtaining satisfactory metrics (*cf.* "Activity indicators." just below) with the constructed indicators (*cf.* Subsection "Indicators construction" below).

(ii) Activity indicators.

- Metrics for Experiment 1:

  Means of the Accuracy, Precision, Jaccard, and F1 scores during the CrossValidation > 0.7, with a standard deviation $\sigma < 0.1$;

  Weighted average of Precision and Recall > 0.7 in the Classification Report;

  ROC AUC > 0.7;

  TP > 0.7 & TN > 0.5 in the normalized Confusion Matrix.

- Metrics for Experiment 2: (highly imbalanced sample)

  Means of the Balanced Accuracy, F1, ROC AUC, Recall, and Precision scores during the CrossValidation > 0.6, with a standard deviation $\sigma < 0.2$ for the reserved data;

Precision, Recall, and F1 scores > 0.6 in the imbalanced Classification Report;

ROC AUC > 0.6

PR AUC > 0.6

TP > 0.6 & TN > 0.6 in the normalized Confusion Matrix.

(iii) Result indicators

- Source code yielding the experiments' results.
- From Experiment 1: an automatic, asynchronous flow-detecting Machine Learning model (based on EduFlow-2).
- From Experiment 2: an automatic, transparent, quasi-real-time flow-detecting Machine Learning model (based only on MOOC users' logs).

(iv) Additional results.

- Publications on the results (*cf.* "Published articles" on page 352).
- Determination of the most relevant features affecting flow detection (*cf.* "Model training" below).

Evaluation

We evaluate both the Proof-of-Concept Machine Learning model, and the Prototype Machine Learning model using relevant Machine Learning metrics, of which the thresholds are described as Activity indicators in Chapter 5(ii) above.

In Experiment 1, we rely on metrics related to simple Machine Learning classification tasks: Accuracy, Precision, Jaccard, and F1 scores (*cf.* Subsection "Classification metrics" above for definitions).

Experiment 2 proved to be much more complex as to evaluate using the same metrics: the extremely imbalanced nature of the input dataset (*cf.* "Model training" below) showed misleading results, either too optimistic or too pessimistic. Thus, metrics adapted or relevant for imbalanced sets were then employed (*cf.* Subsection "Common metrics & measurement tools" above), such as the PR AUC (*cf.* "Precision/Recall (PR) Curve" above).

Details on the evaluation tasks are described in the Results Subsections for the Experiment 1 (Results and metrics), and Experiment 2 (Results and metrics). Metrics results were within the predetermined thresholds, which validated both experiments.

Finally, a comparison between the approached method's results and the postulated research question is performed and backed up by metrics.

Conclusion & perspectives

Finally, we offer an overall balance on this study, *viz.* the steps taken to achieve the goal at hand, the scientific contribution proposed, the limits of our work, and the envisaged research perspectives (*cf.* Conclusion & Perspectives on page 286).

This concludes this Chapter which delved in the approached method in this thesis, closely following the THEDRE method (Mandran & Dupuy-Chessa, 2017) for RHCCS (*cf.* Chapter 5 above). We noted the expected productions, fulfilled them or linked them to the appropriate covered Section in this study, finally generating a flow measurement protocol (*cf.* Subsection "Flow measurement protocol"), which we apply in the Chapter 7 below, Experimentation.

Chapter 6. Trace analysis in a MOOC

This Chapter includes segments of a published article (Ramírez Luelmo, 2020) by this thesis' author. In this Chapter we present the basis for understanding the operational tasks performed in this research project, namely the notions of digital trace, and trace analysis. Both definitions are inscribed within the realm of Learning Analytics[114]:

> Learning analytics is the measurement, collection, analysis, and reporting of data about learners and their contexts, for the purposes of understanding and optimizing learning and the environments in which it occurs. (Siemens, 2013, 2011)

Digital trace

Before inquiring into trace analysis, it is necessary to understand the notion of a digital trace in educational contexts. In France, studies have shown common points when characterizing digital traces (Merzeau, 2013; Michel, 2015). Indeed, in one hand, Merzeau (2013, p. 38) proposed to differentiate between declarative, behavioral, and "calculated"[115] traces (*traces déclaratives* & *traces comportamentales*).

- Declarative traces consist of productions left online with such purpose, ranging from blogs, articles, reviews, online profiles, status updates, comments, messages, photos, videos, and/or personal information explicitly inputted, with many more examples in-between.
- On the opposite side of this spectrum, behavioral traces, while also being intentionally created by users, differentiate themselves in that their public availability is less obvious to the user, *e.g.*, search queries, shopping lists, browsing history, and/or geolocation data.

[114] *« L'analyse des traces d'apprentissages »* (Cherigny et al., 2020).
[115] Unnamed in the original source (2013, p. 38).

Merzeau (2013, p. 38) further details that in such cases, these traces are the result of a contract more or less formal, engaging the user to allow this data collection in exchange for a service.

- Calculated traces are quantitative variables automatically generated by online platforms from the information explicitly inputted, or inferred (calculated or aggregated) from the record of activities performed by the user, *e.g.*, publication frequency, number of contacts/friends/followers/likes/retweets, etc.

Similarly, in the other hand, Michel (2015, p. 19) discerned between automatic and voluntary traces (*traces automatiques* & *traces manuelles, ou volontaires*): automatic traces are collected during the usage of a system whereas voluntary traces are information inputted by the system's user during its activities, such as messages, documents, reviews, etc. (2015, p. 19).

Therefore, considering these fundamental similitudes, plus the definitions of what constitutes a digital trace (Djoudi et al., 2018, p. 4; Iksal, 2012, p. 28; Mille, 2013, p. 7), Pierrot (2018, p. 15) envisaged and employed a rather, compact, simpler, yet comprehensive definition of digital traces (*a.k.a.* traces) to which we adhere: the digital footprints left consciously or unconsciously in a computer environment, structured in such a way to allow their inspection (2018, p. 85), or simply « […] *les empreintes laissées de manière consciente ou non dans un environnement informatique* » (2018, p. 15). An illustration of a digital trace is shown in Figure 6-1, where a log file entry (*i.e.*, "document") is shown.

Figure 6-1

A Screenshot of a Digital Trace

Key	Value	Type
(1) ObjectId("60fc29f31bdb7861d0057821")	{ 15 fields }	Object
_id	ObjectId("60fc29f31bdb7861d0057821")	ObjectId
event_source	browser	String
referer	https://moocgdp.gestiondeprojet.pm/courses/course-v1:LearnGdP+MOOC-GdP...	String
username		String
ip		String
event_type	edx.bi.course.upgrade.sidebarupsell.displayed	String
host	moocgdp.gestiondeprojet.pm	String
accept_language	fr,fr-FR;q=0.8,en-US;q=0.5,en;q=0.3	String
page	https://moocgdp.gestiondeprojet.pm/courses/course-v1:LearnGdP+MOOC-GdP...	String
context	{ 4 fields }	Object
user_id	83266	Int32
org_id	LearnGdP	String
course_id	course-v1:LearnGdP+MOOC-GdP-TC+16	String
path	/event	String
event	{"courseRunKey": "course-v1:LearnGdP+MOOC-GdP-TC+16"}	String
agent	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:84.0) Gecko/20100101 Firefox/8...	String
name	edx.bi.course.upgrade.sidebarupsell.displayed	String
time	2021-01-01T11:59:06.783945+00:00	String
session	e809c33543494e20a4be10b4b39fd9d3	String
(2) ObjectId("60fc29f31bdb7861d0057822")	{ 13 fields }	Object
(3) ObjectId("60fc29f31bdb7861d0057823")	{ 13 fields }	Object
(4) ObjectId("60fc29f31bdb7861d0057824")	{ 13 fields }	Object
(5) ObjectId("60fc29f31bdb7861d0057825")	{ 13 fields }	Object
(6) ObjectId("60fc29f31bdb7861d0057826")	{ 13 fields }	Object
(7) ObjectId("60fc29f31bdb7861d0057827")	{ 15 fields }	Object
(8) ObjectId("60fc29f31bdb7861d0057828")	{ 13 fields }	Object
(9) ObjectId("60fc29f31bdb7861d0057829")	{ 15 fields }	Object
(10) ObjectId("60fc29f31bdb7861d005782a")	{ 13 fields }	Object
(11) ObjectId("60fc29f31bdb7861d005782b")	{ 15 fields }	Object
(12) ObjectId("60fc29f31bdb7861d005782c")	{ 15 fields }	Object
(13) ObjectId("60fc29f31bdb7861d005782d")	{ 13 fields }	Object

Note: In this screenshot, traces (individual log entries) are represented by yellow markers (on the left, uppermost level) as key (left column)/values (right column) pairs in a structured hierarchy.

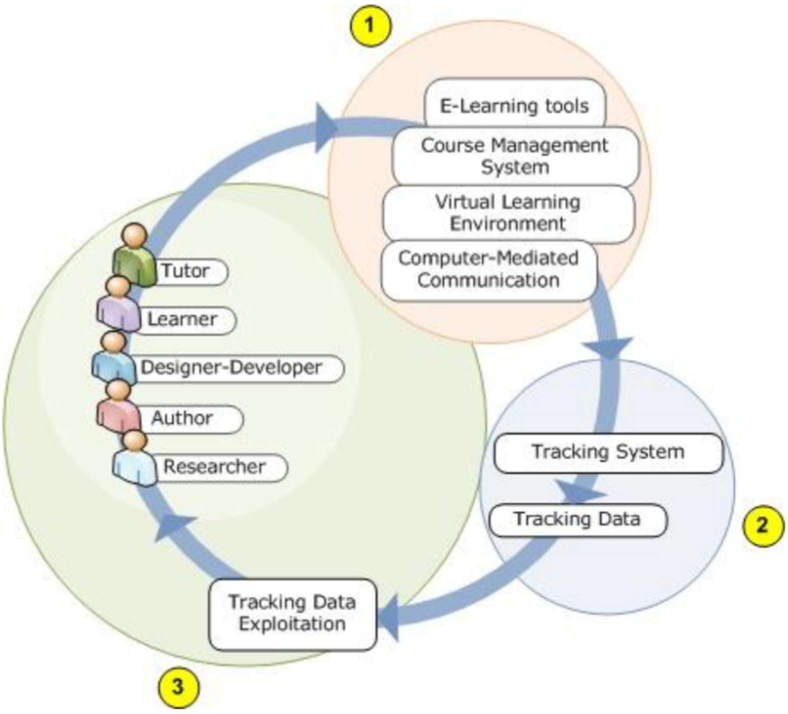
Data tracking

To analyze such traces, a 2009 PhD thesis (May, 2009, p. 26) proposed a general framework representing the life cycle of tracking data in a TEL, quoted as-is in Figure 6-2. It comprises the three following aspects, shown in yellow circles in the same Figure:

1. The TEL platform itself and a study focusing on what is to be tracked and why.
2. The use of a tracking system, focusing on how to observe learners interacting with the TEL, and how to produce relevant and meaningful tracking data. Most modern TEL systems contemplate a mechanism to track and store pre-determined tracking data.
3. Exploiting tracking data, involving operations ranging from simple computations and results display, up to multi-dimensional data analysis and data visualization (2009, pp. 26–27).

Figure 6-2

Example of Life Cycle of Tracking Data in TEL (May, 2009, p. 26)



This framework is exploited in the Iksal (2012, p. 32) HDR[116] thesis by acknowledging that the process of trace analysis is decomposed in four phases (Cherigny et al., 2020; Clow, 2012, 2013; Fayyad et al., 1996; Stamper et al., 2011), of which the last one concerns only TEL researchers:

1. Information collection, *a.k.a.* data collection. It is conformed usually of behavioral data saved in the TEL's databases, or under the form of separated, exported log files.

2. Data analysis. The most important phase of the process concerns the transformation of the trace from collected information to meaningful data for the TEL's users. It implies transformation, combination, enrichment, and/or calculation operations on the collected data.

[116] The "*Habilitation à Diriger des Recherches*" (HDR), "Habilitation to Conduct Research" or "Accreditation to Supervise Research" is the highest French education qualification diploma, obtained after a PhD (Aix Marseille Université, 2023; SciencesPo, 2018).

3. Analysis interpretation. Trace analysis provides knowledge on a given learning situation to come to conclusions. Thus, analyzed data interpretation is a must and should provide insights adapted to the final users of those interpretations.

4. Data and models capitalization. Data sharing should prevent *ad-hoc* tools development and instead facilitate higher-order research. (Iksal, 2012, p. 32)

Please notice the common points between such trace analysis processes and the Machine Learning workflows already explored in Section "Machine Learning workflow"; both basically necessitating a data collection stage, a data processing/analysis stage, and a result's analysis or evaluation. These commonalities facilitate the implementing of both methods when developing Machine Learning models based on trace analysis in a Learning Analytics context.

Data types, as observational data

Bent on the importance of point 4, the DPULS (Design Patterns for collecting and analysing [*sic*] Usage of Learning Systems) project (Choquet, 2005; Verdejo & Celorrio, 2005; Delozanne et al., 2005, 2007; cited by Ngoc, 2011, p. 53) proposed the principles of "capitalization" and "reutilization" (Delozanne et al., 2007, sec. 4; Ngoc, 2011, p. 53) of the know-how and experience of data collection/analysis' researchers/designers already employed in TEL projects.

Ngoc (2011, p. 53) tells us that the DPULS project proposed a typology of "observational data" [*données d'observation*] (2011, p. 53) defining data types[117] according to their usage goals, and origin (2011, p. 53). This typology (original by Choquet, 2005; employed by Choquet & Iksal, 2007, p. 4; employed by Ngoc, 2011, p. 54) is illustrated in the UML (Unified Modeling

[117] In our specific context, it is not to be confused with the most common notion (dating back to the mid 1950's) in Computer Science of data type: "a small list of possible types supported by the [programming] language" (Parnas et al., 1976, p. 149), implying a collection of possible data values, their allowed operations and expected behavior, and their most basic, internal machine representation (primitive). In such definition, a variable would be a primitive, and data types would be "equivalence classes of variables" (1976, p. 149).

Language)[118] diagram (Booch et al., 2005) depicted in Figure 6-3: it shows that Primary and Derived Data are extensions of the Data class, Derived Data is calculated via other primary data or other derived data itself, and where Primary Data is not calculated, but instead it can be:

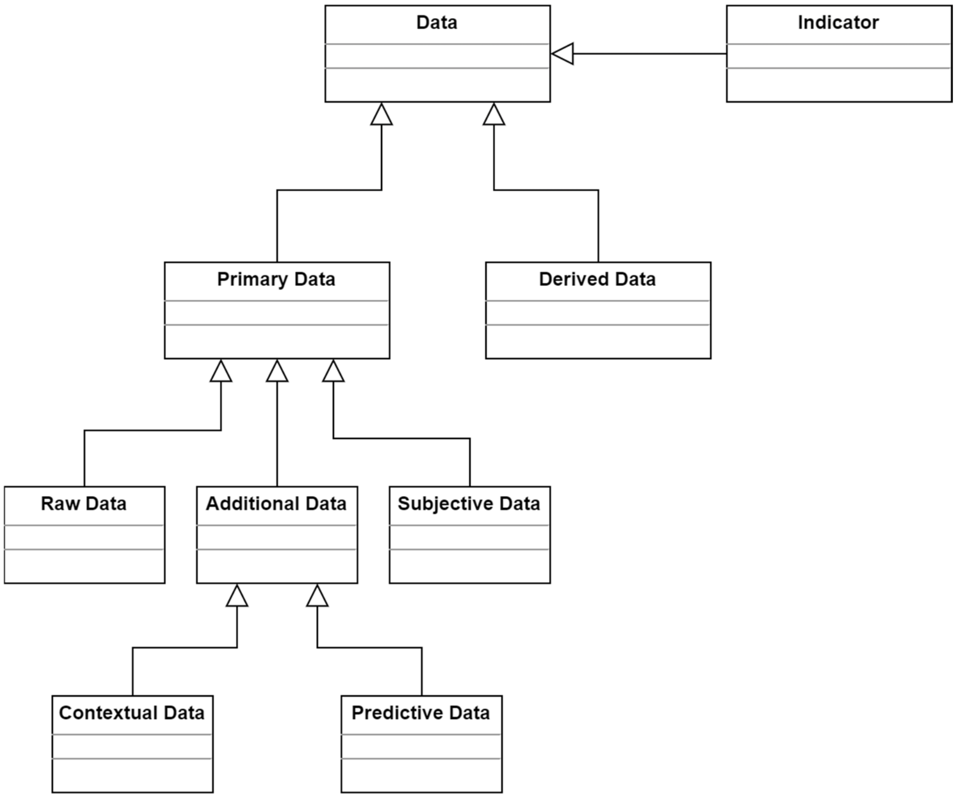
- Raw Data, collected before, during, or after the learning session.
- Additional Data, data employed to set up a derived data, which in turn can be:
- Contextual Data, meta-data describing the learning situation, available before the learning sessions.
- Predictive Data, data produced during the learning situation by the actors involved.
- Subjective Data: data set up by the session analyst. (Choquet & Iksal, 2007, p. 4; Ngoc, 2011, p. 54)

These classifications matter because they help distinguish, among other things, the source of data (*i.e.*, the system itself, the learning actors, or the session analyst), the amount of processing carried out in each data type (*e.g.*, raw data is the less processed type of data, and one could argue that primary data is the most processed type), and the moment any specific data exists (*e.g.*, before or after the learning session).

[118] In a UML class diagram, arrows connecting elements (B →A) are to be read as "B inherits from A", and not as a sequence.

Figure 6-3

UML Observational Data Typology (Choquet, 2005)



Note: This typology was first proposed by the DPULS project (Choquet, 2005) and subsequently retrieven and translated by other authors (Choquet & Iksal, 2007, p. 4; Ngoc, 2011, p. 54; Ngoc et al., 2009, p. 682).

Indicators in TEL

Among these data types, the notion of indicator stands out from the typology:

[An indicator is] a variable that describe 'something' related to: (a) the mode or the process or the 'quality' of the considered 'cognitive system' learning activity (task related process or quality, (b) the features or the quality of the interaction product and/or (c) the mode, the process or the quality of the collaboration, when acting in the frame of a social context forming via the technology based learning environment. (Choquet, 2005; cited by Ngoc, 2011, p. 30; translated by Ngoc et al., 2009, p. 682)

More recently, the Iksal (2012, p. 29) HDR thesis proposed to consider an indicator as a meaningful pedagogical variable, set up or calculated via observed data, qualifying the interaction, the activity or the learning in a TEL, defined by an observational goal, and pedagogically motivated:

> *Un indicateur est un observable significant sur le plan pédagogique, calculé ou établi à l'aide d'observés, et témoignant de la qualité de l'interaction, de l'activité et de l'apprentissage dans un* EIAH. *Il est défini en function d'un objectif d'observation et motive par un objectif pédagogique. (Iksal, 2012, p. 29)*

In our specific flow measurement context, the indicator carries the meaning of flow dimensionality. A dimensionality specifically designed in the flow theoretical model as a flow dimension pedagogically relevant and observable in educational contexts (*cf.* Subsection "Flow in Education (EduFlow & EduFlow-2)" above).

To avoid terminology confusion, please notice that the notion of indicator in the present context of trace analysis differs from the one seen and employed when discussing the THEDRE method (*cf.* Chapter 5 above): in trace analysis an indicator is a pedagogical describing variable whereas in the THEDRE method they are akin to goals or milestones.

Hybrid approach: Hypothesis-guided & Machine Learning

Concerning the analysis of digital traces (*a.k.a.* trace analysis) research has shown that it can be approached from two complementary perspectives. First, the hypothesis-guided, « *guidée par l'hypothèse* », approach, meaning the trace analysis is to corroborate (or refute) a pre-defined hypothesis, theory, or model, implying it is known since the beginning what it is to observe (Iksal, 2012; Pierrot, 2018). The inverse approach; data mining attempts to extract knowledge, information, or a meaningful data structure from the traces (2012, p. 35; Peraya & Luengo, 2019, pp. 1–2).

In this thesis we employ a hybrid approach benefitting from both perspectives:

- In one hand, Machine Learning plays a role regarded by some authors (Iksal, 2012, p. 35; Peraya & Luengo, 2019, pp. 1–2) as Data Mining's; to extract meaningful data structures from the traces (*cf.* Subsection "Scope and importance" above), or as Pierrot explains it: « *[…] de découvrir et faire émerger des éléments « caches » dans les traces, dans le sens où on n'a pas une representation claire de l'information recherché* » (2018, p. 87).

- In the other hand, our hypothesis relies upon a sound, proven, theoretical model (*cf.* Flow in Education (EduFlow & EduFlow-2)), and two measuring scales, describing what we intend to observe (flow), and how (via its defining dimensions in educational contexts), while considering additional suggestions when designing a flow measurement protocol (*cf.* Flow measurement protocol).

This hybrid approach is an additional asset to this research work when combined with the authenticity (real human answers, *cf.* Questionnaire application), and size (*cf.* Data filtering and cleaning) of our input data sample, solidifying this thesis' scientific contribution.

Trace aggregation: classification of entries

Researchers (Cisel, 2017; Iksal, 2012; Pierrot, 2018; Pierrot et al., 2017; Poellhuber et al., 2019; Slouma et al., 2019) have tackled the issue of analyzing digital traces in educational contexts by aggregating them and analyzing the aggregated data forms.

More specifically, they have re-classified and grouped the TEL's developer types of logs into some form of Primary data, carrying meaningful human information, *i.e.*, adding Contextual, Predictive, and Additional data to Raw data to conform Primary data (*cf.* Figure 6-3).

To achieve this, log entries are equated to activity traces (Pierrot, 2018), considered as any trace of the course usage on the hosting platform, « *toute trace de l'utilisation du cours sur la plate-forme d'hébergement* » (Cisel, 2016, p. 390), closely linking the hosting system and the broader definition of digital trace.

Subsequently, analysis and results on the so-created Primary data carry and provide human meaning in the specific context in which they were created, *e.g.*, hypothetical Primary data "number of posts per session" could be obtained by combining the notion of a session (Subjective/Contextual data) alongside the filtering of relevant (Subjective data) posts (Predictive data), in any given forum thread (Contextual data).

Thus, in this example, a hypothetical numerical value "0.8" taken by the Primary data "number of posts per session" now carries the extremely specific, human meaning of "0.8 is the average number of posts containing the text string 'sea' in the post thread 'Global Warming' as replies to other posts (not creating new posts) within a 15-minute interval, without disconnecting from the platform", for instance.

This hypothetical Primary data could itself be employed as the basis for a hypothetical binary indicator "participant engaged in a conversation". In this case, the Primary data value is compared to an empirically deduced threshold to finally conclude on a 0, or 1 (yes/no) value for such indicator, *e.g.*, if Primary data "number of posts per session" > 2 then Indicator "participant engaged in a conversation" = 1, otherwise = 0.

Thus, by creating Primary datum from the aggregation and combination of other data types, and continually contrasting them to predetermined values, indicators are created, which:

- carry human meaning, qualifying the interaction and defined by observational and pedagogical goals,
- are dimensionally simpler to analyze, instead of considering five or more-dimensional data at once,
- and consequently, yield human meaning results.

Nevertheless, despite the efforts carried out to provide meaning to aggregated and combined data we insist that during any direct phenomena observation, the sensors (Di Mitri et al., 2018) employed for measurement, whether they constitute a fixed-length stick, or a particle accelerator, lack the ability to make interpretations or to assign meaning to the data collected, or computed.

This remains particularly true in the contexts of Machine Learning for Learning Analytics: the learner's cognitions, emotions, beliefs, motivation, or even learning outcomes are latent attributes that can only be inferred and not measured directly (Di Mitri et al., 2018), a statement perfectly aligned with flow measurement research (*cf.* Subsection "Measurement Attempts of Flow" above).

Moreover, the extremely recent study paper concluded that "mouse click frequency alone cannot be used to predict the flow experience" (Muramatsu et al., 2023, p. 1288), strongly suggesting the need for multi-dimensional approaches[119] when measuring flow in any TEL system, and thus confirming the suitability of our multi-instrument, multi-dimensional, hybrid trace analysis, flow measurement approach.

Therefore, we conclude the present Chapter on Data tracking, where we reviewed the notions of digital traces, along data types and data tracking. We also presented the existing approaches for trace analysis, and the currently employed applied method, based on data aggregation and re-categorization. This is the primary method employed in this research project when approaching trace analysis, detailed in 0 below.

We also conclude Part II – Method, primarily driven by the THEDRE research method (Mandran & Dupuy-Chessa, 2017), extensively detailed on Chapter 5. Additionally, Part II covered the basis behind trace analysis in a MOOC, which are the ultimate main data source employed during experimentation (*cf.* Subsection "Logs data filtering, cleaning, and aggregating").

The following Part III concerns our proposed approach, *i.e.*, the execution of the planned phases detailed on Chapter 5, fully illustrated in Figure 5-1. We cover a tri-partite selection: the

[119] *Cf.* also "Multimodal learning" (Shani et al., 2023).

flow measurement instruments, the experimentation terrain, and the Machine Learning method to implement our approach.

Part III

Our proposed approach

*Our proposed approach

Part III is concerned with the execution of the Experimentation and Evaluation macro tasks shown in Figure 5-1, comprised in a single Chapter 7 Experimentation.

Based on the appropriate indicators and research needs detailed in Chapter 5 above, we first determine the flow measurement instruments to employ, and we justify this selection (*cf.* Section "Flow measurement instruments identification" below). Second, we determine and justify the selection of the research terrain to this study (*cf.* Section "Terrain: the MOOC "*Gestion de Projet*"" below), and third, that of the Machine Learning method (*cf.* Section "Towards a flow-detecting Machine Learning model" below).

Then, we implement these three elements into a two-part experimentation consisting of a Proof-of-Concept, reviewed in Experiment 1 (Experiment 1 – Proof-of-Concept), aiming to verify the proper working of these elements when paired together, and of a Prototype, reviewed in Experiment 2 (Experiment 2 – Prototype), materializing a flow-detecting model which ultimately facilitates answering our proposed research question (*cf.* Introduction).

Also, both experiments share a common phase of data collection, covered in detail in Data collection below.

Chapter 7. Experimentation

This Chapter extensively details the two proposed experiments designed to approach our research question.

It includes a common data collection phase to both proposed experiments (*cf.* Section "Data collection" below), which then follow as a Proof-of-Concept, and a Prototype named experiments. Each comprises their own goals and resulting metrics thresholds, detailed in Sections Experiment 1 – Proof-of-Concept and Experiment 2 – Prototype, respectively. Also, specific discussions and conclusions are addressed by each experiment.

We begin this Chapter by determining the three elements required for the Experimentation macro task depicted in Figure 5-1: the most-adapted flow measurement instruments (covered in Section "Flow measurement instruments identification"), the research terrain (covered in Section "Terrain: the MOOC "*Gestion de Projet*""), and the Machine Learning method (covered in Section "Towards a flow-detecting Machine Learning model"). We point out that this Chapter features text from published articles by this thesis' author (Ramírez Luelmo, 2022; Ramírez Luelmo, El Mawas, Bachelet, et al., 2022).

Flow measurement instruments identification

The present Section presents the rationale leading us to select two flow measurement instruments (and their accompanying scales and theoretical models, if any), based on their characteristics face to our research needs, and the instruments themselves. We pool them from the previously described flow measurement instruments (and their scales) historically employed in educational contexts (*cf.* Section "Flow in Educational Contexts" above).

Firstly, we review the considerations to account for when designing flow measurement protocols previously explored in Section "Considerations When Measuring Flow" above. Then, we confront them with the surveyed flow scales, with our constraints, but mostly with our research needs. This successive confrontation excludes scales until two candidates remain: Flow-Q and EduFlow-2.

Constraints: research needs & considerations

We aim to detect flow in a MOOC, *i.e.*, an educational, online, distant setting, accessed at the time, place and frequency of choosing of a multitude of participants. Chosen flow measurement instruments, along the research terrain, should respect the following constraints:

- Ideally, designed with educational contexts in mind.
- Reliable, proven to effectively measure flow.
- Not interfere with the participant's flow state, *e.g.*, post-event measures.
- Adapted to online, distant settings, accessible without staff intervention, *i.e.*, not requiring tasks such as "by raising hands …", "ask participants to step forward …", etc.
- Adapted to adults, *i.e.*, not children and teenagers.
- Cost effective, *i.e.*, not requiring additional devices (pingers, beepers, glasses, helmets, gloves, sensors of any kind), infrastructure (*ad hoc* headquarters, additional software purchases, servers' setup and maintenance), staff (survey team, specialized engineers), adaptations (children → adults, basic education → high education, work domain → education domain), translations (Korean → French, German → French), etc.
- Ideally, designed and validated in French, or at most drafted for a nonnative English speaker audience.
- Be considered a "short" instrument (by the flow research community).

Furthermore:

- At least two different types of flow measure instruments should be employed: unidimensional (*a.k.a.* a-dimensional), and multidimensional.

- The application and data collection processes associated to each instrument should be reliable and straightforward, given the research context, *i.e.*, not necessitating *in-situ* distribution, nor a non-standard protocol/algorithm/data format (*cf.* footnote [117] on data type on page 191 in Chapter 6 above).

Flow measurement tools confrontation

Being the most popular flow measurement instrument (*cf.* Appendix 1. – The Experience Sampling Form) employed in research since its inception, we dedicate a prominent space to the ESM (*cf.* Subsection "Experience Sampling Method (ESM) and the ES Form" above).

Despite its scientific success, the ESM has proved to pose some issues (Magyaródi et al., 2013, p. 87; Moneta, 2021; Nakamura & Csíkszentmihályi, 2009; Rheinberg et al., 2003):

- Firstly, it can be intrusive, requiring the participant to wear an electronic device at all/specific times, depending on the research protocol (Jackson et al., 2008, p. 562), and because of its dependence on self-reports, it could unwillingly expose private and/or sensitive information[120] (Magyaródi et al., 2013, p. 87).

- Another disadvantage is its costly execution (Magyaródi et al., 2013, p. 87), which not only includes the devices themselves' nominal and operating costs, effectively limiting the number of devices and thus, of participants, but also the costs of a minimal operating training for the participants.

- On the conceptual area, the prevalent ESM experience has been shown (Moneta, 2021, p. 33) as somehow "imposing" flow on the participants instead of letting them "explicitly to report whether or not they experienced flow at the time they were beeped" (2021, p. 33).

[120] Maybe even illegal activities.

Steps taken to overcome a few of these issues such as the Day Reconstruction Method (Kahneman et al., 2004) still require further development (Moneta, 2021, p. 30).

Albeit being a shorter scale, the FKS is meant to be applied in tandem with the ESM method, which immediately limits its use in our research context.

Just like the ESM, the FSS2 and the DFS2 suffer from a lengthy reporting process, comprising many items. Such many questions can demotivate individuals when answering the measure instrument, leading to inconsistencies in reporting or simply not enough participants.

Measurement instruments such as the Flow Scale and the Learning Flow Scale are still in early adopting stages. We could not find any validation source for the FS in the literature, nor the scale itself. Furthermore, the FS was specifically designed with sports and psychotherapy in mind, a context too distant from our own. The Learning Flow Scale suffers from a poor adopting rate, being unknown outside of Korea and only existing in Korean.

While the Flow in Human-Computer Interaction has been historically employed to study flow in educational contexts (Heutte, 2019), its specific experimentation and validation context (work-related and computer use) partially (distant, online) match our own research context. Indeed, while some MOOC learners enroll because of work reasons (*cf.* Chapter 3 above), it cannot be said so of all participants while learning goals do generalize.

The WOLF-S initially seemed like a great candidate for measuring flow in educational contexts (university & high school settings), being a multidimensional, short measurement instrument. However, modeling flaws have been pointed out in its parent model, WOLF (covered on page 82). Furthermore, it has not been validated in French. Thus, remodeling and language-validating the measurement instrument remains well outside the scope of this research project.

The EduFlow-2 measure instrument was designed (and proven) specifically for educational contexts, in MOOC (online, asynchronous, distance learning) and classroom (offline, synchronous, presential learning) situations. It differentiates dimensions relevant to cognitive processes, while remaining a short (12-items) scale, also available in French.

Flow-Q is dimension-agnostic, *i.e.*, general-purpose flow detection measurement, comprising only three items, taking a binary scale, and returning a binary score. It is universally admitted by flow researchers as a valid and proven flow measurement instrument. Additionally, Flow-Q has been translated and validated in French.

Thus, thanks to the literature review seen in Section "Flow in Educational Contexts", we have identified two measurement instruments adapted to our research context which also respect the research needs: the EduFlow-2 and the Flow-Q measurement instruments and their respective scales.

Then, we identify the experimentation terrain for their application. The following Section presents the MOOC "*Gestion de Projet*", a French MOOC with sufficient learners' attendance (accounting for the MOOC drop-out rate), already in-place research protocols (described on page 212), and an experienced administrative staff, among other qualities.

Terrain: the MOOC "*Gestion de Projet*"

The French MOOC « *Gestion de Projet* » (Project Management, "GdP" for short) was launched in 2013 by Rémi Bachelet, within the realm of the École d'Ingénieurs Centrale Lille[121]. It is the first certified xMOOC in France (Delpeyroux & Bachelet, 2015, p. 2; Verzat & Bachelet, 2020, p. 51).

As of November 2023, this online learning platform[122] had 335 360 enrolments, among which 62 651 students[123] fully completed either the basic or the advanced tracks. Half of the active learners enroll through their university, while the other half do so of their own will, with one of the best completion rates in the francophone world (Chermann, 2020).

[121] https://ecole.centralelille.fr/
[122] https://mooc.gestiondeprojet.pm/
[123] http://bit.ly/2LozEFI (URL shortening provided and maintained by the MOOC GdP staff)

While the MOOC can be inscribed within the École's own cursus if students are enrolled by their professors, the subject interests professionals as well: special sessions dedicated to the enterprise world (Bachelet, 2019; Chermann, 2020). As such, the MOOC's proposed training on Project Management is free if undertaken on a personal note but incurs a fee for universities and private companies. For the former, 1 and 2 ECTS[124] credits are delivered for the basic and advanced tracks respectively, within the MOOC agreement with the Centrale Lille university (Verzat & Bachelet, 2020, p. 57). Like many other MOOCs business models, the MOOC's fee-based offer goes from a certification upon successful completion, up to a certified, webcam-proctored exam (Verzat & Bachelet, 2020, p. 54). The proctored exam relies on the PSI RPNow[125] validation system. This proctored assessment constraints the learner to access documents or any other kind of external help while providing additional time to compensate for the eventual technical difficulties.

Because of this target public duality, the MOOC GdP benefits from double categorization. In one hand, it can be seen both as a specialist MOOC, having as a primordial pre-requisite to be inscribed in the Centrale Lille (superior education) and thus, targeting university students. In the other hand, it can also be seen as a profane MOOC; with no precise pre-requisites and open to all public (Vrillon, 2017) when aimed at the individual level.

In a similar fashion, it can be both categorized as a Lecture-MOOC because of its embedding within the École Centrale Lille, with ECTS deliverance, proctored exams, and on-campus students' exchanges, and as a conventional MOOC, when targeting thousands of online user worldwide (Ebner et al., 2020).

Furthermore, according to the targeted discipline domain, it is undoubtedly an intermediate, specialized MOOC, demanding between two to four hours of weekly estimated effort and focused in the Law, Economics and Management fields (Vrillon, 2017).

[124] https://www.study.eu/article/what-is-the-ects-european-credit-transfer-and-accumulation-system
[125] https://systemcheck.rpexams.com/

If we consider the previously reviewed MOOC categorizations (Gilliot et al., 2013; Pilli & Admiraal, 2016), we can conclude that the MOOC GdP is mostly a quadrant-IV type of MOOC (more open than closed and distributed at a large scale), with an obvious leaning towards being an iMOOC instead of an xMOOC because of its successful dual public targeting requiring more openness all variables considered, as we show in Table 7-1:

Table 7-1

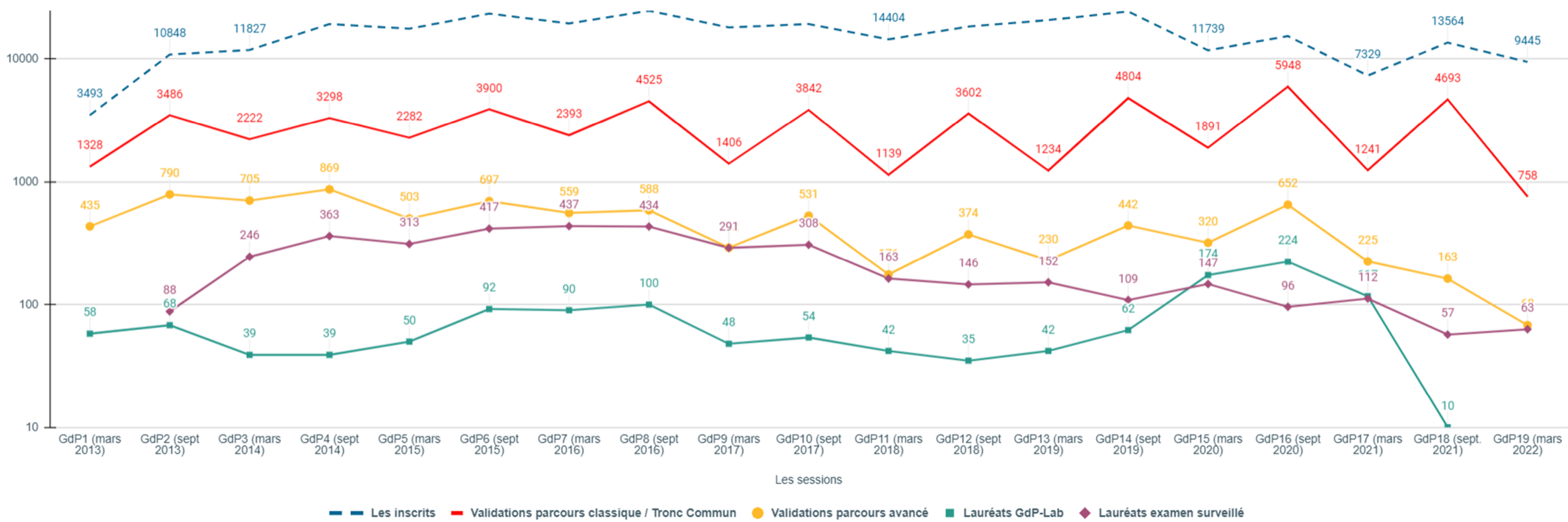
MOOC GdP's Pedagogical Openness/Closeness Categorization

Pedagogical dimension	MOOC GdP
Learning goals	O
Choice of resources	C
Organization of the learning activities	O
Organization of the group's work	O
Collaborative co-production	O

We plotted in Figure 7-1 the number of inscribed MOOC participants (blue, dashed, first top line), those who validated the common branch and the basic track (orange, continuous, second top line), those who validated the advanced track (yellow, continuous, third line), and those who succeeded the final, proctored exam (violet, continuous, fourth line). A green line depicts those completing the additional GdP Lab, not described here. Data for this plot goes back to the MOOC's inception in 2013 up to March 2022.

Figure 7-1

MOOC GdP Participants' Evolution



Organization

Organizationally speaking, the MOOC GdP programs two sessions per year: 1st session spans the September – November period and the 2nd session comprises the March – May period (although precise dates vary each year). It is noteworthy to highlight that historically, the number of participants validating the first half of the MOOC has been consistently larger during the 1st session compared to the 2nd session, with over a 110% increase (Bachelet, 2019), *i.e.,* the period of September – November encompasses more validating learners for the common branch than during the period March – May. This is better appreciated in Figure 7-1.

Each session is comprised of nine weeks plus an initial pre-opening week which does not count to the week numbering, usually employed for getting used to the MOOC's tools and the community; the other learners and the trainers. This weekly organization is represented in the first column (Week) of Table 7-2.

The MOOC unlocks pedagogical modules (or units) every week automatically, on a Monday's noon, although participants can freely choose to achieve it at their own pace. However, to successfully complete the MOOC, learners must finish at least the common branch by the end of the 9th week. The common branch is also known as the first part, the first half or the first moment of the MOOC, with the remaining of the MOOC known as the second part, second half or second moment. Table 7-2 shows in its second column "Modules unlocked" an example of the units being unblocked every week in the basic track ("+" indicates an additional module).

Table 7-2

Example of Unlocking Modules for the Basic Track

Week	Modules unlocked	Description
0	Personal presentation	Get acquainted with the MOOC
1	+Project Management fundamentals	Common branch (minimum for certification)
2	+The essentials of Project Organization	
3	+Advanced tools	
4	+Risk Management	

4-7	Certification exam		Free certification exam, not proctored
5	+Chosen module A	+Chosen module B	Choice of at least two specialization modules among 15 modules proposed
6			
7			
8			
9	Final exam		Proctored exam
…	All previously seen modules		Previously accessed modules do remain accessible

Note: The unlocking mechanism is per calendar week and not per individual work time (in weeks).

The MOOC offers two possible tracks: basic and advanced. The basic track comprises the Common branch (four modules) plus at least two out of 15 specialization modules plus the free certification exam.

Table 7-3

MOOC GdP's Basic Track Organization

Week / Content	1	2	3	4	5	6	7	8	9
Common branch									
Free Certification exam									
Two specialization modules									
Final proctored exam									

Note: Darkened areas show suggested scheduling.

The advanced track comprises the basic track plus a Case study comprising the delivery of three products, three peer evaluations on other learners' products (and thus, receiving three peer evaluations), and three self-evaluations, in addition to the final proctored exam. Examples of what three products are tasked to be delivered could be a Conceptual map, a Planning, or/and a Meeting's report. Both tracks are depicted in Table 7-3 and Table 7-4, respectively (accounting only the main 9 weeks); the advanced track Case study is shaded differently to mark the difference between the basic and advanced tracks.

Table 7-4

MOOC GdP's advanced track organization

Week Content	1	2	3	4	5	6	7	8	9
Common branch									
Free Certification exam									
Two specialization modules									
Case study	P1	P2	P3						
		P1:PE	P2:PE	P3:PE					
			P1:SE	P2:SE	P3:SE				
Final proctored exam									

Note: P1 refers to Product one; PE means Peer evaluation, and SE means Self-evaluation.

Content pedagogical configuration

Each pedagogical module is composed of six chapters, each with one video of about five minutes time length. Videos are subtitled and transcribed, providing a handy jump-start functionality to any line in the transcription text. They also can be watched in full screen, at different speeds (0.5x – 2.0x, in increments of 0.25), with/without subtitles, in High Definition or in Standard Definition, as well as freely searchable by the progress bar.

Although the videos are stored on the YouTube[126] video platform and are publicly accessed (provided the video URL), they are not searchable via the own platform's tools. It is noteworthy to highlight that the corresponding YouTube channel ("*Gestion de projet*"), despite

[126] https://www.youtube.com/

featuring no videos, no playlists, no community exchanges, and no intra-channels, accounts for almost two thousand subscribers.

A training questionnaire follows each video where the answers can be obtained immediately upon request. Forum participation, given activities (non-graded assignments) and synthesis facilitate practicing, reviewing, and exchanging with the other participants.

At the end of each week, an evaluation allows for that week's badge. Once the common branch has been successfully completed, a free, non-proctored exam is proposed. If this exam is successfully passed, the option of purchasing a certification is given. Participants having the French status "jobseeker" can apply to be exonerated of this fee upon proof. An additional way of obtaining the certifications for no fee is proposed, by participating to unrelated-to-the-MOOC online research surveys.

The MOOC's organization team preconizes daily working sessions of about 20 minutes (about 2h20 of estimated weekly effort).

As a research platform

The MOOC GdP comprises a Research & Development team that precedes the work described in this thesis (Verzat & Bachelet, 2020, p. 56). This team accounts Rémi Bachelet (University Lecturer), François Bouchet (University Lecturer), Rawad Chaker (University Lecturer) and Jean Heutte (University Professor), among others. Together, they develop and implement research instruments revolving around research subjects such as motivation, engagement, attrition, etc. in MOOCs. This research organization fulfills human-in-the-loop decisions-making to strict social values such as fairness, and equality (Chancellor, 2023, p. 78).

The Research & Development team from the MOOC GdP allows for three distinct periods (P1, P2, and P3) for measurement instruments application. This allows for evolution observation, often required when applying consecutive psychometric measure instruments to the same participants. Because of administrative reasons and to alleviate cognitive burden, these application periods are fixed and cannot last longer than those shown in Table 7-5, with their distribution remaining the same for all seen sessions.

Table 7-5

MOOC GdP's Measurement Instruments Application Periods and Contents

Periods \ Week	0	1	2	3	4	5	6	7	8	9	…
P1	Demographics										
P2				Measurement instruments							
P3					Measurement instruments						

Demographic data (sex, birth year, country of residence, occupation and highest academic degree obtained) are surveyed only during P1, *i.e.,* from the beginning of week 0 until the end of week 4.

Finally, from a technical point of view, the MOOC GdP is hosted by the MOOCit[127] platform, itself offering its services on the Open edX[128] Learning Management System, which is essentially the heart of the MOOC GdP: "The LMS is the most visible part of the Open edX project" (edX Inc, 2023). The hosting solution for the entirety of the project is Amazon Web Services[129].

The Open edX's general documentation[130] provides basic information on the platform while the EdX Research Guide[131] offers more in-depth information on the inner workings of the Open edX Learning Management System, *viz.* the description of the JSON field names of students' events stored in the tracking logs[132].

It is important to highlight here that, because of security reasons, we had no access to the Event Tracking API[133], nor to the Analytics module of the MOOC GdP, nor to code deployment

[127] https://moocit.fr/
[128] https://openedx.org/
[129] https://aws.amazon.com/fr
[130] https://docs.openedx.org/en/latest/
[131] https://edx.readthedocs.io/projects/devdata/en/latest/index.html
[132] https://edx.readthedocs.io/projects/devdata/en/latest/internal_data_formats/tracking_logs/index.html
[133] https://edx.readthedocs.io/projects/edx-developer-guide/en/latest/analytics.html

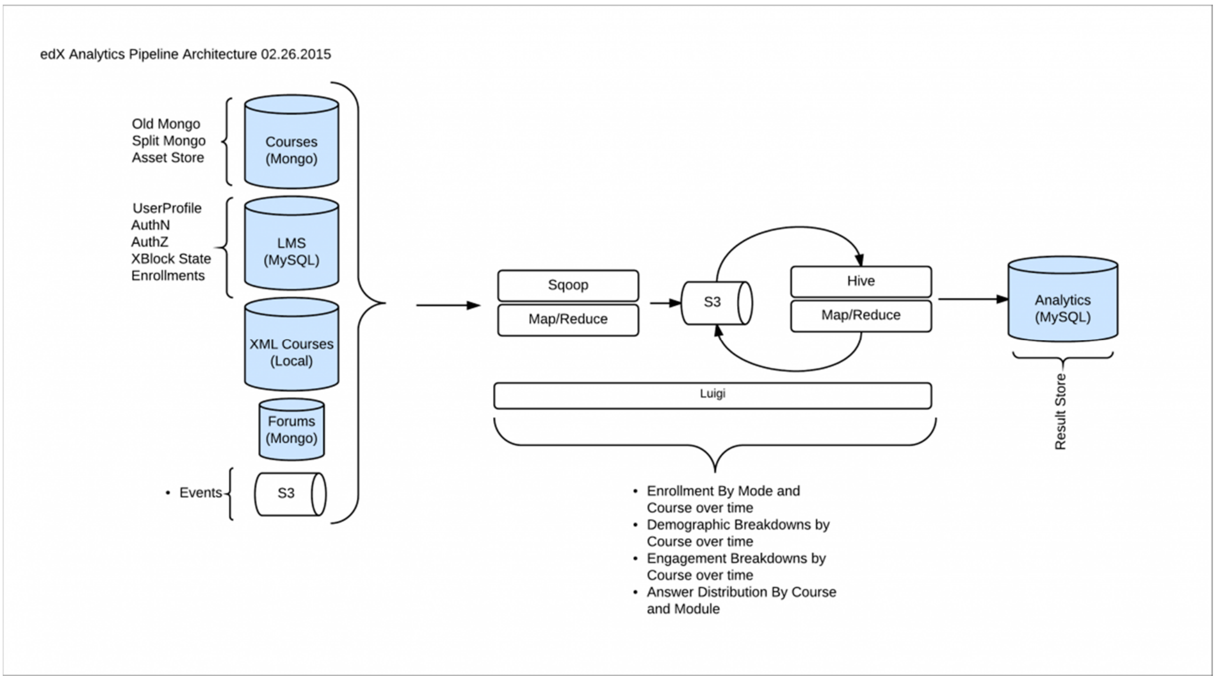
to exploit the Event Tracking library. Still, lack of access to neither of these tools did not affect the ultimate course of this thesis, as our collected log data (*cf.* Subsection "Traces: log data" below) proved to be a cold store ultimately originated from the Event Tracking API. In the case of the Analytics module, the predetermined, already aggregated, processed data it provides is unrelated to our research context and well too distant from its raw data status. Lastly, the Event Tracking library is still under development at the time of writing of this study and so far, it can only emit and track developer's own *ad hoc* events, which renders it useless for our research purposes.

Locations for the cold store Event Tracking API and for the Analytics module are illustrated in Figure 7-2, where the Analytics module lies at the right end of the Figure, on a MySQL[134] database, and the log events are stored in an Amazon Simple Storage Service (S3) bucket storage resource (in white) on the bottom left of the Figure. The log events are stored in JavaScript Object Notation[135] (ecma, 2017) (JSON) documents, themselves stored into log files (digital traces).

134 https://www.mysql.com/
135 https://www.json.org/json-en.html

Figure 7-2

edX Analytics Pipeline Architecture 02.26.2015 (edX Inc, 2023)



As we have shown in Section "Terrain: the MOOC "*Gestion de Projet*"" above, the MOOC *« Gestion de Projet »* successfully fulfills the applicable constraints and research needs mentioned in Subsection "Constraints: research needs & considerations" and those on the flow measurement protocol shown in Subsection "Flow measurement protocol" as well, notably accounting a reliable learners' attendance, not necessitating additional staff nor tools, accounting for a Research & Development team experienced and firmly established, with measurement protocols and tools already set in place.

Furthermore, considering the know obstacles to flow (*cf.* Subsection "Known Obstacles to Flow" above) and understanding the above mentioned conditions in which the MOOC GdP is held, we can state that, up to the extent of our knowledge, all participants in this study seemingly were in condition to experience flow.

In this Section we presented this thesis' research terrain, the MOOC GdP. We also covered its organization, and pedagogical configuration, demonstrating to cover all our applicable constraints and research needs.

Once the flow measurements (Flow-Q & EduFlow-2) and the research terrain (MOOC GdP) identified, the following Section covers the still-missing Machine Learning method detection component, and thus it details how these three components will be mobilized towards obtaining a flow-detecting Machine Learning model.

Towards a flow-detecting Machine Learning model

This Section covers the last component remaining to identify, the Machine Learning method, while reviewing relevant phases of a Machine Learning workflow: the training and production phases. Moreover, this Section explains the reason behind splitting the Experimentation into Experiment 1: Proof-of-Concept, and Experiment 2: Prototype, describing both experiments' purpose and sequencing. Finally, based on the chosen flow measurement instruments theoretical models, we construct indicators susceptible to be found in the MOOC learners' traces.

We begin by reminding the reader that our goal is to detect flow in a MOOC using a Machine Learning trained model by employing uniquely the learners' digital traces, without requiring any action that would be conducive to interrupting a potentially existing flow.

Machine Learning method detection

To extract meaning of input data, we employ Machine Learning to systematically discern in a multidimensional space non-evident similarities and thus, to classify participants into two classes: class 0 and class 1.

In our specific research context, we assign the meaning of flow to the '1' value and no-flow to the '0' value. This meaning assignation follows closely the choice of the flow measurement instruments and their consequent flow models. Indeed, the Flow-Q scale, universally recognized in the flow researchers community (*cf.* "FlowQuestionnaire"), renders a binary flow score of flow presence or flow absence, while the EduFlow-2 measurement instrument (*cf.* "Flow in Education (EduFlow & EduFlow-2)") supports both the flow-continuum, and the binary flow presence/absence conceptual flow model frameworks, of which we choose the binary form to set both instruments to the same scale[136].

We take a brief pause here to mention that, with the aim of avoiding repetition, we employ the following pairs of terms as equivalent, all being correct within the chosen flow theoretical models:

- 1 – flow & 0 – no-flow, as in "state" (preferred)
- 1 – flow presence & 0 – flow absence, as in "state"
- 1 – having flow & 0 – not having flow, when referring to individuals
- 1 – in flow & 0 – not in flow, when referring to individuals

Among the many Machine Learning methods, Multi-variate Logistic Regression is a Machine Learning technique adapted to the needs and constraints of our research context, namely:

- Being a Supervised learning method, Logistic Regression requires labelled data (*a.k.a.* the known target) to be included in the dataset. In our case, flow absence/presence is determined mostly by the Flow-Q measure instrument, an all-or-nothing flow measurement instrument.
- Our Machine Learning target is also binary (flow absence/presence) and Logistic Regression is a binary classifier. Moreover, the final classification depends on a statistical class membership probability prediction ranging from 0 to 1 in a mathematical continuum, *i.e.*, if conceptually mandatory, flow detection can then be easily switched from a binary detection to a continuum detection.
- Multi-variate Logistic Regression admits classification with more than one independent variable, and in the following experimentations we present either four or 23 independent

[136] As in range, spectrum.

variables: the EduFlow-2 measure instrument four dimensions, or the indicators issued from the traces aggregated data.

- Logistic Regression is not a black box model, still human-inspectable and human-understandable, *i.e.*, the weights and biases of the model are easily accessible and not overwhelmingly problematic to verify mathematically (*cf.* Chapter 4(i) above), even for a non-expert. This is particularly noticeable when Logistic Regression is compared to other much more complex Machine Learning techniques (*e.g.*, Neural Networks, ensembles, or Deep Learning).
- Logistic Regression is a computational simpler Machine Learning model than other Machine Learning techniques, while still offering adequate prediction performance, *e.g.*, does not consume significant amounts of power, nor requires long training times, nor costly software or specialized hardware.
- Logistic Regression is easily updatable if incoming data changes in shape (additional dimensions, *e.g.*, further measurements) or in quantity (number of participants).

Thus, for our experimentation, we employ Logistic Regression as the final classifier in our designed Machine Learning pipeline.

In compliance with the upcoming European Artificial Intelligence Act (Artificial Intelligence Act, 2021, sec. 5.2.4), we acknowledge the transparency obligations our proposed Machine Learning model might face, given its intended purpose of recognizing an emotional state through automated means.

Experimentation sequencing

Once all components have been identified (flow measurements instruments Flow-Q & EduFlow-2, the MOOC GdP as research terrain, and the Logistic Regression Machine Learning model), we can mobilize them all in the Experiment task (*cf.* Figure 5-1 above).

The Experiments task is separated into two subprocesses, a Proof-of-Concept subprocess, and an ensuing Prototype subprocess.

The Proof-of-Concept subprocess validates (or refutes) the choices of flow measurement instruments (Flow-Q and EduFlow-2), and of the Machine Learning method (LogisticRegression), relying heavily on the amount of input samples issued from the terrain (MOOC GdP). Once the Proof-of-Concept has shown all choices perform correctly in tandem supported by several metrics (*cf.* its own Subsection "Results and metrics"), the Prototype subprocess can then begin.

This time, the Prototype subprocess trains again a LogisticRegression Machine Learning model with even additional[137] data: the MOOC learners' digital traces aggregated data, in the form of constructed indicators. Therefore, the resulting Machine Learning trained model detects this time flow using solely the learners' digital traces, without any flow measurement instrument application.

Both the Proof-of-Concept and the Prototype subprocesses follow a typical Machine Learning workflow (*cf.* Figure 4-2 in Section "Machine Learning workflow" above) comprising phases such as a Model requirements stage, a Data processing stage, a Feature engineering stage, a Model training stage, etc., some of which were covered just before the current Chapter (*cf.* "Method" in Chapter 5 above).

We illustrate (further below) in Figure 7-16 and Figure 7-27 the training (*a.k.a.* learning) phase, and in Figure 7-19 and Figure 7-39 the production (*a.k.a.* deployment) phase, for both Experiment 1 and 2 respectively.

During the training phase, the Logistic Regression Machine Learning model learns to recognize flow via many examples of what might characterize flow and no-flow, *i.e.*, what flow "looks like" and what no-flow "looks like".

[137] The existing and validated data of the flow measurement instruments during the Proof-of-Concept subprocess is employed again.

During the production phase, the trained Logistic Regression Machine Learning model is used to detect flow on unseen participants' data (Figure 7-19 and Figure 7-39, for Experiment 1 and 2, respectively).

It is noteworthy to mention that both subprocesses also share input data, which renders some of the stages' contents (*e.g.*, requirements, data processing, etc.) also common to both.

Indicators construction

We construct indicators according to the theoretical flow model identified. According to the EduFlow-2 theoretical model (*cf.* Section "Flow in Educational Contexts" above), four possible dimensions could arise in educational contexts, for which, in consultancy with the authors (J. Heutte, personal communication, December 2022), we construct the following aggregated indicators susceptible to be found (being this one of our main limitations) in the MOOC's learners' digital traces:

(i) Cognitive control (D1). We likened D1 to its previously acknowledged component "feeling of control", and expand this consideration over the MOOC platform itself, *i.e.*, we posit a "richer", more varied usage of the tools provided by the platform would point to a feeling of control.

Furthermore, we considered employing the cognitive state of the learner, usually found within the Learner Model (*cf.* footnote[37] above) (Bodily et al., 2018; Bull & Kay, 2010; Kay & Kummerfeld, 2019) and actively tracked by the MOOC, but digital traces available to us did not account for the learners' knowledge, nor for a standardized pedagogical path.
Still, we attempted to create a standardized path indicator carrying the information of the absolute position of the learner within the MOOC's proposed pedagogical path. This hypothetical position would depend on the context of the type of edX event considered. However, such effort proved futile as during the tests of the training phase (*cf.* Chapter 7(ii) below), the algorithm could not make sense of it.

Therefore, in this study we attempt to account D1 by the diversity of actions showcased in the logs as considered both by the edX platform itself and by our own categorization (*cf.* div_events and div_edx_cat in Table 7-19 and Subsection "Logs data filtering, cleaning, and aggregating" below).

(ii) Immersion & Time Transformation (D2). We likened the first part of D2 to the immersion experienced when actively interacting in the MOOC forums. We noted the total number of log events belonging to our category "forum participation" (*vs.* the more passive "forum reading" category).

The second part of D2 is entirely time-connected: we likened it to the time spent in the MOOC, both by the length of the sessions as well as the total number of sessions.

Therefore, in this study, we attempt to account D2 by the number of forum participation events, the number of valid sessions created, and the minimum, maximum, and average session length time in seconds.

(iii) Loss of self-consciousness (D3). After consulting with the authors of the EduFlow-2 measurement instrument, we convened to consider D3 to the experience of actively staying connected longer to the MOOC during what we might consider an atypical period for pedagogical instruction: over midnight on a Friday or a Saturday. This meant counting the number of longer-than-average sessions that started on a Friday and spilled over to a Saturday, and those that started on a Saturday and spilled over a Sunday

However, limiting ourselves to looking for flow in these two very specific days of the week seemed shortsighted and probably reflected our own biased learning experience. Then, we expanded this indicator to account for any session spilling over the following day, regardless of its total duration, because time spent in the MOOC was already accounted for by D2.

Thus, in this study, we attempt to account D3 by the number of sessions in which their last event occurred on a different date of their first event, *i.e.*, the session which started one day and ended another day.

(iv) Autotelic experience (D4). We initially considered a self-rewarding experience would prompt for consecutive, fast-paced, sequence of events and thus we proposed an AverageActionsPerMinute indicator, focusing only on non-passive actions, *e.g.*, reading, video-watching, etc.

However, this proved to be non-descriptive indicator because the final aggregation employed during the model training is calculated per user, in which case our proposed indicator becomes meaningless outside the scope of an entire session (or even outside the scope of a shorter period), *i.e.*, any given user has many sessions, with potentially different flow states, which cannot be reduced to a single numerical value.

But more importantly, flow (the "autotelic experience", *cf.* Subsection "The Challenge-Skill Balance" above) depends on the individual perception of the difficulty of the activity faced, and therefore should not instill a generalized "pace" of actions, even within the context of the same individual, because different activities present different challenges, thus different paces, which could all potentially lead to flow. If anything, this indicator could be meaningful if the activity context is known, but more specifically, if the timespan of the activity is known, as to distinguish flow pacing and no-flow pacing (an individual pacing indicative of flow or of no-flow). This is so far difficult to identify in the digital traces available to us.

Thus, in this study, we are unable to account D4 by any reasonable proxy found in the digital traces.

This concludes the present Section, which overviews the planning of steps to be taken to construct the experimentation. We also presented the reasoning behind the construction of indicators in accordance with the theoretical flow model identified and the available digital traces. A summary of the indicators transformed into fields can be found in Appendix 14 on page 397.

Data collection

This Section describes both Experiment 1 and Experiment 2 data collection processes. For the questionnaires' data collection, common to both experiments, we follow the Flow measurement protocol we designed and presented above. For Experiment 2, besides the questionnaires' data already processed during Experiment 1, we employ the MOOC's learners' digital traces, which we explain how they are collected from the MOOC hosting platform.

Questionnaire application

For the application of our selected Flow measurement instruments (FlowQ and EduFlow-2) and being limited to the application periods shown in Table 7-6, we chose to maximize the time to gather data over equal-lengths, smaller time data collection periods. We gathered the respondents' data at each period's closing date (shown in Table 7-6, column "Closure & Collection").

The chosen flow measure instruments were asked during P2 and P3; with FlowQ first (three items) and then EduFlow-2 (12 items). The specific scales applied are found in Appendix 2. – The FlowQuestionnaire and in Appendix 8. – The EduFlow & EduFlow-2 measure instruments, respectively. Demographics were in turn gathered during P1.

Table 7-6

Measurement Instruments' Application Periods and Data Collected

Period	Opening	Closure & Collection	Data collected
P1	Start of Week 0	End of Week 4	Demographics
P2	Start of Week 3	End of Week 4	Flow-Q, then EduFlow-2
P3	Start of Week 4	End of Week 11	Flow-Q, then EduFlow-2

Note: This scheduling was employed for all sessions.

It is relevant to remark that other psychometric instruments beyond the scope of this research work (chosen and managed by the Research & Development team from the MOOC

GdP) are applied jointly as well during these same applications periods, *i.e.,* the total cognitive burden for the survey's participants is not determined solely by the measurement instruments presented here and usually, two or three more surveys are performed jointly.

All questionnaires (comprising our and other psychometric instruments) share the following field fields (last column name changed nomenclature likely due to a software update), shown in Table 7-7. Please notice that:

- The field submitdate is a timestamp for when the entirety of the questionnaire was finalized, *i.e.*, partial answers (questionnaires abandoned mid-way) are saved by the surveying software, which might include the entirety of our chose flow measurement instruments or not. Individual item verification for empty values is required to discern if the measurement instrument item was completed or not.
- The id, token, and email fields work as key fields, allowing us to link learners from both the MOOC and the surveying platforms (Open edX & LimeSurvey), and from subsequent questionnaires.

Table 7-7

Questionnaires' Common Fields

Field code	Description
id	LimeSurvey's sequential id number
submitdate	LimeSurvey's timestamp when the entirety of the questionnaire was finalized
token	Internal Student ID from the Open edX platform
email	Student's email in the Open edX platform
lastname/surname	Student's username in the Open edX platform

Note: These fields were common to all sessions.

Also, demographic fields are common to all questionnaires (field QR1Sexe was later edited to admit non-binary values with an additional text field for 'Other'), show in Table 7-8. Scrollable lists (List) allow for closed answer selection and yield an integer equivalent to its value in the list. For the sake of readability, available values per list are not mentioned here but detailed in Appendix 10.

Table 7-8

Questionnaires' Common Demographic Fields

Field code	Possible values	Item
QR1Sexe	1=Man, 2=Woman; string=user entry, for Other	Êtes-vous?
QR1AnneeNaissance	List: 1905-2010 years	Quelle est votre année de naissance?
QR1PaysRegionResiden	List: 191 countries	Quel est votre pays de résidence ? (et non d'origine). Tapez la première lettre pour aller directement au nom du pays.
QR1Statut	List: 10 statuses	Quel est votre statut ?
QR1NiveauScolaire	List: 7 levels	Quel est le plus haut niveau scolaire que vous ayez terminé (formation générale ou professionnelle) ?
QR1FIGdP	List: 3 answers	Avez-vous reçu une formation à la gestion de projet au cours de votre cursus académique ?
QR1FCGdP	List: 4 answers	Si vous avez terminé vos études, suivez-vous dans le cadre de votre travail :

Note: These fields were employed for all sessions.

In the context of thesis, for each tracked session (2020a, 2020b, 2021a, 2021b) of the MOOC GdP (GdP15, GdP16, GdP17, and GdP18), and for each application period (P1, P2, and P3), we coded the field names of the applied measurement instruments' items as shown in Table 7-9:

Table 7-9

Field Name Coding per Session

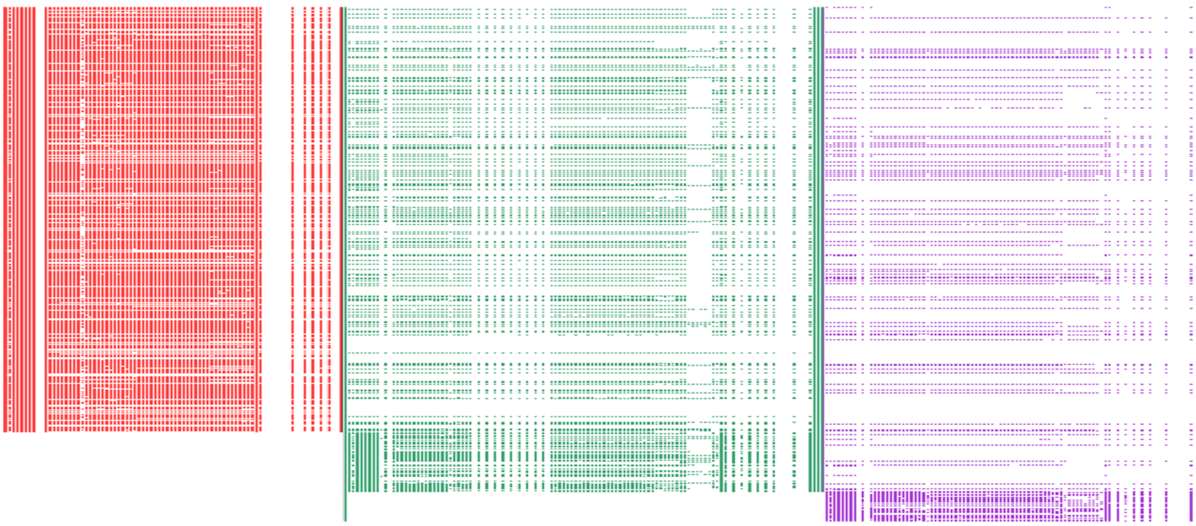
GdP (Session)	P1 codes	P2 codes	P3 codes
GdP15 (2020a)	Common demographic fields	FlowMOOC, QR2Flow	QR3FlowMOOC, QR3Flow
GdP16 (2020b)		QR2FlowMOOC, QR2EduFlow	QR3FlowMOOC, QR3Flow
GdP17 (2021a)		QR2FlowMOOCac, QR2EduFlow	QR2FlowMOOCac, QR3Flow
GdP18 (2021b)		QR2FlowMOOCac, QR2EduFlow	QR2FlowMOOCac, QR3Flow

Questionnaires items' names and variables are inconsistent during the data collecting phase, but they were subsequently all homogenized and corrected during the Data filtering and cleaning step.

We show in Figure 7-3 the silhouette of a typical MOOC session raw data. It shows the participant answers to all the psychometric instruments applied by the Research & Development team from the MOOC "*Gestion de Projet*". Please, notice the dilution of MOOC participants as time passes on (from P1 to P2 to P3, from left to right): fewer participants answer the measurement instruments during P3 than during P1.

Figure 7-3

Graphical Representation of One Typical MOOC Session (Ramírez Luelmo, El Mawas, Bachelet, et al., 2022, fig. 5)



Note: Coloring differentiate application periods P1, P2 and P3 (from left to right: red, green, and violet), blanks represent null data, and non-blanks represent data, each horizontal line means to represent a participant, with columns correspond to the participant's answers. Furthermore, the bottom "jagged" data rows indicate participants who started answering the measurement instruments late, *e.g.*, not answering during P1 but participating during P2 and P3.

One check question was placed in the middle of the EduFlow-2 measure instrument (but not in the Flow-Q questionnaire, being too short) to verify if participants read all the items, followed directives (*e.g.*, "Please, select 4 for this item"), and were not simply answering randomly, a phenomenon already described in the literature (Obadă, 2021). Actual questions and item codes are shown in Table 7-10. Please note that internally and only visible to us, the item code includes the correct answer (*e.g.*, expected value for item named "DET2" is "2", and for "DET4" is "4"), to facilitate automatic batch processing verification during the data clean-up phase.

Table 7-10

Check-Questions Field Coding

Item code	Question	Approximate translation
DET2	Répondez "2 - Très peu d'accord" à cette phrase	Answer "2 - Disagree" to this sentence.
DET4	Répondez "4 - Moyennement d'accord" à cette phrase	Answer "4 - Agree" to this sentence.

Full dictionaries of variables for all questionnaires, for all items, for all sessions were kept but are not shown here, for the sake of brevity and readability.

Initially we gathered a total of 9 448 participants' answers, which after a rigorous cleaning-up data process (*cf.* Experiment 1 – Proof-of-Concept), yielded an input data sample constituted of 1 589 trustworthy participants' questionnaires self-reported answers (n = 1 589). For publication purposes of this research work, all data was anonymized by removing any personal data and/or attributes (Ferreira Marques & Bernardino, 2020).

Our gathered data spans four MOOC Sessions, ranging from March 2020 – December 2021, *i.e.*, two years of data collection. We merged all four sessions, three period's (P_n) data, into a single CSV file using the common lastname/surname field as key, and calculated scores for both flow measure instrument while also keeping the self-reported individual items' results of the EduFlow-2 measure instrument.

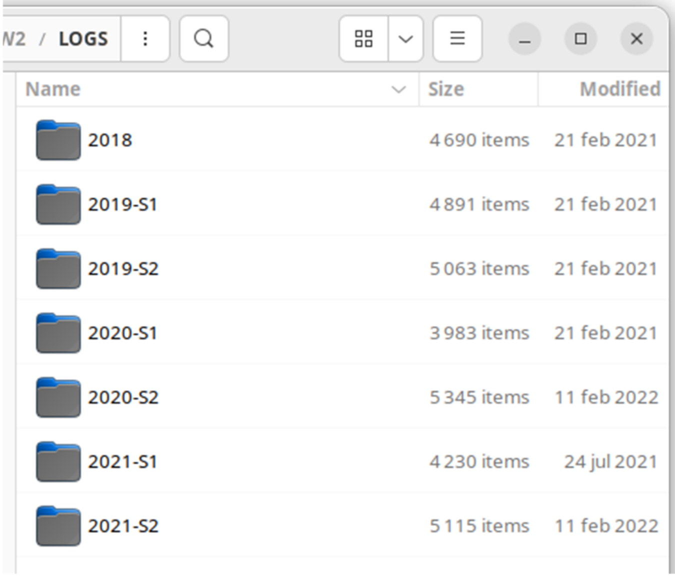
Traces: log data

Log data collection occurred as a batch (*cf.* Chapter 4(iii) above), according to the MOOCit's policies (*cf.* Subsection "As a research platform" above). The MOOC's log events were fully recorded and exported as massive log files, stored in an S3 bucket. We were granted access to the S3 bucket to download available archived log files in two moments:

1. Logs accounting from 2018 up to 2020: 23 976 files and folders, totaling 182.9 GiB (thousands of uncompressed files.
2. Logs accounting from 2018 up to 2021: 38 459 files and folders totaling 34.7 GiB, originally in a single tarball (*.TAR[138]) compressed file.

Figure 7-4

Second Moment's Data Download Partial Screenshot



138 https://www.libarchive.org/

Second data access happened to also include first moment's data (in folders named 'diff') making it the most complete data source of both, and therefore rendering the first moment data access redundant. This two-moment, incomplete log download necessarily delayed all possible log data treatment. Decompressed data from the second moment access is illustrated in Figure 7-4 above.

Additionally, we discarded data which we could not associate to the applied flow measurement tools, *i.e.*, the 2018 folder, accounting for log data stored before this study took place.

Thus, compressed second-moment data can be resumed as follows (difference with previous total files and size is due to redundant and additional 2018 data):

Table 7-11

Second Moment's Data Delivery

Session	Contents & size
2021a	4 229 compressed files, 1.1 GiB
2021b	5 115 compressed files, 2.9 GiB
2020a	3 983 compressed files, 2.0 GiB
2020b	5 344 compressed files, 4.2 GiB
2019a	4 891 compressed files, 1.5 GiB
2019b	5 063 compressed files, 3.6 GiB

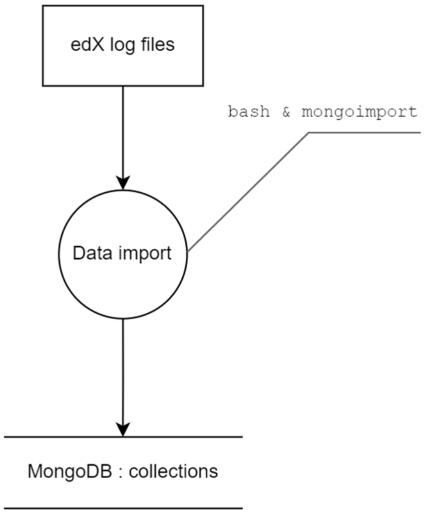
Files were compressed in the gzip[139] file format (*.GZ) at a heavy ratio (~12:1). Log files were originally in JSON text format, as documents. To limit file conversion issues and data structure loss, we employed the document-oriented database software MongoDB[140], supporting JSON-like storage.

Please notice that a file and a document are not the same, especially in the context of MongoDB's vocabulary[141]. In general, a JSON file contains many documents.

139 https://www.gnu.org/software/gzip/
140 https://www.mongodb.com/try/download/community
141 https://www.mongodb.com/docs/manual/core/document/

Figure 7-5

DFD for Importing edX Log Files into MongoDB



Decompressing log files and importing them into MongoDB was done in a batch recursively using the mongoimport command. The data import process is schematized in the Data-flow Diagram (DFD[142]) (T. H. Tse, 1986; T. H. Tse & Pong, 1989) shown in Figure 7-5: multiple edX log files are inserted into multiple MongoDB collections.

A command-line example for a local import of the 2020b collection is in turn illustrated in Figure 7-6. It goes through every file in the local system's tree structure and attempts to insert the documents within into the specified MongoDB Collection:

Figure 7-6

Example of a Bash Command to Import Logs Files

```
for f in ./*; do mongoimport -vvv --host localhost:27017 --db mooc_gdp --collection 2020
b --file $f; done
```

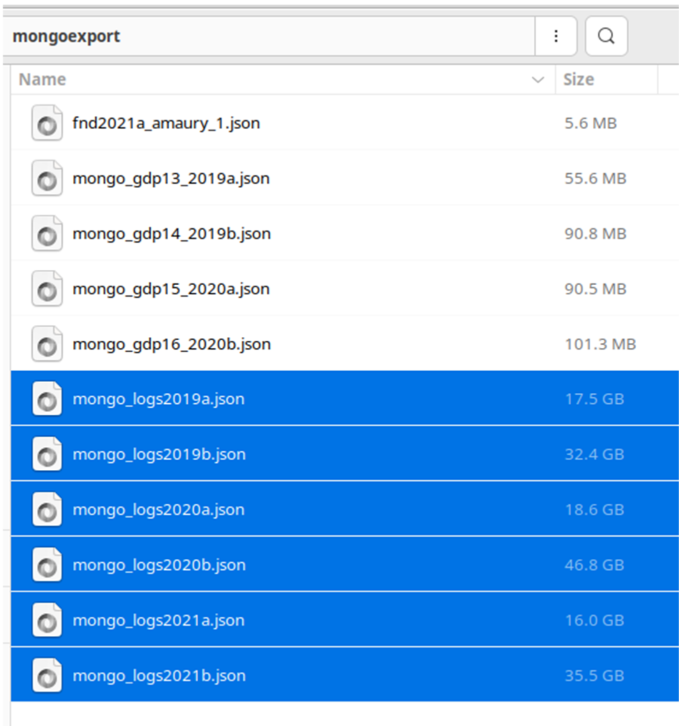
[142] DFD notation preferred over flowcharts or UML activity diagrams because of their simplicity.

Note: The previous example recursively decompresses and imports the log files into a local MongoDB database into a 2020b Collection.

Once the files were imported into MongoDB they were re-exported as individual, uncompressed JSON files for backup purposes, highlighted in blue in Figure 7-7.

Figure 7-7

Backed-Up Selected Log Collections Partial Screenshot



Note: Backed-up log collections were exported as JSON files (in blue).

Precise data consumption for each collection is reported by MongoDB as shown in Table 7-12, where the number of documents is also shown (number spacing was manually added for readability purposes).

Table 7-12

Individual Log Collections' Data Consumption (as Reported by MongoDB)

Collection	Size in bytes	Documents count

logs2019a	17 104 594 011	13 493 833
logs2019b	31 655 054 380	22 828 314
logs2020a	18 157 058 807	13 878 530
logs2020b	45 830 199 548	35 227 286
logs2021a	15 610 334 022	12 082 251
logs2021b	34 579 556 781	24 469 633

In accordance with the Flow measurement protocol previously designed, full file backups were effectuated after each of the previous steps (except for the last step, which is a backup step itself):

1. Downloading logs from the AWS server.
2. Unarchiving from a single *.TAR file, where it applies.
3. Decompressing and importing into MongoDB, via the mongoimport command (from many *.GZ files spread over a tree structure).
4. Extracting and exporting individual JSON files, via the mongoexport command.

Finally, subsequent log cleaning and filtering (not covered in this Subsection but in Subsection "Data filtering and cleaning" below) led to a single MongoDB collection accurately named 'logs', accounting for the previously shown six collections, which data consumption follows in Table 7-13:

Table 7-13

Entire Logs Collection Data Consumption (as Reported by MongoDB)

Collection	Size in bytes	Documents count	Average Object Size
logs	86 381 368 614	47 266 376	13 493 833

This concludes the present Section, concerned with the processes applied for and during data collection for both the questionnaire data and the log data. Data collection mostly followed the Flow measurement protocol formerly designed, spanning a two-year period. Because of this lengthy data collection period (and its accompanying data cleaning) we could only employ Offline Learning (*cf.* Subsection "Main Definitions") for both Experiment 1 and Experiment 2.

Questionnaire data is employed in both Experiment 1 just below, and Experiment 2, while log data (the digital traces) are employed only during Experiment 2 – Prototype.

Experiment 1 – Proof-of-Concept

In this Section we describe Experiment 1 – Proof-of-Concept. We detail the data filtering and cleaning effectuated before feeding the cleaned-up data as training data for the Logistic Regression Machine Learning model. Resulting metrics show the pertinence of both flow measurement tools when working in tandem. Extensive parts of Experiment 1 appear in several publications by this thesis' author (Ramírez Luelmo, 2022; Ramírez Luelmo, El Mawas, Bachelet, et al., 2022; Ramírez Luelmo, El Mawas, & Heutte, 2022), with one being awarded "Best Doctoral Paper" (Ramírez Luelmo, El Mawas, Bachelet, et al., 2022) and received an invitation for extending it into a book chapter (submitted and decision pending).

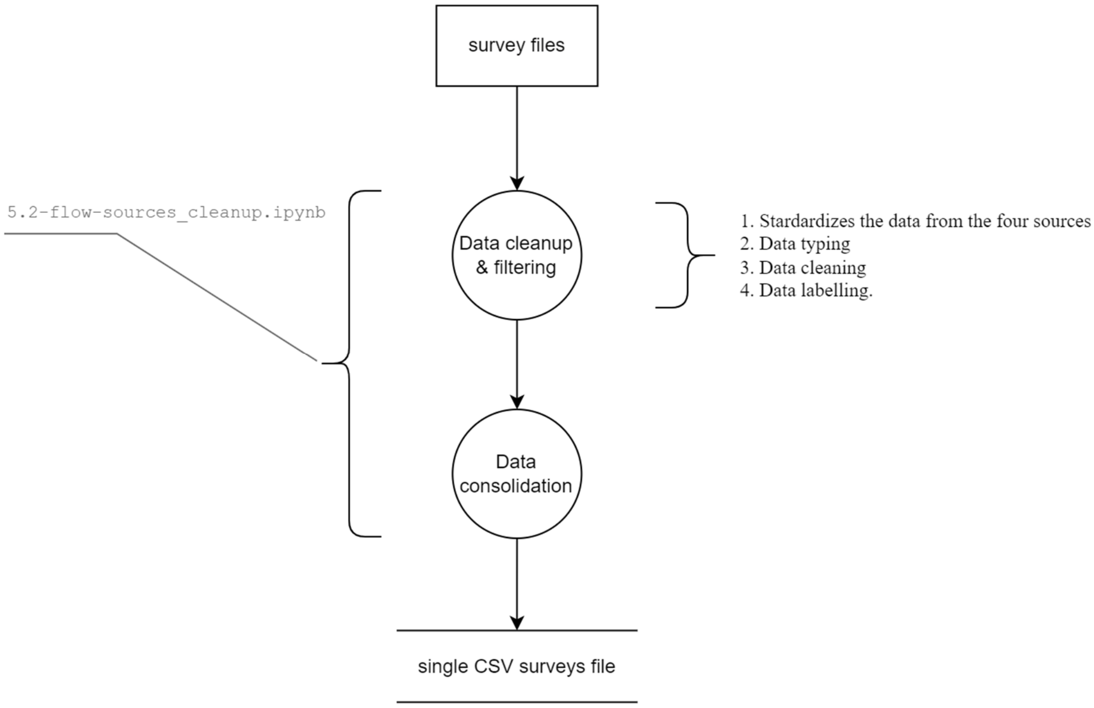
Data filtering and cleaning

We start with a broad picture of the data flow processes performed, as illustrated the DFD shown in Figure 7-8: all surveys' data (CSV files) are pre-processed (data typing and labelling), cleaned up, and consolidated into a single CSV file. This is performed in a Jupyter's 5.2-flow-sources_cleanup.ipynb notebook. Consolidated resulting file was employed as source dataset for publications (Ramírez Luelmo, 2022; Ramírez Luelmo, El Mawas, Bachelet, et al., 2022).

Then, we proceed with detailing the filtering and cleanup process. We completely discarded data rows of respondents who answered incorrectly to any of our two check items (*cf.* Subsection "Questionnaire application" above), if they chose multiple genders at once, or if they setup an unrealistic birth year, *e.g.*, selecting "man" and "woman" and "other" at once, or "1916", etc.

Figure 7-8

Experiment 1: DFD for the Data Cleaning Process



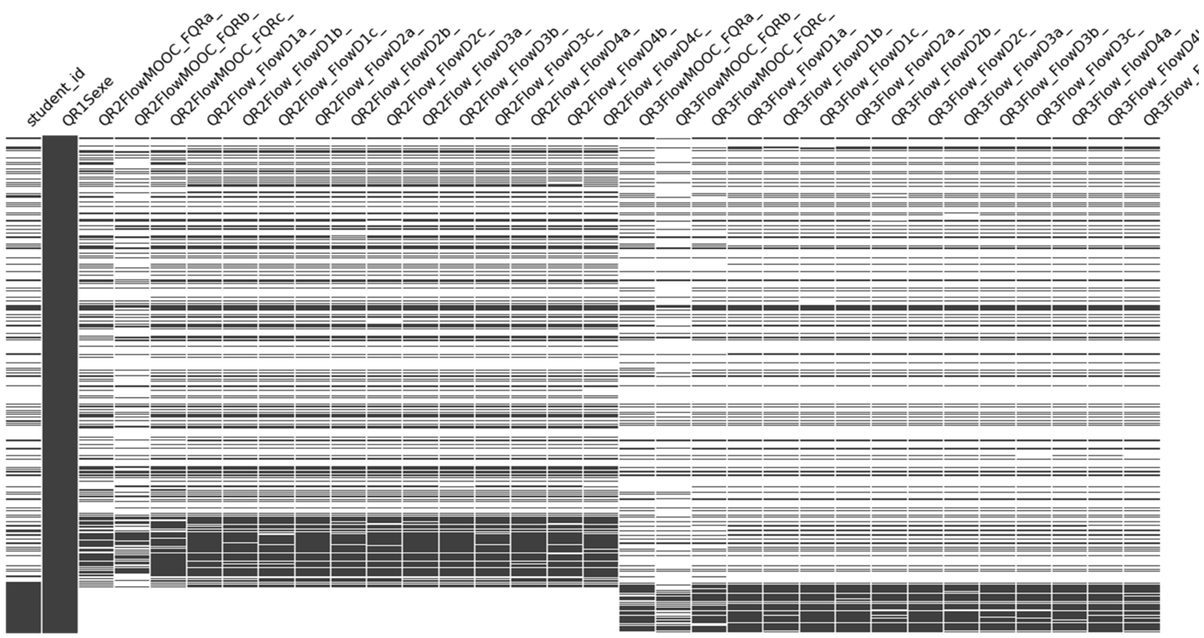
Also, we individually verified all flow measurement items for empty values to discard incomplete answers, *i.e.*, a learner answering partially a questionnaire (the entirety of psychometric instruments applied by the Research & Development team of the MOOC GdP) did not convey immediate elimination if our flow measurement instruments were fully answered, *e.g.*, we keep the data of a learner fully answering Flow-Q and EduFlow-2 even if abandoning the remainder of the questionnaire.

Items names were all homogenized and corrected during data labelling. First, Flow-Q items (coded FlowMOOC, and FlowMOOCac) were prefixed instead correctly prefixed with string “FlowQ”. Then, EduFlow-2 items (coded EduFlow, Flow) were instead correctly prefixed “EduFlow2”. Individual items were homogenized as Ia, Ib, and Ic, and D1a, D2a, D3a, D4a, D1b, D2b, D3b, D4b, D1c, D2c, D3c, D4c, respectively for each measurement instrument, *e.g.*, field name “QR2EduFlow2.D4c” corresponds to item D4c (fourth flow dimension, third item) of the EduFlow-2 measurement instrument applied during the second period P2.

Figure 7-9 depicts a graphical representation of a given raw data sample (here, GdP17) before clean-up and scoring. It clearly shows the missing data (white spaces) *vs.* the available data (dark areas), and only a few selected fields are shown.

Figure 7-9

Graphical Representation of a Given Raw Data Sample



Note: This representation reflects a given raw data sample before cleaning it up.

Preliminary analysis of partially collected data already pointed out to trends we would see in the totality of the cleaned-up dataset. For instance, GdP17's collected data showcased better quality data (higher proportion of participants with reliable answers to the survey) during QR3 than during QR2, and a clear target data imbalance. Both phenomena can be appreciated in Figure 7-10 and Figure 7-11, for QR2 and QR3 respectively (GdP17).

In both Figures, a simple Linear Regression Machine Learning model was employed to determine how well the GdP17 dataset could be divided into two distinct classes, represented by the two vertical lines of blue points for no-flow (left) and flow (right). An orange diagonal represents the linear trend the model should follow approximately: no-flow blue dots should concentrate at the origin (bottom-left) of the orange diagonal (0, 0) while flow blue dots should instead agglomerate at the end (1, 1) (top-right).

Figure 7-10

Linearity Assumption During GdP17 – QR2

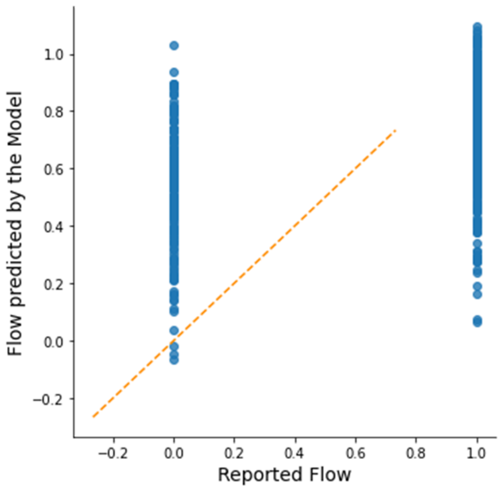
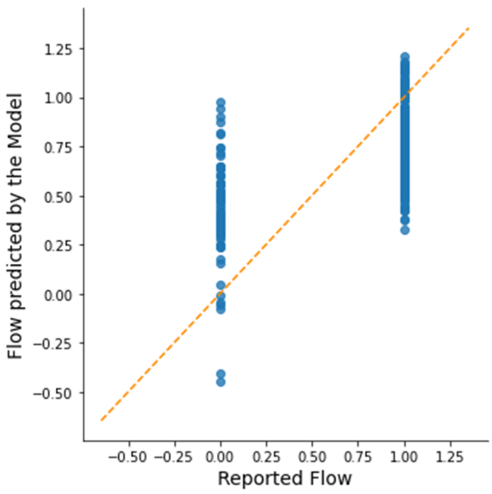


Figure 7-11

Linearity Assumption During GdP17 – QR3



In one hand, during QR2, Figure 7-10 shows no-flow blue dots not concentrating at the origin of the orange diagonal but instead spreading over a broader range (~ 0.2 – 0.8) and situated closer to the flow dots concentration (~ 0.4 – 1). In the other hand, during QR3, Figure

7-11 shows that no-flow dots are less spread (~ 0.3 – 0.55) when compared to QR2, while favorably flow dots also agglomerate closer to 1.

After filtering, cleaning up and consolidating the data, out of an initial pool of 9 448 participants to the Research Questionnaires, we accounted 1 589 trustworthy participants' questionnaires self-reported answers (n = 1 589).

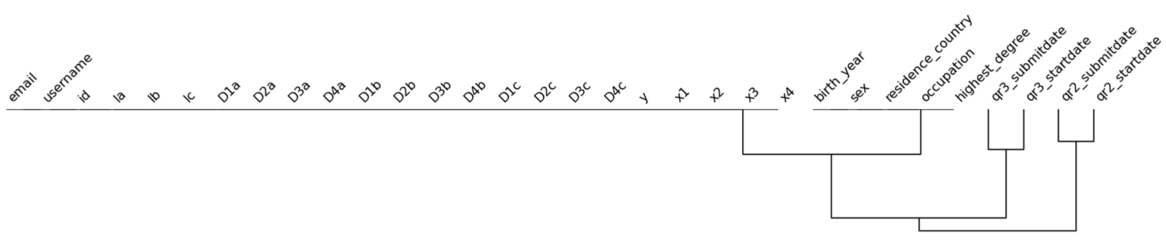
A simple dendrogram representation of this cleaned-up survey dataset is depicted in Figure 7-12. It clearly shows automatic class distinction of input data into six classes, which we group into three main classes featuring (from left to right):

- flow data (platform identifiers, raw surveys' answers, and partial/full flow scores),
- demographic data (gender, birth year, country of residence, occupation and highest studies degree achieved), and
- timestamp data (survey start/end).

These favorable preliminary results hint to clear distinctions and relationships on the nature of features of our input data, *e.g.*, demographic feature residence_country grouped alongside the flow class would point to an issue with the input dataset.

Figure 7-12

Simple Dendrogram of the Cleaned-Up Survey Data

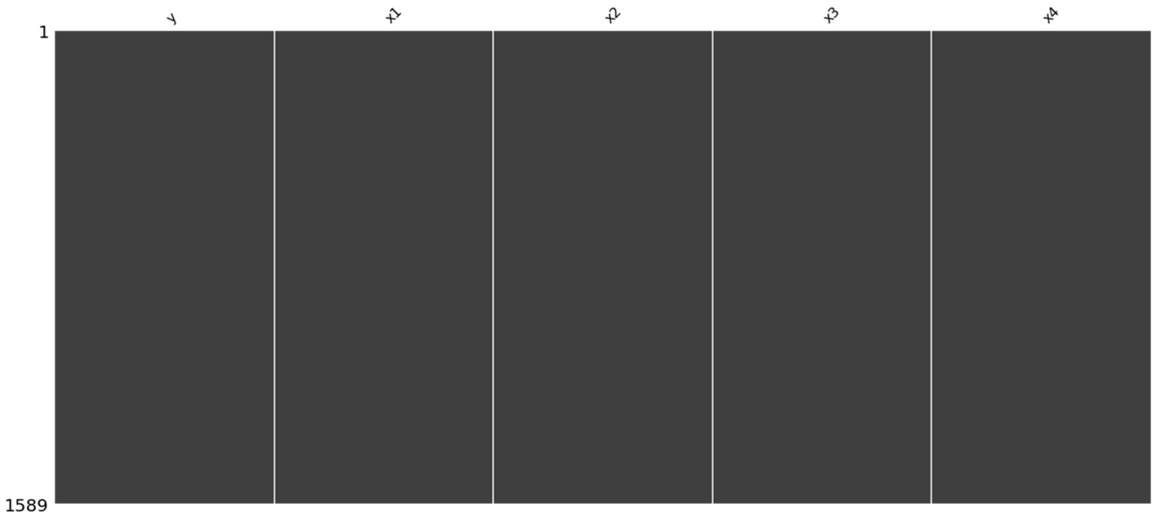


For Experiment 1, we employed the simple, binary scoring (y, representing flow presence (1) or flow absence, *a.k.a.* no-flow (0), a design choice of the instrument's authors) of the Flow-Q measurement instrument, and the EduFlow-2 measurement instrument's four-dimensional partial scores (x_1, x_2, x_3, x_4, representing the degree of agreement to a pedagogical-related activity situation).

Figure 7-13 shows the graphical representation of the data shape of the cleaned-up, labelled and consolidated data sample employed for Experiment 1: gray areas represent data, arranged in columns, with no visible horizontal white spaces, representing missing data. The first column indicates the known target (*y*), followed by the four feature columns.

Figure 7-13

Graphical Representation of the Cleaned-Up Sample



Note: This represents the all-respondents (n = 1589) CSV file, with the target (y), the four features (x_i) and no missing data.

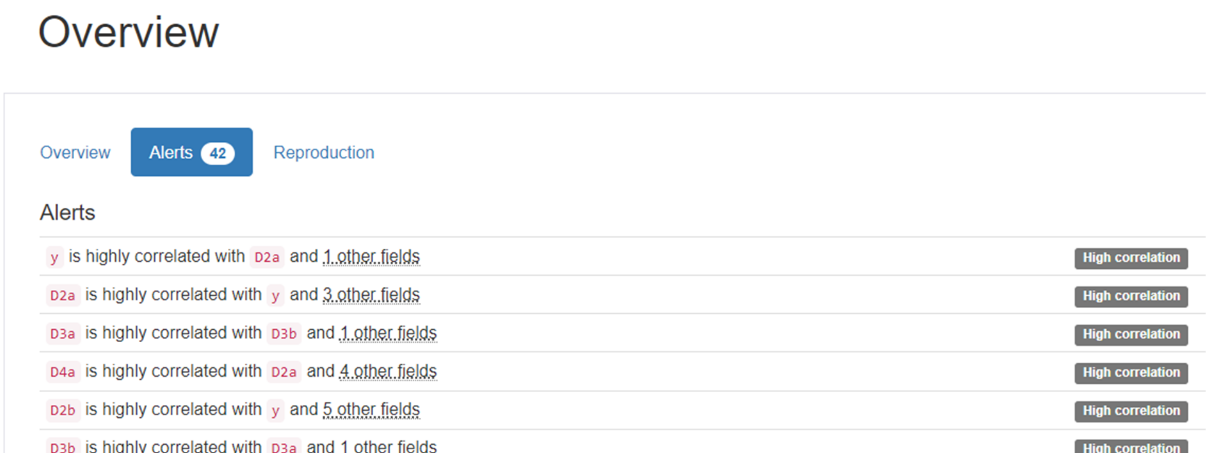
Furthermore, a download link to an interactive, stand-alone HTML Pandas Profiling Report[143] of the anonymized multi-dimensional input data sample is available at the end of Appendix 9, featuring Spearman's, Pearson's, and Kendall's correlations heatmaps. The report on our input dataset raised 42 alerts, of which the top, partial view can be seen in Figure 7-14. The

[143] https://pypi.org/project/pandas-profiling/

42 alerts correspond to various combinations of highly correlated features, *e.g.*, D2b is highly correlated with y and D2a, D4a, D4b, D2c, and D4c (unseen, expanded fifth line in Figure 7-14).

Figure 7-14

Pandas Profiling Report on the Input Data

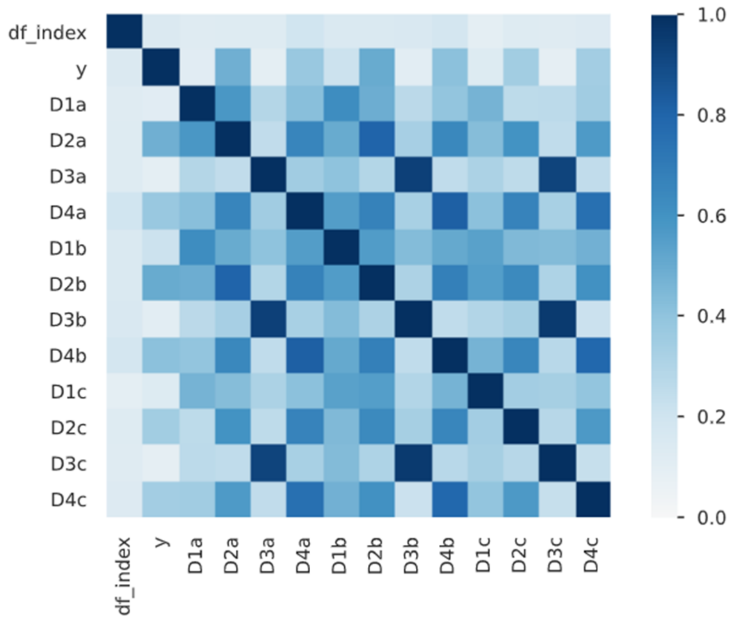


Note: The Pandas Profiling Report (*cf.* Appendix 9) includes more information that we do not include here for brevity reasons.

These alerts are far from representing a problem and instead they embody one more step in the proper direction. As previously seen during our literature review on flow measurement instruments, flow dimensions are expected to be highly correlated (*cf.* Section "Considerations When Measuring Flow" above), which is the desired case for our data. This behavior is clearly observed in Figure 7-15 where the ϕ_K correlation (Baak et al., 2019) heatmap for Flow-Q target y and D_{iq} EduFlow-2 dimensions ($i = \{1, 2, 3, 4\}$, $q = \{a, b, c\}$) shows a few $\phi_K > 0.4$ and predominant (besides same-variable correlations) $\phi_K > 0.6$ correlations (0 – 1). Additionally, Spearman's, Pearson's, and Kendall's correlations heatmaps are also depicted in Appendix 9, featuring the same behavior.

Figure 7-15

ϕK Correlation Heatmap for Flow-Q target y and EduFlow-2 Dimensions



In a more traditional approach, we would explain the correlations pairs, what they represent in our flow research context and conclude the experimentation. However, having highly correlated variables is a known fact (*cf.* Section "Considerations When Measuring Flow" above) that represents another contributing reason to employ Machine Learning to make sense of high dimensional data, as these statistical pairs correlations cannot further explain complex multi-dimensional relationships.

Model training

Instead, we employ a Logistic Regression classifier with a binary target (1 – flow presence, 0 – flow absence) to be determined by four independent variables (four EduFlow-2 dimensions d1, d2, d3, d4, arisen from the 12 surveyed items), represented in the dataset as x_1, x_2, x_3, x_4).

Specific to the Logistic Regression binary classifier in the scikit-learn python library[144], the probability of the positive class $P(y_i = 1|X_i)$ is predicted as:

[144] https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression &

$$\hat{p}(X_i) = \operatorname{expit}(X_i w + w_0) = \frac{1}{1 + \exp(-X_i w - w_0)}.$$

The classifier aims to minimize the following cost function:

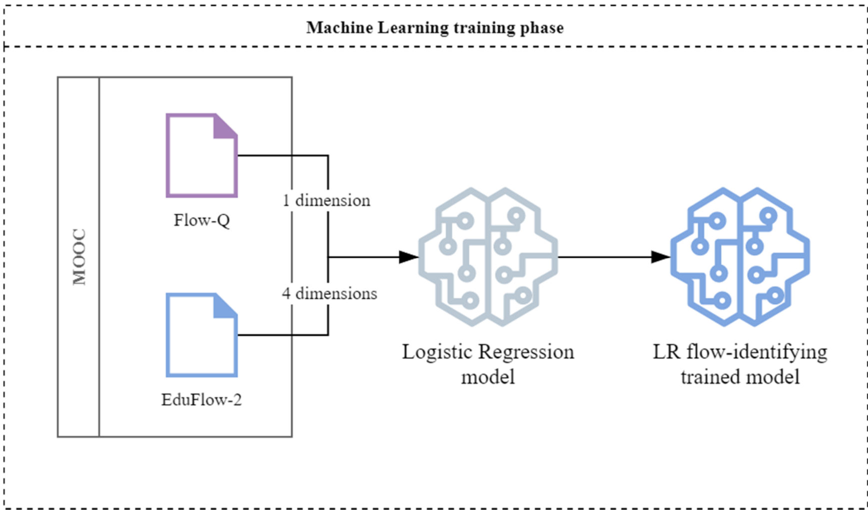
$$\min_{w} C \sum_{i=1}^{n} (-y_i \log(\hat{p}(X_i)) - (1 - y_i) \log(1 - \hat{p}(X_i))) + r(w).$$

where $r(w)$ is the regularization term (*cf.* footnote [144] above) of the classifier.

Experiment 1 training phase is illustrated in Figure 7-16, where the MOOC GdP (on the left) yielded a flow score and four flow dimensions, which were in turn fed to a Logistic Regression [Machine Learning] model, resulting in a Logistic Regression flow-detecting trained model.

Figure 7-16

Experiment 1: Training Phase



This resulting trained model was then evaluated to corroborate or refute the performance of the flow measurement instruments and the Logistic Regression model when working in tandem, and the pertinence of the dataset size (*cf.* Subsection "Results and metrics" below).

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

No production phase was planned for the resulting Proof-of-Concept model, but a hypothetical deployment scenario is shown in Figure 7-19, in Subsection "Discussion and conclusion".

All experiments employed the scikit-learn (Pedregosa et al., 2011) built-in Machine Learning Logistic Regression classifier for Python[145] (Van Rossum & Drake Jr, 1995). This specific experiment was run in a Google Colaboratory (Google Colab) remote instance (Bisong, 2019b) which was detected by CodeCarbon as a Python 3.7.13 on a Linux-5.4.188+-x86_64 running Ubuntu-18.04-bionic with 12.683 GB RAM, two Intel® Xeon® CPU @ 2.20GHz and no GPUs, in the United States (USA, District of Columbia or Nevada).

Available data was randomly divided into training and testing sets at a 70/30 ratio. The test set was employed only when the model finalized training to evaluate its performance on previously unseen data. No hyperparameter optimization took place so no data was reserved for that purpose.

We manually designed a pipeline chaining the PolynomialFeatures and StandardScaler pre-processors to the LogisticRegression classifier. The PolynomialFeatures pre-processor was given a degree argument of 2 (default). The LogisticRegression solver was left to the "lbfgs" default, and its multi_class parameter was set to "multinomial". Permanent caching of the pipeline was effectuated in an external Joblib[146] file.

Multiple instances of the Machine Learning Logistic Regression model were trained with the Flow-Q (y) and EduFlow-2 scores (x_1, x_2, x_3, x_4) from the training set (70%). Additionally, a 10-fold Cross Validation took place for each random instance training to assess its performance

[145] Python has shown to be one of the two most employed languages in data science (Stack Overflow, 2020), constantly raising in usage from 2015 to 2020, in 2019 being the "fastest-growing major language" (*Stack Overflow Developer Survey 2019*, 2019), all-in-all positiong itself among the top three most commonly-used programming languages (*Stack Overflow Developer Survey 2021*, 2021; *Stack Overflow Developer Survey 2022*, 2022; *Stack Overflow Developer Survey 2023*, 2023).

[146] https://joblib.readthedocs.io/

across distinct data slices and determine the best model fit using the means of Precision, Jaccard, and F1 metrics.

Among those trained Machine Learning Logistic Regression models instances, this research study presents the results of the one where the Accuracy, Precision, and ROC AUC scores on the test set (30%) are all simultaneously higher (value differences between other instances < 5%). This resulting pipeline is depicted in Appendix 12, where the weights employed for each step are appreciated. It is also digitally available at the locations determined by the CIREL-Trigone laboratory.

Results and metrics

All the results and metrics shown here were performed on the reserved test set (30%), which the finished model was not confronted with at all during training.

During the evaluation 10-fold Cross Validation, means of metrics relevant to regression and classification (Accuracy, Precision, Jaccard, and F1) were calculated. These metrics are shown in Table 7-14.

Table 7-14

Means of Metrics Applied During the 10-Fold Cross-Validation

Test	Mean	Standard deviation
Accuracy	0.78	0.02
Precision	0.80	0.01
Jaccard	0.72	0.02
F1	0.83	0.01

The classification report of the resulting Machine Learning flow-detecting model is shown in Table 7-15 below. Scores for flow presence detection are clearly higher than for the flow absence detection (second and first rows respectively of the Classification Report).

Table 7-15

Classification Report for the Resulting Machine Learning Flow-Detecting Model

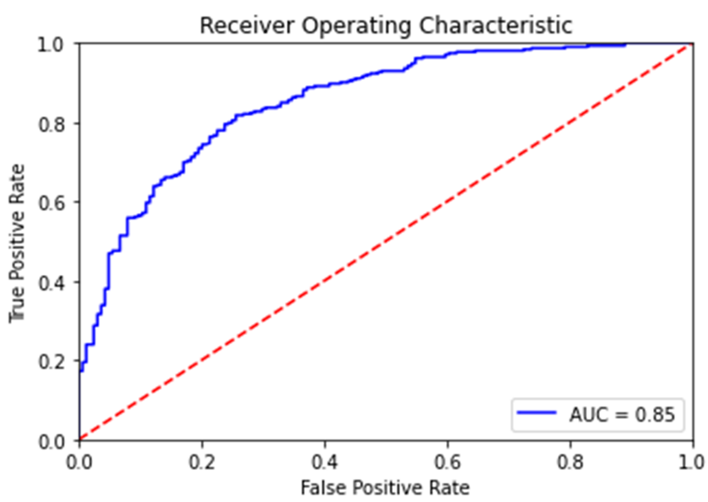
	Precision	Recall	F1-Score	Support
Flow absence	0.74	0.63	0.68	164
Flow presence	0.82	0.88	0.85	313
Accuracy	-	-	0.8	477
Macro avg.	0.78	0.76	0.77	477
Weighted avg.	0.79	0.80	0.79	477

Note: All scores were rounded up at the source.

The ROC curve of the resulting ML flow-detecting model features an AUC of 0.85 (blue curve line), shown in Figure 7-17 below.

Figure 7-17

ROC Curveof the Flow-Detecting Machine Learning Model

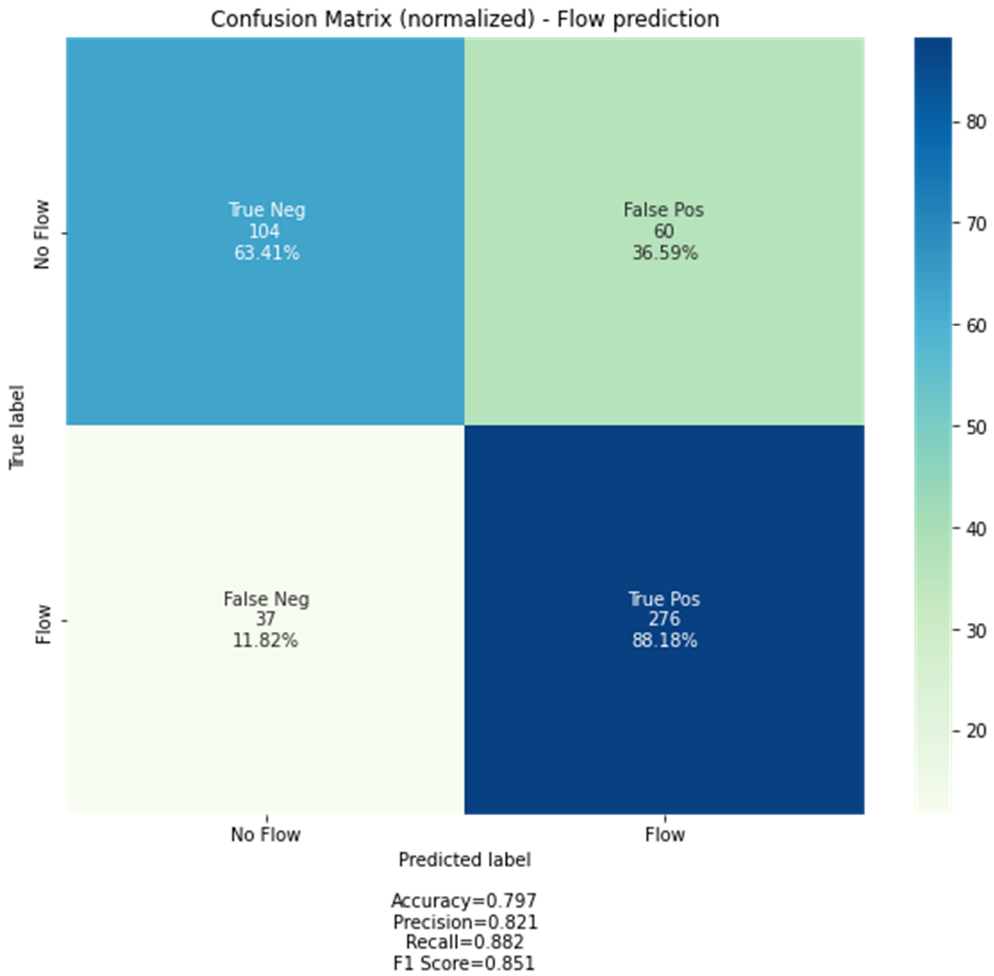


Note: The ROC is represented by the continuous blue curve (AUC = 0.85) *vs.* a hypothetical random classifier curve (dotted straight red line).

The normalized Confusion Matrix (Figure 7-18) of the resulting flow-detecting Machine Learning model shows a combined Accuracy of 0.797, with a larger proportion (88%) of correctly predicted cases in the True Positives cell, compared to the True Negatives cell (63%).

Figure 7-18

Normalized Confusion Matrix for the Flow-Detecting Machine Learning Model



Discussion and conclusion

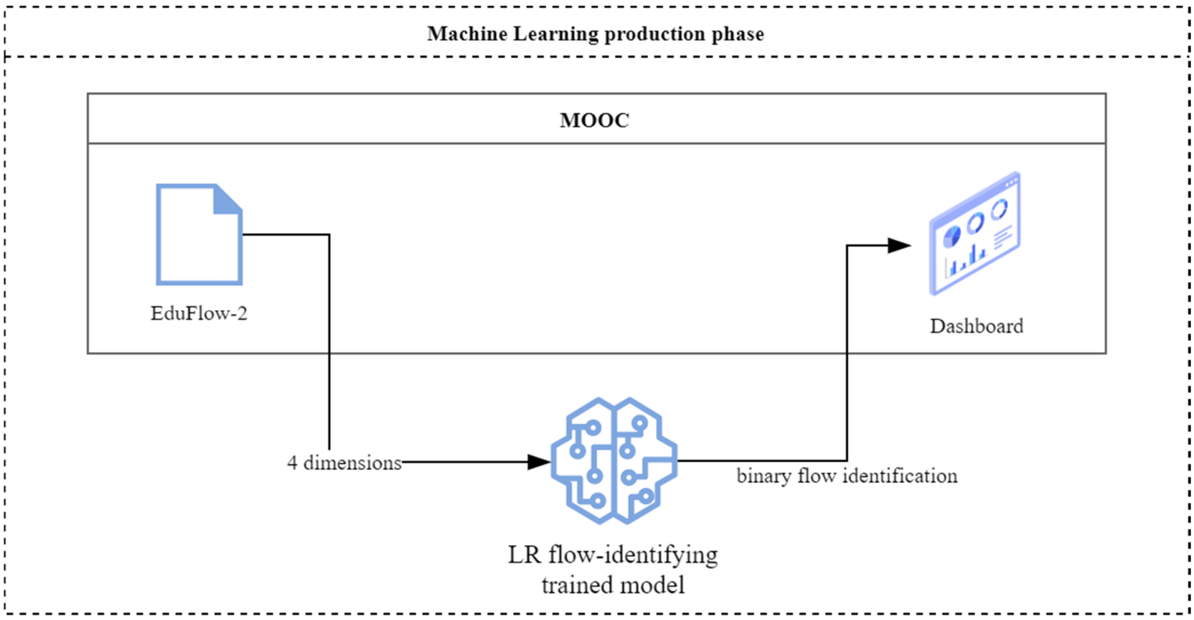
Metrics' results above the predefined thresholds (*cf.* Chapter 5(ii) above) displayed in this experimental Proof-of-Concept showed the correctness of both chosen flow measuring instruments and the selected Machine Learning method. The association of both the results of Flow-Q and the dimensional analysis of the EduFlow-2 model via a LogisticRegression Machine Learning model allow for successful flow detection in participants via either instrument (*i.e.*, both measurement instruments detect flow correctly when used on their own, independently of each other, as per their respective authors) as well as when paired together, dimensionally contributing to flow detection.

Moreover, the resulting Machine Learning model (Ramírez Luelmo, El Mawas, Bachelet, et al., 2022) detects flow asynchronously and automatically by using only the EduFlow-2 instrument, while featuring very acceptable metrics (> 0.8) for a participant's self-reported-based Machine Learning model.

This is illustrated in Figure 7-19, where a hypothetical deployment of this model back in the MOOC GdP would allow for flow detection using the results of newly applied EduFlow-2 measurements, and their displaying on the MOOC's learners' or instructors' dashboard. If not hypothetical, such deployment would soon likely be legally obliged to disclose learners that their flow state (as an emotion) is being recognized through automated means.

Figure 7-19

Experiment 1: Hypothetical Production Phase



Yet, if we focus on this hypothetical case, scores for the detection of flow presence are clearly higher than for the flow's absence (first and second rows of Table 7-15), which is likely due to the target imbalance hinted at during training. The target data imbalance itself might be associated with 1) the way the psychometric tests are drafted, and 2) human bias (*a.k.a.* the nature of respondents).

Indeed, the Flow-Q measure instrument quotes situations where presence of flow is described, but it does not describe absence of flow, which is a situation none of the reviewed flow measurement instruments tackles (*cf.* Section "Flow in Educational Contexts" on page 60) but is more noticeable on an adimensional measurement instrument. This alone might explain the noticeable skew of the resulting model towards detecting the presence of flow better than its absence. We (and the questionnaire's designers) assumed that absence of flow as being the opposite of the presented text quote.

Concerning the human bias, we hypothesize that respondents might feel more inclined to answer Flow-Q's items positively if they clearly self-identify with the items' text (Flow-Q asks to self-identify with described life experiences), but instead, respondents might feel more inclined to leave the question unanswered (blank) if they do not self-identify with it, instead of simply answering 'No'. We came to this conclusion because of how such participants behaved in psychometric instruments beyond the scope of this research work.

Human bias could be expressed in a less conscient manner. Indeed, we noticed that during the data cleanup phase, removed participants tended to belong to the P2 periods (as P1 concerns demographics only). We think that, just like during expected MOOC dropout, participants more committed to the MOOC completion answer our questionnaires more accurately, hence a larger proportion of P3 respondents ended up in the final sample, compared to P2.

Nevertheless, as a validating stage for Experiment 2, results (all metrics ~ 0.8) confirm a consistent relationship between both flow measurement instruments and ratify the pertinence of both flow measurement tools when working in tandem. Ensuing Section "Experiment 2 – Prototype" promptly benefits from these favorable results by developing a Machine Learning flow-detecting trained model.

Experiment 2 – Prototype

The construction of the prototype is conditioned to the validation of both flow measurement instruments (Flow-Q & EduFlow-2) working in tandem, alongside the Logistic Regression classifier, covered in the previous Section.

Instead, the present Section covers the extensive data processing applied to the MOOC's learners' digital traces (logs), the complex phase of training an imbalanced dataset, and the explanation of its results and metrics.

Logs data filtering, cleaning, and aggregating

Besides the already cleaned-up questionnaire's data employed during Experiment 1, we filter and clean up the MOOC logs, prior to their aggregation.

First, documents in the logs were discarded when showing missing usernames and/or missing session[147] identifiers, a phenomenon acknowledged in the Open edX documentation:

Occasionally, an event is recorded with a missing or blank context.user_id value. This can occur when a user logs out (or the login session times out) while a browser window remains open. Subsequent actions are still recorded in the log system but the system cannot supply the user identifier. EdX recommends that you ignore these events during analysis[148].

We removed such invalid documents (line 26) concurrently to the merging of all Sessions logs (2020a, 2020b, 2021a, 2021b) into a single collection (line 36) using the MongoDB query depicted in Figure 7-21 (partial source code of 1-202xy_MergeAll.js). This process is illustrated in the DFD shown in Figure 7-20 where multiple collections are consolidated into a single MongoDB collection.

[147] Please notice that differentiate in this Section the meaning of "session", as the time period internally determined by the Open edX platform that starts with a user login, from a GdP "Session" (initial uppercase) employed so far.
[148] https://edx.readthedocs.io/projects/devdata/en/stable/internal_data_formats/tracking_logs.html

Figure 7-20

DFD for Cleaning-Up, Filtering, and Merging All Collections

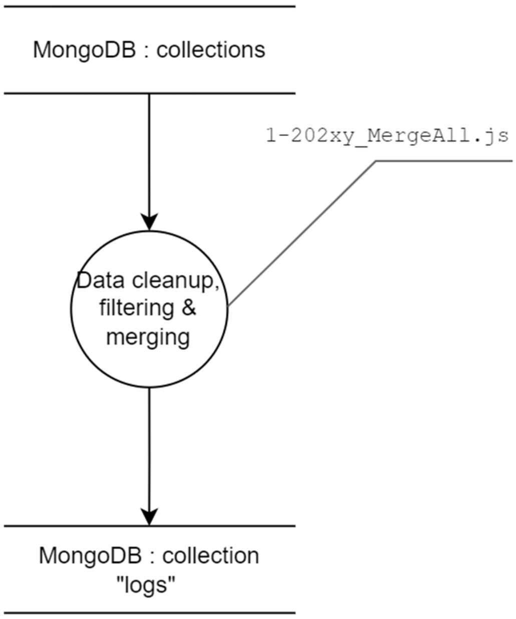


Figure 7-21

MongoDB Query for Merging All Collections

```
*/

db.logs2020a.aggregate( [ // It starts with 2020a
  //
 { $set: { _id_origin: "2020a" } }
 , { $unionWith: { coll: "logs2020b", pipeline: [ { $set: { _id_origin: "2020b" } } ] } }
 , { $unionWith: { coll: "logs2021a", pipeline: [ { $set: { _id_origin: "2021a" } } ] } }
 , { $unionWith: { coll: "logs2021b", pipeline: [ { $set: { _id_origin: "2021b" } } ] } }

 // Remove invalid documents:
 , { $match: { username: { $ne: "" }, session: { $ne: null } } }

 // Limit the total number of Documents:
 //, { $limit : 500 }

 // Sort by the original Collection, then by the original _id, then by the time field
 // This stage takes hours to complete (over 2h!!)
 , { $sort: { _id_origin: 1, _id: 1, time: 1 } }

 // Save the whole export into a new Collection in tmp
 , { $out: { db: "output", coll: "logs" } }

], { allowDiskUse: true }
)
```

Note: This query also filters out invalid usernames and invalid sessions.

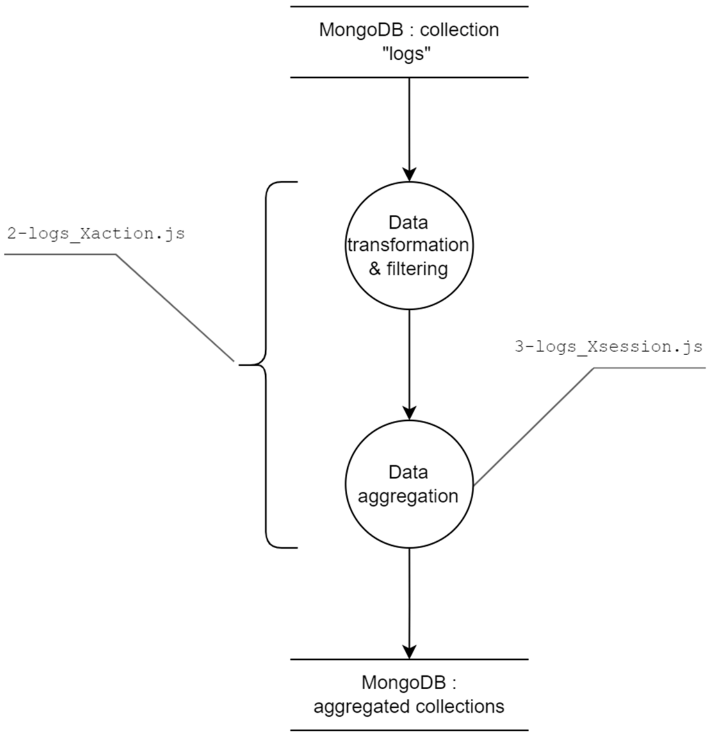
Then, to construct the indicators expressed in Subsection "Indicators construction" above, we filtered, re-categorized (data transformation and extraction), combined, and aggregated Raw log data into Derived data, to join the data sample employed in Experiment 1.

1. Filtering: we only work with data from learners whom we have survey data of,
2. Transforming: we extract and/or transform relevant portions of Raw data, at the learner's individual actions' level. This process yields xACTION Derived data, *e.g.*, calculating the weekday from a date.
3. Aggregating:
    a. We group the resulting xACTION Derived data in the scope of each learner session, *i.e.*, a learner's actions are accounted xSESSION, for all his/her sessions, *e.g.*, counting the number of navigation-related events.
    b. We group the resulting xSESSION Derived data in the scope of each individual learner (xUSER), *e.g.*, calculating minimums, maximums, and means of values for all sessions.
4. Merging: we finally calculate the final aggregations with the Experiment 1 survey input dataset, yielding the needed flow indicators again, xUSER.

These processes are depicted in the DFD in Figure 7-22 (processes one to three), where the merged collection "logs" is processed by two consecutive scripts (2-logs_Xaction.js & 3-logs_Xsession.js). Final aggregation and merging (process 4) is illustrated in Figure 7-25 and performed by the 4-agg_mongo_db.py script.

Figure 7-22

DFD for Data Transformation, Filtering, and Aggregation



One examples of a simple data transformation is parsing datetime fields and converting them to Unix time[149] for straightforward comparison between distinct platforms' datetime values, *i.e.*, the Open edX (logs) and the LimeSurvey (flow measurement data) platforms.

Another such example is extracting partial information from a value, *e.g.*, creating a new value that contains the weekday name (Saturday, Sunday, etc.), or categorizing existing log values to bestow human meaning.

In that regard, it is important to highlight that we face a MOOC, a commercial one for that matter (featuring a commercial facet and a working business model clearly described), not designed specifically with research purposes in mind. Therefore, the data harvested from the MOOC does not originate from a scientific device specialized in learning data collection, but it is instead a general log system.

149 Seconds elapsed since 00:00:00 UTC on 1 January 1970, the beginning of the Unix epoch.

That is to say that such MOOC's log data inherently lacks "thickness"[150] (Peraya & Luengo, 2019, p. 3) which would describe in detail the pedagogical activity surrounding the MOOC's participant's actions, which in turn would likely facilitate the identification of learners' psychological state (Abyaa et al., 2019) and thus, facilitate flow measurement.

Therefore, to impart some "thickness" to the available logs, we assign human meaning to the types of log events present in the traces by aggregating and categorizing them, a method already employed in trace analysis (*cf.* Subsection "Trace aggregation: classification of entries" above), *i.e.*, we add Contextual data to Raw data to create Additional data to set up in turn Derived data. Contextual data comes from the Open edX documentation on students' events stored in the tracking logs (*cf.* footnote [132] above).

Indeed, among the many fields created and tested for this study, for the sake of brevity we only delve into two: the events re-categorization and the MOOC's path locator. The former was successfully employed for data aggregation while the latter constitutes an example of dismissed Derived data. Other considered fields are described in Appendix 14. – Constructed fields for trace analysis.

Understanding both examples is not crucial for understanding the resulting modelling dataset and the reader can safely skip to Chapter 7(iii) below: Resulting modelling dataset.

(i) Example: events re-categorization. We proposed ten event categorizations in a newly created field STRING_edx_cat (a summary describing the events included is shown in Table 7-16, with full event details located in Appendix 13). Regrettably, the Open edX did not provide documentation for all events we found in the logs, and we were bound to create a category for such case "cat_general". Now, while some of these could be guessed from their name (and could

[150] « *épaisseur* », in the original text.

hypothetically been employed in this study), the lack of documentation prevented us from properly exploiting the fields in the branches attached to these events.

Table 7-16

Proposed Categorization of Open edX Student Events

Category name	Examples of types of edX events
cat_forum	Searching the forums, viewing the threads, voting answers.
cat_forum_post_comments	Creating a post, a thread or replying to a post.
cat_video	Play, pause, stop, change speed of the video; show/hide the transcript/subtitles, language settings, etc.
cat_admin	Events that can only be initiated by the staff, such as add/remove users from cohorts/groups, create/edit exams, create/edit certificates, etc.
cat_assessments	Starting a proctored exam, submitting it, counting the attempt, etc.
cat_problem_sessions	Answering online exercises and getting a hint.
cat_system	Events generated by the MOOC itself (not the learner, not the staff), such as the notification of finished calculating grades, finished creating a certificate, finished assigning content, finished reindexing, etc.
cat_navigation	Saving/removing/accessing user bookmarks; changing the course tab, and/or moving forward/backwards on the pedagogical sequence, for a given module.
cat_general	A general container for all events not described in the official Open edX documentation.
cat_unknown	A catch-all category for unaccounted events. None showed up.

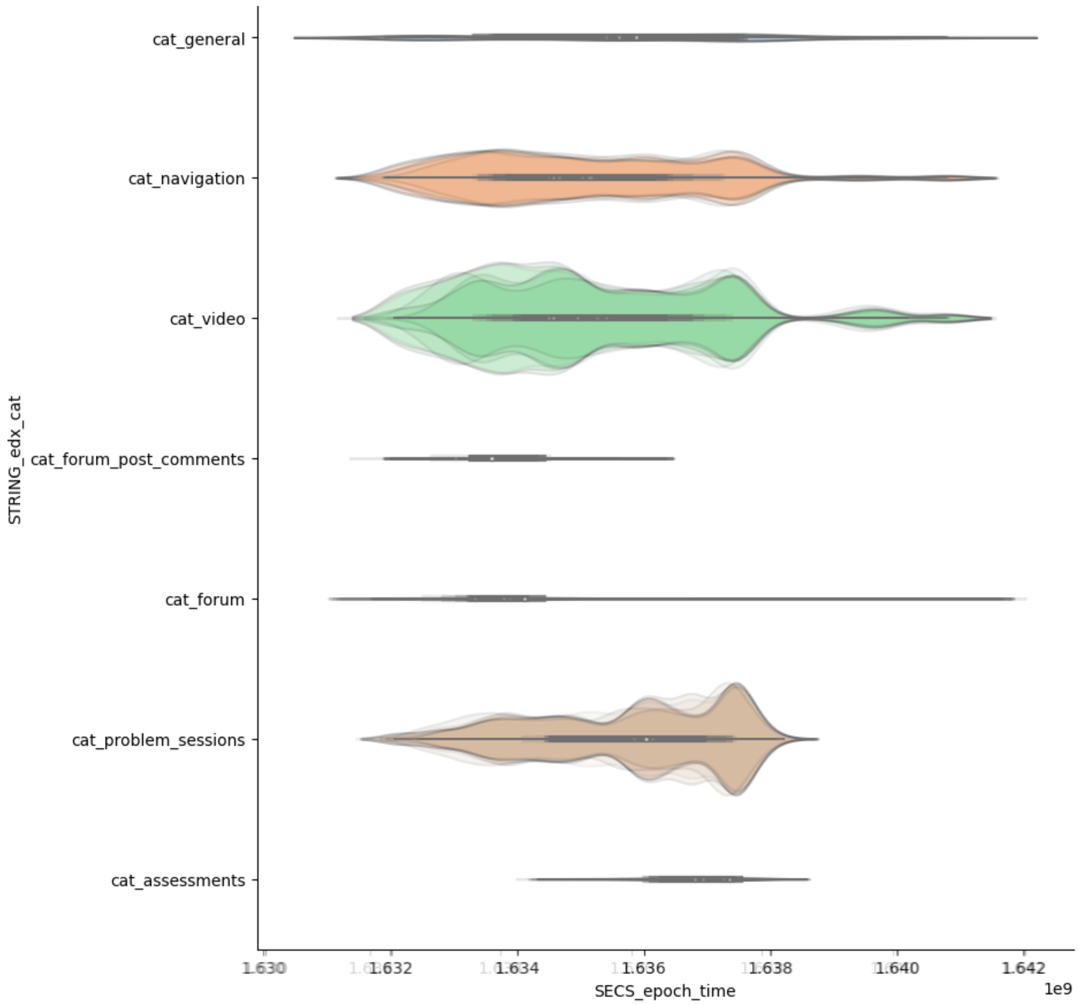
We disregarded all events belonging to categories MOOC learners could not possibly be the origin of, *i.e.*, cat_admin & cat_system, such as events initiated only by the staff (create cohort, create proctored exam, etc.), by the system itself (reindexing finished, finished calculating grades, etc.), or not explained in the official documentation ("cat_general", and "cat_unknown").

This categorization allowed us to focus the trace analysis on six families of events instead of attempting to extract meaning from ~90 singular-type edX events, of which some are very likely to come one after another, *e.g.*, play_video, pause_video, justifying thus their family-grouping.

The outcome of this categorization is better understood when looking at Figure 7-23 and Figure 7-24, where we depict six superimposed violin plots[151] of randomly selected learners from the MOOC's college (university), and open-public demographics respectively.

Figure 7-23

Superimposed Violin Plots of Six Selected Learners (College)



[151] A violin plot shows the distribution of quantitative data across several levels of one (or more) categorical variables. The violin plot features a kernel density estimation of the underlying distribution (Hunter, 2007; Waskom, 2021).

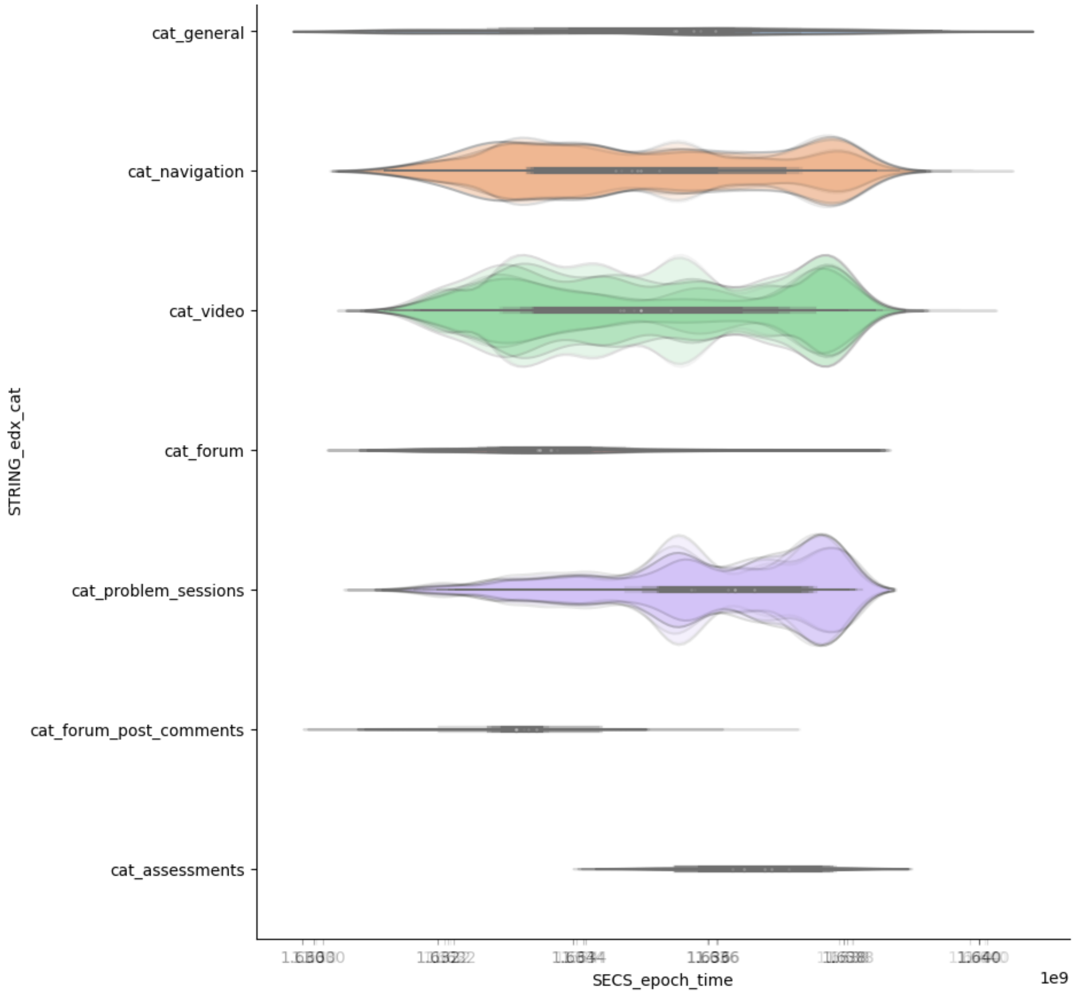
Note: The randomly selected participants (*n* = 6) employed in this plot belong to the college demographic.

Proposed re-categorization (STRING_edx_cat) values are shown on the ordinate, while Unix time (SECS_epoch_time) is shown on the abscissa. Superimposing participants' plots is roughly attempted via graph's transparency. It is important to remark that exact values do not matter here but the overall figure shape. Data employed for plotting comes only from the second MOOC Session of 2021.

Both figures consistently share shape with the rest of the MOOC participants belonging to the same demographic. Colors are meaningless, and unfortunately categories swap order between both figures. We include the dismissed cat_general violin plot to make obvious its reduced role when facing other, more prominent categories.

Figure 7-24

Superimposed Violin Plots of Six Randomly Selected Learners (Open-Public)



Note: The randomly selected participants (n = 6) employed in this plot belong to the open-public demographic.

Please note that college (university) learners display similar shape patterns both in the time axis (bottom), and per category. This is not the case for the open-public MOOC participants where we had trouble aligning their time axis. Being unbound by curricula is the most likely explanation for such heterogenous practices.

Also, please note that both figures attest to extensive video, navigation, and exercise-solving tasks, when compared to forum and evaluation-related (during both tests and proctored) activities.

(ii) Example: absolute path-locator. Another proposed field is the proposed absolute path-locator STRING_edx_path which attempts to capture in a single text string the absolute location (visited web page) of the learner within the most common path sequence of the MOOC. We posit that the difference between a chain of multiple STRING_edx_path fields and the MOOC's most-taken path would yield an indicator of how much learners deviate from the majority's path, suggesting in turn an indicator for flow's D1.

Thus, the field STRING_edx_path was composed by the concatenation of contextual data located in the branch context.path of the newly constructed STRING_edx_cat field, *i.e.*, the proposed path depends on the family of events it originates from. The pseudocode[152] for creating field STRING_edx_path is detailed in Table 7-17.

Table 7-17

Pseudocode for Creating Field STRING_edx_path

STRING_edx_cat	STRING_edx_path pseudocode
cat_video	event.code [IF IT EXISTS] OR event.id
cat_navigation	CONCAT(event.old, event.new) [IF IT EXISTS]
cat_navigation	CONCAT(event.current_url, event.target_url) [IF IT EXISTS]
cat_forum	CONCAT(event.category_id, event.target_username, REPLACE(event.title, " ", "_"))
cat_forum_post_comments	CONCAT(event.category_id, event.thread_type, LENGTH(event.body), event.options.followed) (NOT SEEN)
cat_problem_sessions	LAST(event.problem_id, "@") [IF IT EXISTS]
cat_problem_sessions	LAST(event.problem, "@") [IF IT EXISTS]
cat_assessments	event.exam_id
cat_general	context.path

So, for each of our categories, we select (and construct) data from fields that might represent the learner's visited web pages within the MOOC. We aim to discard details too specific to the learner itself, such as their unique event id, or the total length of the video selected, or the text title course of the MOOC, or the location of the navigation pane on the page, or the exact text posted on a forum (as we don't analyze their content), etc.

152 All CONCAT operations require an appropriate separator ("_" or "/" or "*", depending on the content of the original concatenated strings.

However, preliminary Feature Selection analysis on partial input training datasets showed such text strings were inadequate for flow detection because:

- Despite our efforts, such text strings were not simple enough to be processed as-is per the Logistic Regression classifier.
- Re-encoding such text strings would have implied an additional, distinct variety of Machine Learning methods which tend to generate more features and thus spread values into a larger hyperspace[153] which would also hinder the intended white box approach.
- These text strings carried information rather specific to the MOOC's staff organization and not to the actual path taken by the learner.
- Obtaining a "difference" (or "distance") from any given reference point necessitated first to calculate such reference point, which could not be known before the end of the MOOC.
- But most importantly, because employing such text strings in any way called into question if flow's D1 could really be hampered by the learner remaining in the confines of a given path in the MOOC.

However, this reflection still brought up the notion of temporality for validating the learners' digitals traces.

Indeed, we concluded that we could not exploit traces generated after our flow measurement tools have been applied because we could not demonstrate nor if they corresponded to a flow-inducing nor to a flow-induced behavior in the MOOC, *i.e.*, we were constrained to exploit log raw data generated up to the moment before submitting the surveys.

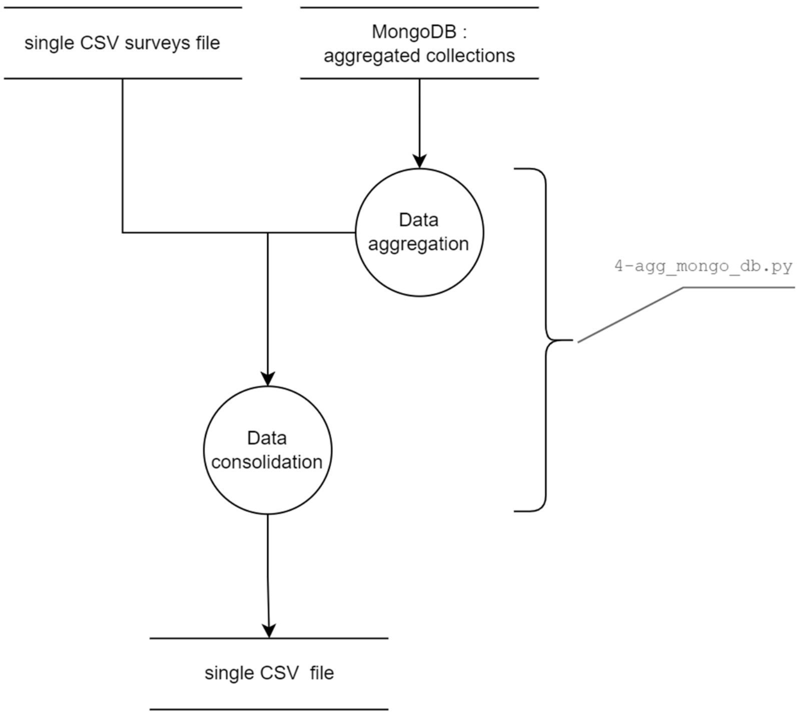
[153] *Cf.* the curse of dimensionality.

Thus, we effectively limited log data processing to learners we knew their survey's submitting timestamp, which we reflected in a newly created best_before field. By comparing any given learner's session start time to this value we accounted or dismissed the entire session.

This process is performed by script 4-agg_mongo_db.py and illustrated in the DFD in Figure 7-25 where we see the CSV surveys file from Experiment 1 is employed during the data aggregation phase to filter out irrelevant learners' digital traces from the already aggregated collections in MongoDB. Aggregated data is finally merged into a final CSV file.

Figure 7-25

DFD for Final Data Aggregation and Merging



Preliminary Feature Selection analysis on partial input training showed a few of the considered indicators hindered target prediction as noise and were dismissed during this cleaning phase. A summary of such indicators is shown in Table 7-18 while a more complete list is found in Appendix 14. – Constructed fields for trace analysis.

Table 7-18

Constructed Fields Discarded During FeatureSelection Analysis

Indicator field name	Description	Dismissal reason
BOOL_multiday_session	Per-session field: activated if the current session crosses midnight.	Non informational: contained in other features.
BOOL_saturday_q	Per-session field: activated if the session started on a Sunday.	Non informational: contained in other features.
time_diff	Per-row field: Minutes between first and last connection.	Non descriptive: sessions of different lengths can happen in-between.
STRING_time_day	Per-action field: In French, the name of the timeslot of the day when the action happened.	Non informational: it becomes diluted when considered per-user.
STRING_edx_path	Per-action field: attempt to merge several fields in one concatenation reflecting a path positioning.	Unreliable and too specific to the MOOC because of labels and edX identifiers.
AverageActionsPerMinute	Per-session field: Average number of actions per minute for that session.	Non descriptive as a cumulative of sessions.

(iii) Resulting modelling dataset. Finally, to give the Machine Learning model a sense of minimums and maximums, we provided the total number of valid events and sessions generated in the MOOC, totaling the list of indicators shown in Table 7-19:

Table 7-19

Fieldnames: Indicators from Aggregated Data

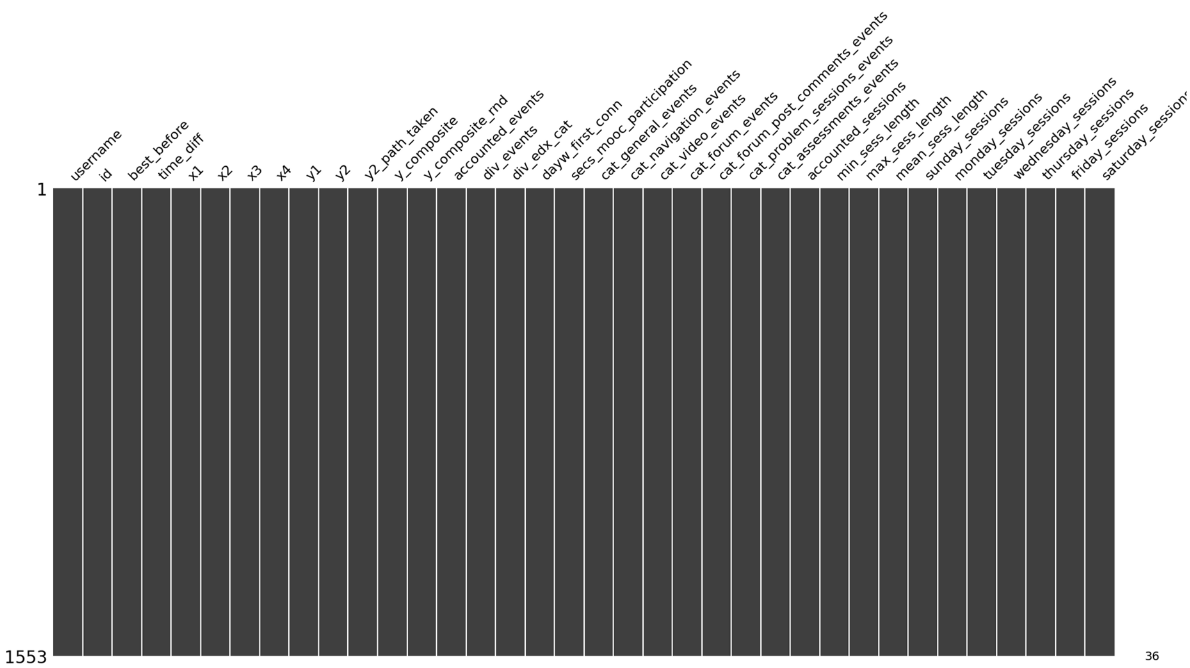
Indicator	Description
accounted_events	Total number of valid events generated.
div_events	Event's diversification: total number of different types of events created.
div_edx_cat	Categories' diversification: total number of distinct families of events created.
dayw_first_conn	Day of the week the user first logged in.
secs_mooc_participation	Sum of all seconds spent logged in.
cat_navigation_events	Number of navigational events.
cat_video_events	Number of Video-related events.
cat_forum_events	Number of Forum-reading events.
cat_forum_post_comments_events	Number of Forum-participating events.
cat_problem_sessions_events	Number of Exercise-solving events.
cat_assessments_events	Number of Exam-solving events.
accounted_sessions	Number of logged sessions.
min_sess_length	Minimal duration of a logged sessions, in minutes.

max_sess_length	Maximal duration of a logged session, in minutes.
mean_sess_length	Average duration of a logged session, in minutes.
sunday_sessions	Count of sessions started on a Sunday.
monday_sessions	Count of sessions started on a Monday.
tuesday_sessions	Count of sessions started on a Tuesday.
wednesday_sessions	Count of sessions started on a Wednesday.
thursday_sessions	Count of sessions started on a Thursday
friday_sessions	Count of sessions started on a Friday
saturday_sessions	Count of sessions started on a Saturday.

Thus, the shape of the modelling dataset CSV file is shown in Figure 7-26, including anonymized identifiers, flow measurements, and 23 features. These indicators align themselves with those widely made available by MOOCs and employed by researchers on the field, as recently discovered by a literature review on engagement indicators issued from log data in MOOCs (Sharif & Ramakrisnan, 2023, tbl. 9): videos (and their related interactions) are the primary source of pedagogical indicators, seconded by events related to the discussion forums, followed by assignments (private interactions between learner and instructor), finalizing with quizzes interactions (Sharif & Ramakrisnan, 2023, sec. Discussion).

Figure 7-26

Graphical Representation of the Cleaned-Up Dataset



Note: This representation is for a n = 1 553 participants CSV file.

Interestingly, limiting digital traces' exploitation at the time of submitting the surveys brought up two situations:

- first, we entirely discarded a few participants who answered both our flow measurement instruments without generating relevant MOOC events (*a.k.a.* either answering "too early" or never logging in the MOOC);
- second, we discarded participants who did not finish the entirety of the Research & Development's staff psychometric tests (even if they did complete ours) because we did require the survey's submitting timestamps for filtering out raw log data.

Additionally, one learner had malformed log data which also accounts for the discrepancy (36) in the final headcount between Experiment 1 (n = 1 589, *cf.* Figure 7-13) and Experiment 2 input datasets (n = 1 553, *cf.* Figure 7-26).

Based on the conclusions from Experiment 1, a new target y was calculated from weighing both flow measurement instruments at a 70/30 ratio, with the EduFlow-2 score prevailing over that of Flow-Q because of its educational component.

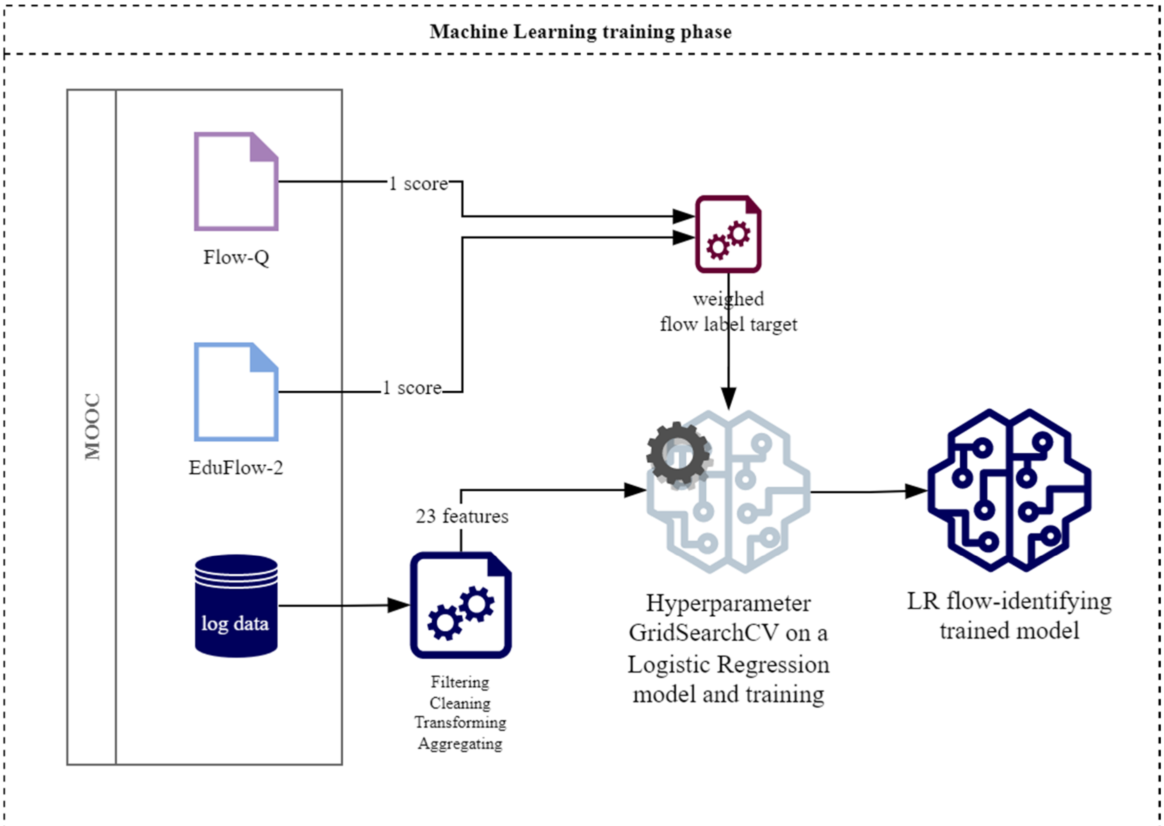
However, this newly calculated target y showed a high data imbalance distribution (frequency of '1' values more than doubled frequency of '0' values), which we tackled during the model training phase.

Model training

The training phase of Experiment 2 is illustrated in Figure 7-27: the MOOC GdP yields two flow scores (weighed-in to constitute a single flow label) plus 23 indicators ("features") issued from the filtering, cleaning, transforming and aggregating of the MOOC's learners' digital traces (*cf.* Subsection "Logs data filtering, cleaning, and aggregating" above). These are in turn fed to a GridSearchCV process on a Logistic Regression pipeline (which includes diverse preprocessors), yielding a distinct Logistic Regression flow-detecting trained model capable to identify flow using only the MOOC's learners' digital traces of an entirely different input sample.

Figure 7-27

Experiment 2: Training Phase



However, due to the impossibility of accessing the MOOC for the deployment phase of the resulting trained model, we simulated it by first randomly[154] splitting and reserving 30% of data created in previous step "Logs data filtering, cleaning, and aggregating" as test data (shown in Figure 7-28 in light orange).

The remaining 70% modelling data (light blue) is further randomly divided into training (gray) and evaluation (red) sets during training at an 80/20 ratio. The evaluation set is reserved

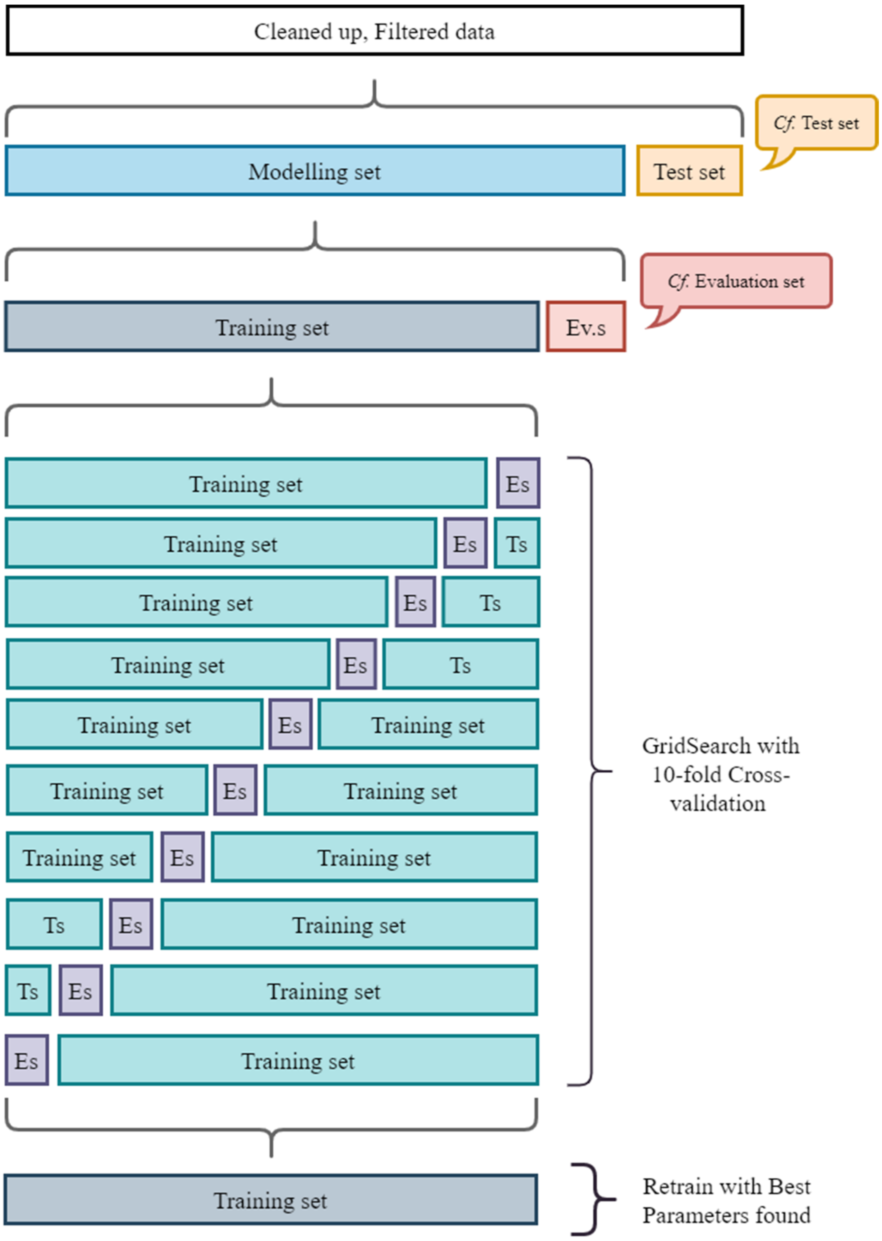
[154] While accounting for target data imbalance.

for metrics assessment at the end of model training (*cf.* Subsection "Cross-validation & data splits").

The training set happens to be divided again internally (cyan and dark violet) during the 10-fold Cross Validation task of the hyperparameters GridSearch but then re-conformed (gray, bottom) to train the automatically selected hyperparameters yielding the best metrics during training.

Figure 7-28

Sequence of Modelling/Testing & Training/Evaluation Data Splitting

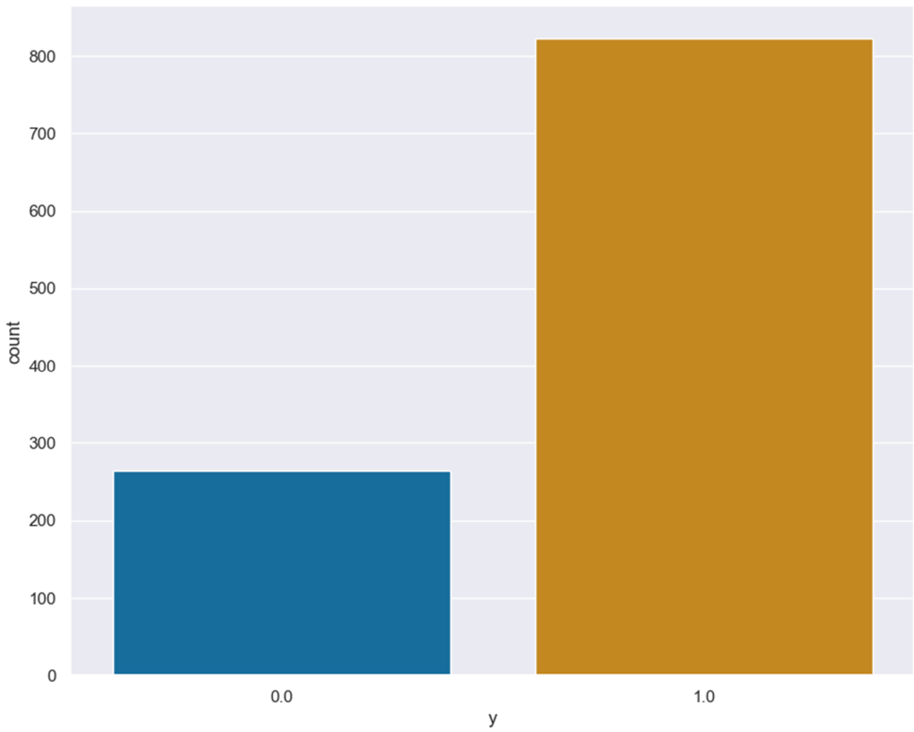


Note: Figure is not at scale. To be read top-to-bottom. Rectangles represent data splits, with data imbalance and split percentages intentionally not depicted for the sake of clarity. Text balloons point to nomenclature definitions.

Contrary to Experiment 1, preliminary testing showed that high data imbalance distribution in target *y* (shown in Figure 7-29) did hinder model training and evaluating metrics. This is a rare case where the positive class (1 – flow presence) is majorly represented instead of being the minority class[155].

Figure 7-29

Target (Flow Absence vs. Flow Presence) Imbalance in the Final Dataset



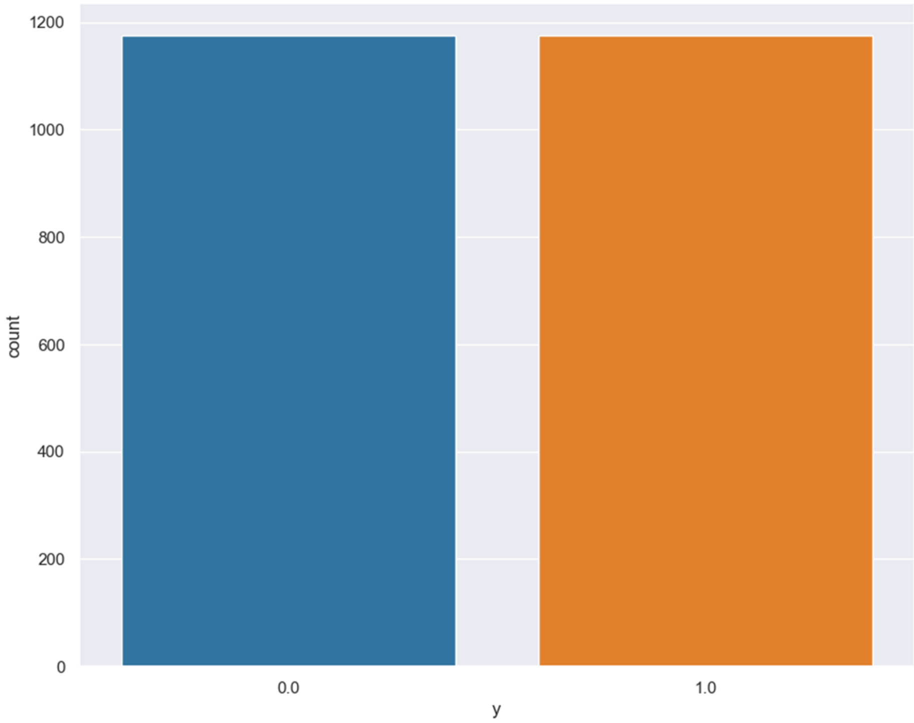
We tackled this issue by employing the SMOTE algorithm implementation in the imblearn library (Lemaître et al., 2017). It is selectively applied during training, and *k*-folding, but not during evaluation nor testing, by the GridSearchCV task automatically. An example of this

[155] Most imbalanced datasets commonly showcase reduced examples of the positive class, *e.g.*, during fraud detection, fraudulent transactions tend to be orders of magnitude smaller than usual transactions.

algorithm's results is shown in Figure 7-30, where the target *y* balance has been restored for training purposes. Also, all data splits shown in Figure 7-28 account for the prevailing data imbalance by stratifying the split.

Figure 7-30

Example of a Target-Balanced Dataset

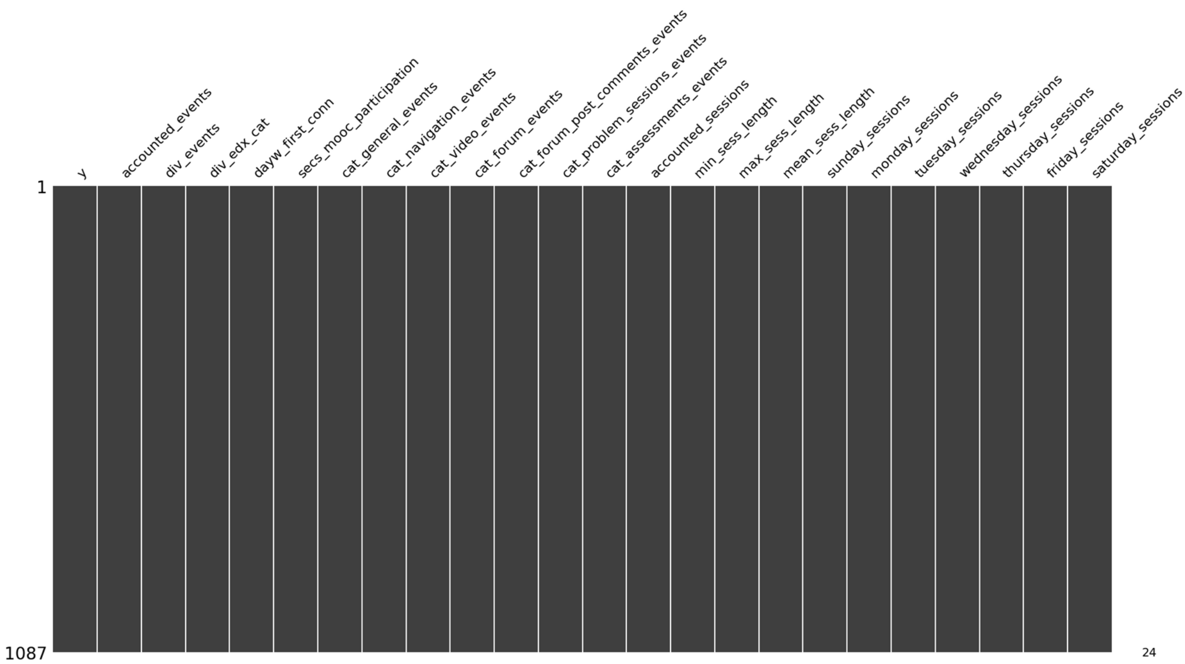


Note: The balance is restored after applying the SMOTE algorithm.

Thus, the modelling dataset (blue in Figure 7-28) included 23 features and one target *y*, ready for training a Machine Learning model. A graphical representation of this modelling dataset is shown in Figure 7-31 also showcasing no missing data.

Figure 7-31

Graphical Representation of the Modelling Dataset



Note: This representation is for the 70% (*n* = 1 087) CSV file.

Experiment 2 was run in a Google Colaboratory (Google Colab) remote instance using a local kernel which was detected by CodeCarbon as a Python 3.10.6 on a Linux-5.15.0-56-generic-x86_64-with-glibc2.35 running Ubuntu-20.04-focal with 30.99 GB RAM, 12 Intel(R) Core(TM) i7-10850H CPUs @ 2.70GHz and one 1 x Quadro RTX 4000 GPUs, in France (hauts-de-france). Permanent caching of the pipeline was effectuated in an external Joblib file.

We employed the generic Logistic Regression classifier from scikit-learn built-in Machine Learning Logistic Regression classifier for Python (*cf.* Subsection "Model training"). To tackle the impending data imbalance issue, we employed the specifically designed substitute Pipeline class from the imblearn library (Lemaître et al., 2017). This drop-in substitution to scikit-learn's own Pipeline class specifically balances the dataset during fit but not during transform and sample methods.

Based on previous experience from Experiment 1, we manually designed a base pipeline as shown in Figure 7-32. Pipeline begins with the function transformer $\log(1 + x)$, then it is followed by the StandardScaler preprocessor, the SMOTE algorithm sampler, the Principal Component Analysis (PCA) preprocessor, and finally concludes with the LogisticRegression classifier. Descriptions of these tasks' roles are found in Table 7-20.

Table 7-20

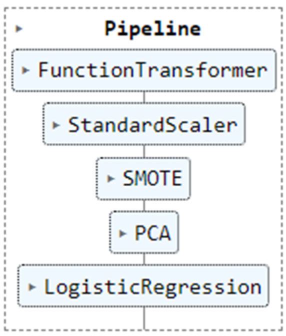
Roles' Description for the Base Pipeline Steps

Pipeline step	Role description
Transform log(1+x)	Accentuates features' values further (normalization)
StandardScaler	Removes means and scales to the unit variance
SMOTE sampler	Selectively re-balances input data
PCA	Reduces data dimensionality
LogisticRegression	Classifier for training

Fixed hyperparameters were the LogisticRegression classifier's class_weight='balanced', and all functions' random_state=42[156].

Figure 7-32

Base pipeline for Training



A GridSearchCV task exhaustively combed through 144 combinations of the base pipeline hyperparameters shown in Table 7-21 to determine the one yielding best metrics' values. Each combination constituted a trained instance of the Machine Learning Logistic Regression

156 "Answer to the Ultimate Question of Life, The Universe, and Everything" (Adams, 2005).

model and thus it was evaluated via an internal, stratified 10-fold Cross Validation task, totaling 1 440 trained models $(4 \cdot 2 \cdot 3 \cdot 3 \cdot 2 \cdot 10 = 1440)$.

It is noteworthy to mention that initially we considered further hyperparameters to evaluate, but experience from both Experiment 1, and current Experiment 2's training showed that the more successful models tended to dismiss the same set of hyperparameters. Nevertheless, we did comb through 3 600 combinations of hyperparameters of the base pipeline several times before noticing this trend and mending it. By reducing the hyperparameter space, we also decreased training and testing time, as well as our CO_2 emissions.

Table 7-21

Hyperparameters to Evaluate During the GridSearchCV Task

Hyperparameter	Pipeline step name	Values
n_components	PCA	7, 10, 17, ''mle'
svd_solver	PCA	'auto', 'full'
solver	LogisticRegression	'newton-cf', 'lbfgs', 'liblinear'
max_iter	LogisticRegression	300, 500, 1000
multi_class	LogisticRegression	'auto', 'multinomial'

Out of the 1 440 envisaged trained instances, 240 failed due to invalid combinations of hyperparameters as expected.

Once the GridSearchCV task completed and the hyperparameters yielding best metrics were found, the pipeline was automatically re-trained with the entirety of the training dataset (*cf.* previously shown Figure 4-7). The resulting trained pipeline with its currently found hyperparameters is shown in Figure 7-33.

This trained pipeline constitutes our proposed Logistic Regression flow-detecting trained Machine Learning model. It yields a flow/no-flow label in an automatic, transparent, and real-time[157] fashion, based on the learner's recent behavior on the MOOC, without applying any

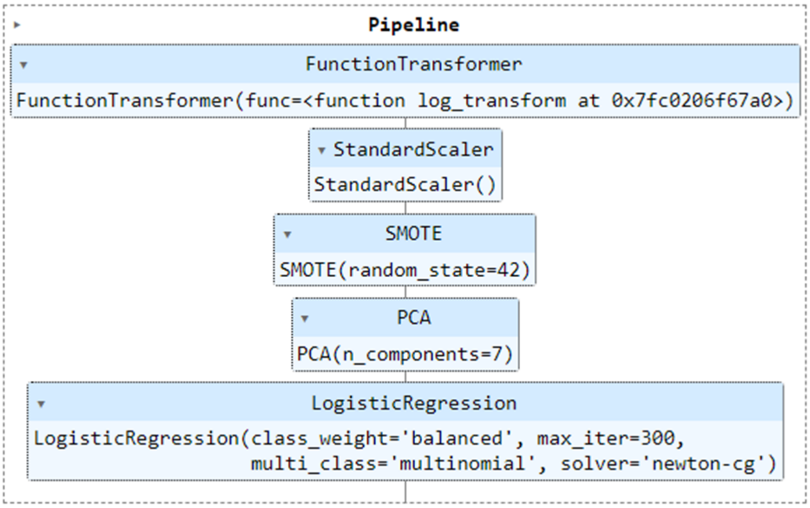
[157] Accessing, retrieving, and aggregating the necessary data for features' creation takes orders of magnitude longer than the milliseconds the model requires to yield a flow label.

psychometric test nor interrupting any activity, pedagogical or otherwise, but by simply pressing a button or accessing a webpage.

This trained model is digitally available by addressing the Trigone-CIREL laboratory research staff and it does not require any specialized intervention once set up as a secured, remote-access API.

Figure 7-33

Resulting Pipeline from the GridSearchCV Task



Acknowledged difficulties (*cf.* Related tasks above) when estimating FLOP numbers (Sevilla et al., 2023, p. 31) were clear to us as both the Python interpreter (Van Rossum & Drake Jr, 1995), and the Google Colaboratory tool (Bisong, 2019b, p. 59) (among other tools employed) add layers of machine code translation to our own model calculations before reaching any CPU. Thus, in the spirit of reporting FLOP numbers, our resulting Machine Learning model reports an estimated CO_2 consumption, tiny enough as to, without a doubt, allow us to overlook actual FLOP numbers reporting.

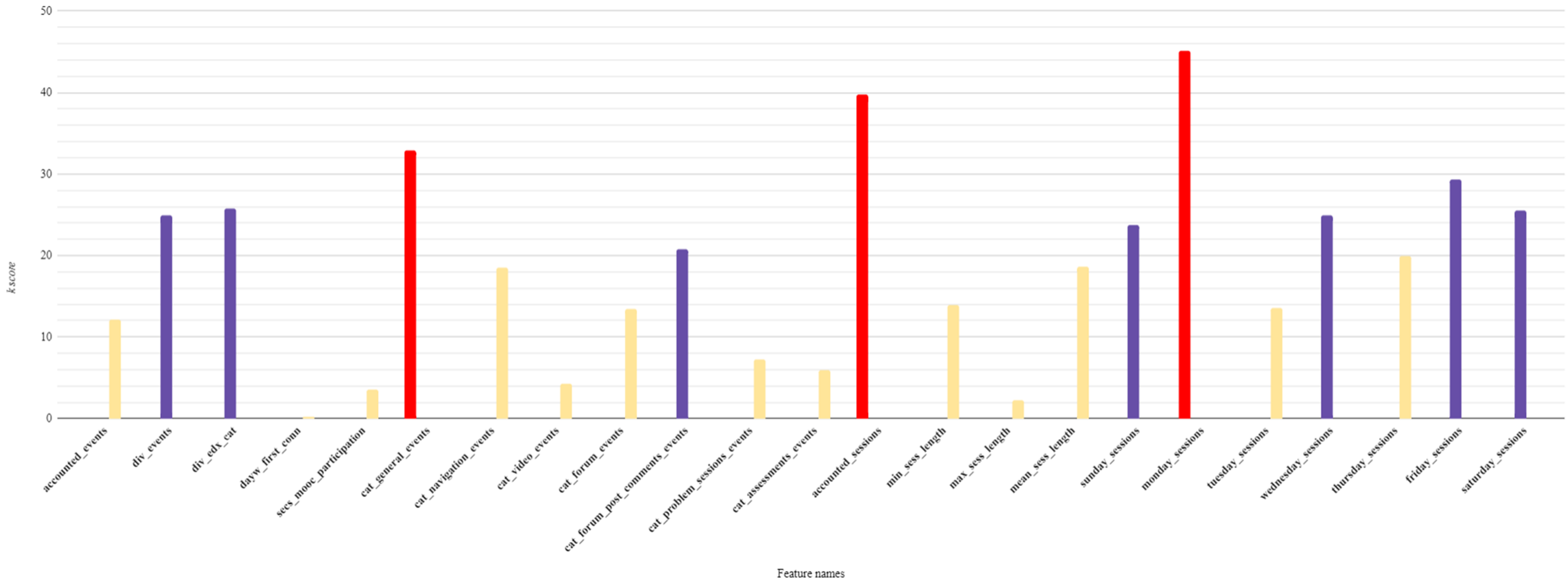
Indeed, the entirety of the previously described process barely generated 0.000075149363060589600 Kg. eq. CO_2, *i.e.*, 75 milligrams of CO_2-equivalent, as reported by

CodeCarbon. Yet, it is important to highlight that similar (or smaller) emissions were generated from the many preliminary tests and prior analysis (~ 50) conducive to this final Experiment 2.

Moreover, prior to training the actual model, under an exploratory light, we employed two ANOVA-based methods (SelectKBest and SelectPercentile: 30) for feature selection in the input sample dataset. Both methods yielded identical *k*-scores for feature importance, which we categorized in high (red), low (deep blue), and medium (yellow) importance, as can be appreciated in Figure 7-34. According to both methods' results, features "number of sessions started on a Monday", and "total number of created sessions" contributed the most to determine the presence of absence of flow, followed by the "number of undocumented events in the general category".

Figure 7-34

FeatureSelection k-scores



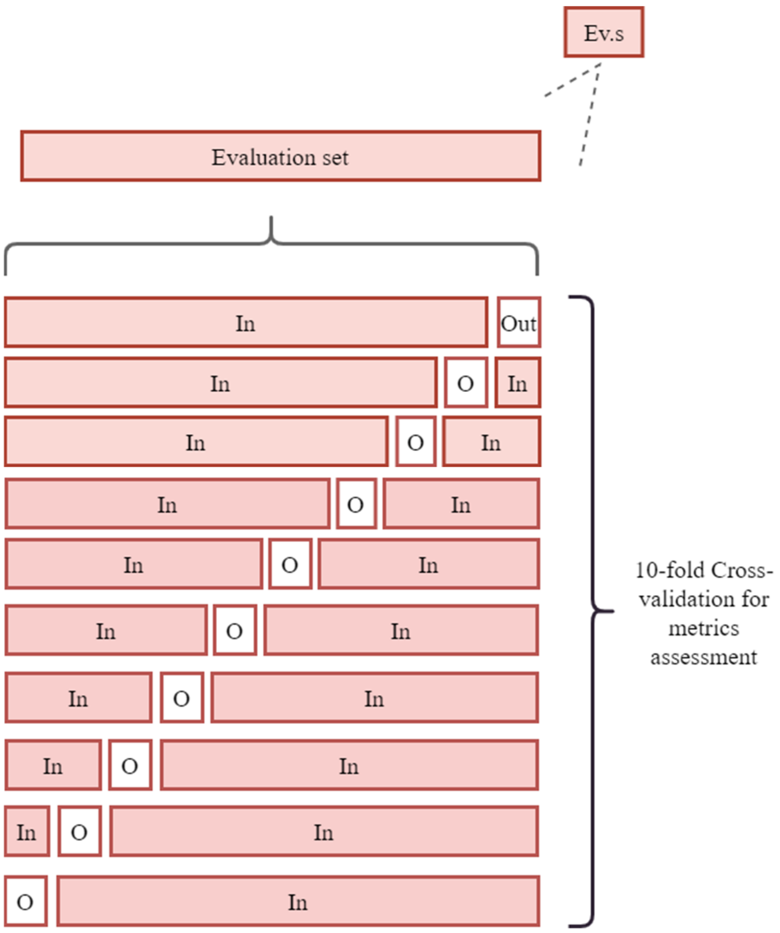
Note: *k*-scores are categorized in High (red), Medium (deep blue), & Low (yellow), according to their respective values.

Results and metrics

(i) On the evaluation set. The resulting Logistic Regression flow-detecting trained model was assessed on the evaluation set (20% out of 70%, in red, in already-shown Figure 7-28) using a stratified 10-fold Cross Validation (illustrated in Figure 7-35), following a classical Machine Learning workflow (*cf.* Section "Machine Learning workflow").

Figure 7-35

Cross Validation Task for Metrics Assessment (Evaluation)



During the stratified 10-fold Cross Validation, means of metrics relevant to imbalanced datasets (accuracy, balanced_accuracy, precision, recall, f1, roc_auc) were calculated. These metrics' values are shown in Table 7-22.

Table 7-22

Means of Metrics Applied During the Stratified 10-fold Cross Validation (Evaluation)

Test	Mean	Standard deviation
Accuracy	0.572	0.077
Balanced accuracy	0.587	0.089
Precision	0.824	0.062
Recall	0.560	0.113
F1	0.660	0.080
ROC AUC	0.634	0.051

Likewise, we used the specially crafted classification report method for imbalanced datasets from the imblearn library, which yielded the results shown in Table 7-23.

Table 7-23

Classification Report for Imbalanced Datasets (Evaluation)

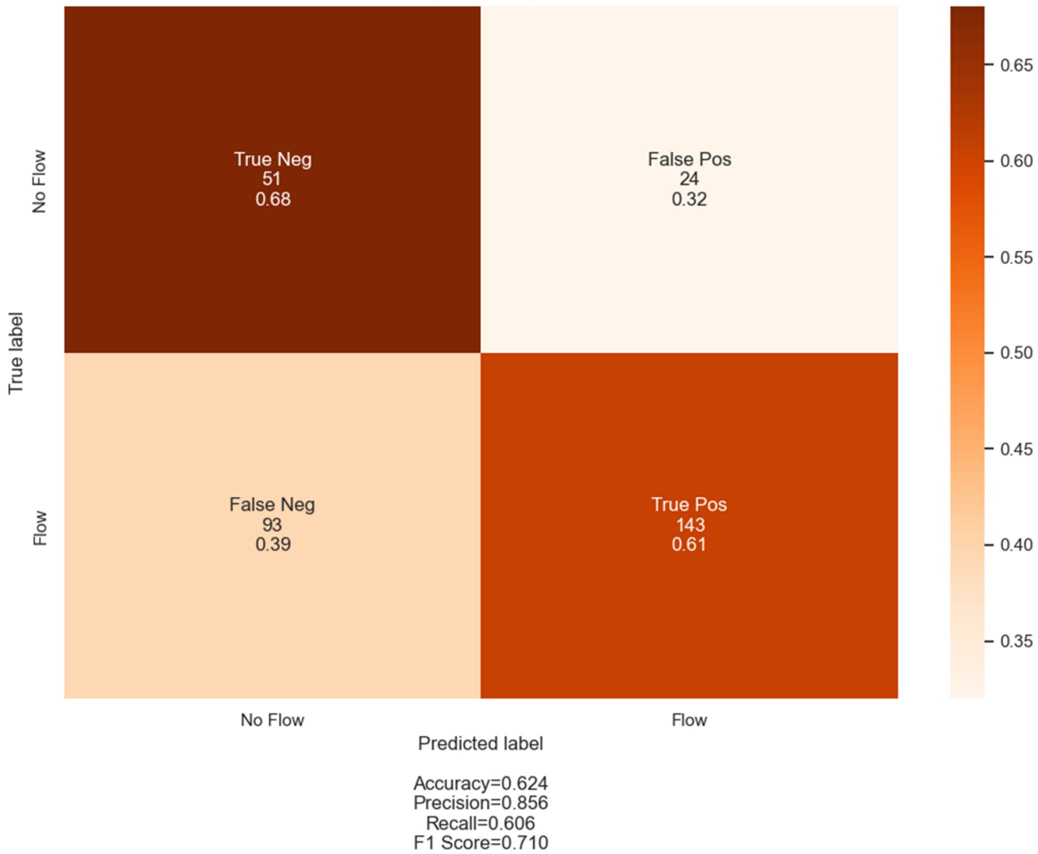
	Precision	Recall	Specificity	F1	Geometric mean	IBA	Support
No-flow	0.35	0.68	0.61	0.47	0.64	0.42	75
Flow	0.86	0.61	0.68	0.71	0.64	0.41	236
Average & totals	0.74	0.62	0.66	0.65	0.64	0.41	311

Note: All scores were rounded up at the source. IBA stands for Index Balanced Accuracy of the geometric mean.

Aside from the precision and f1 scores, other metrics behave in similar ways for both targets. Precision and f1 scores are clearly higher for flow than for no-flow (values in bold), which is mostly due to the dataset imbalance, *i.e.*, the model detects flow better than it does no-flow because it was given more training examples of what characterizes flow than of what no-flow looks like.

Figure 7-36

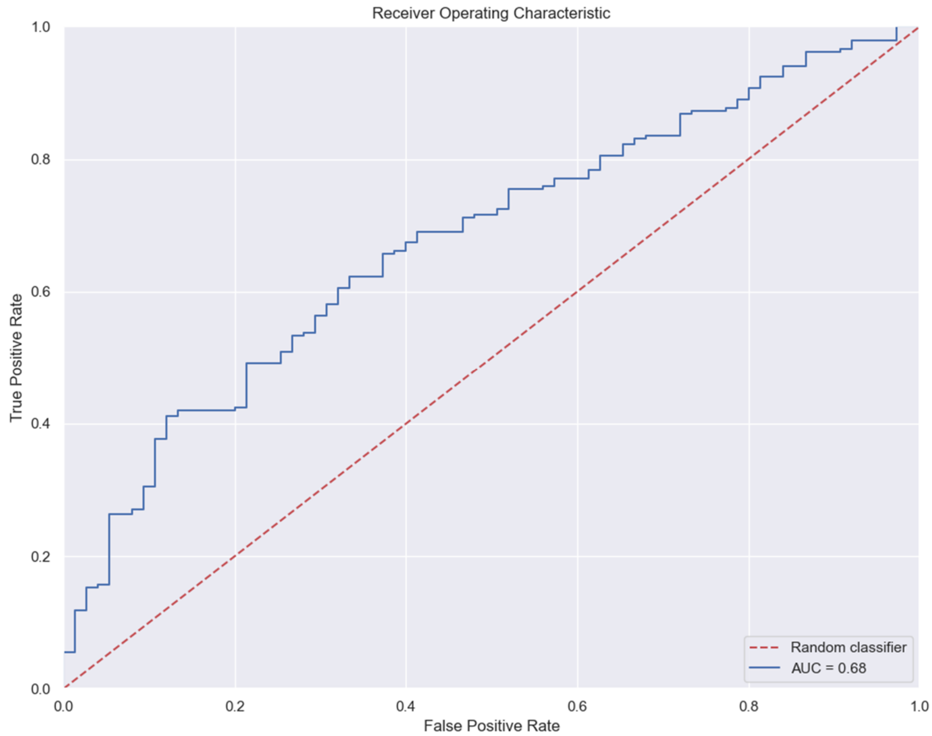
Normalized Confusion Matrix (Evaluation)



These flow-detecting results are better illustrated in the normalized Confusion Matrix (Figure 7-18): the model systematically detects True Positives and True Negatives (darker squares) better (61% & 68%) compared to the incorrectly identified cases (32% & 39%).

Figure 7-37

Evaluation ROC Curve



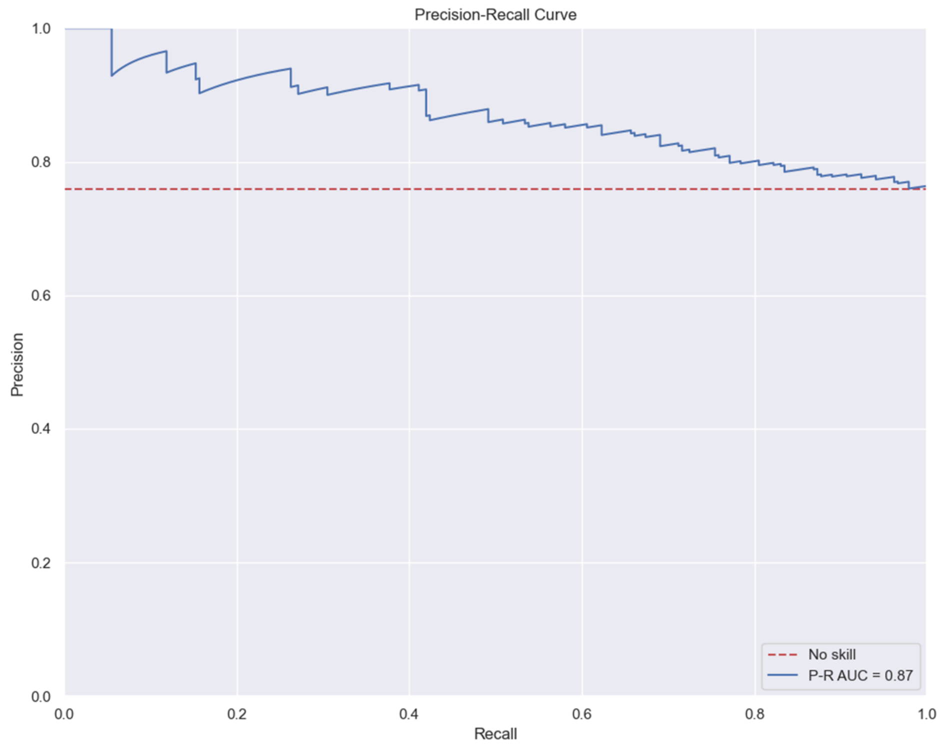
Note: ROC curve (AUC = 0.68) is depicted in blue *vs.* a hypothetical random classifier, in red.

This overall, general model behavior is better appreciated on the ROC, which features an AUC = 0.68, shown in Figure 7-37: while the model (blue line) always performs better than a random classifier (red line), it corrects itself often, visibly struggling to reach a broader AUC (the top-left corner).

Specializing in imbalanced datasets, the PRC gives a more complete understanding of the model performance on our imbalanced target. In our case, the PRC (depicted as a blue zigzagging line in Figure 7-38) features a quite successful AUC = 0.87 and constantly remains above the no-skill threshold (red line at ~ 0.76) automatically calculated from the imbalanced evaluation set.

Figure 7-38

Evaluation PRC Curve



Note: The PRC Curve (AUC = 0.87) is depicted in blue *vs*. a hypothetical no-skill model, in red.

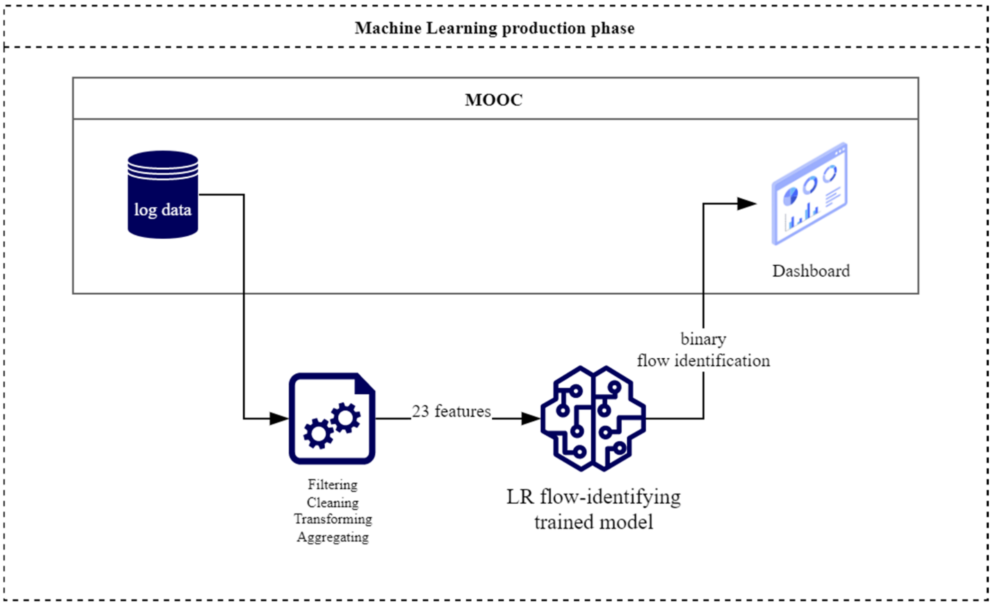
(ii) On the test set. The test set is the reserved dataset split before model training (30%, light orange in previously shown Figure 7-28). We reserved it to simulate the production phase (*a.k.a.* deployment) by assessing the model's performance on entirely untouched[158] unknown data.

This production phase would have unwound as shown in Figure 7-39: the resulting Logistic Regression flow-detecting trained model would yield a flow/no-flow label to a hypothetical MOOC flow dashboard when fed the 23 features issued from the MOOC's processed logs.

[158] Besides the mandatory stratified data 70/30 split.

Figure 7-39

Hypothetical Production Phase of the LR Flow-Detecting Trained Model



Again, a stratified 10-fold Cross Validation was performed on the test set to obtain means of metrics relevant to imbalanced datasets (accuracy, balanced_accuracy, precision, recall, f1, roc_auc), shown in Table 7-25.

Table 7-24

Means of Metrics Applied During the Stratified 10-fold Cross Validation (Test)

Test	Mean	Standard deviation
Accuracy	0.589	0.113
Balanced accuracy	0.601	0.142
Precision	0.829	0.104
Recall	0.580	0.108
F1	0.678	0.096
ROC AUC	0.656	0.145

The classification report for imbalanced datasets from the imblearn library yielded the results shown in Table 7-25.

Table 7-25

Classification Report for Imbalanced Datasets (Test)

	Precision	**Recall**	**Specificity**	**F1**	**Geometric mean**	**IBA**	**Support**
No-flow	0.34	0.69	0.58	0.46	0.63	0.40	113
Flow	0.85	0.58	0.69	0.69	0.63	0.39	353
Average & totals	0.73	0.61	0.66	0.63	0.63	0.40	466

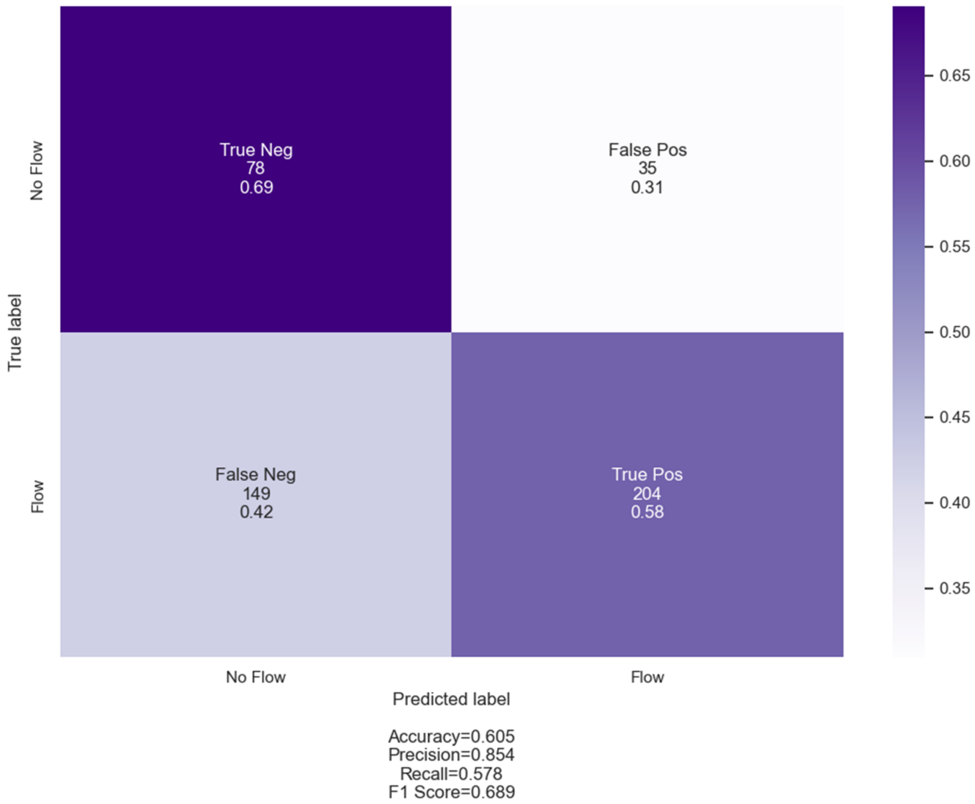
Note: All scores were rounded up at the source. IBA stands for Index Balanced Accuracy of the geometric mean.

Similar behavior is seen when assessing on the test set: metrics behave in similar ways for both targets with precision and f1 scores noticeably higher for flow than for no-flow (values in bold). This similar behavior on the new, unknown test dataset attests to the ability of the model to effectively detect flow without overfitting, despite the extreme data imbalance.

This behavior is again noticed when reviewing the normalized Confusion Matrix shown in Figure 7-40: the model still identifies both the positive and the negative classes (darker squares) better (58% & 69%) than the incorrectly identified cases (31% & 42%) on an entirely unknown dataset.

Figure 7-40

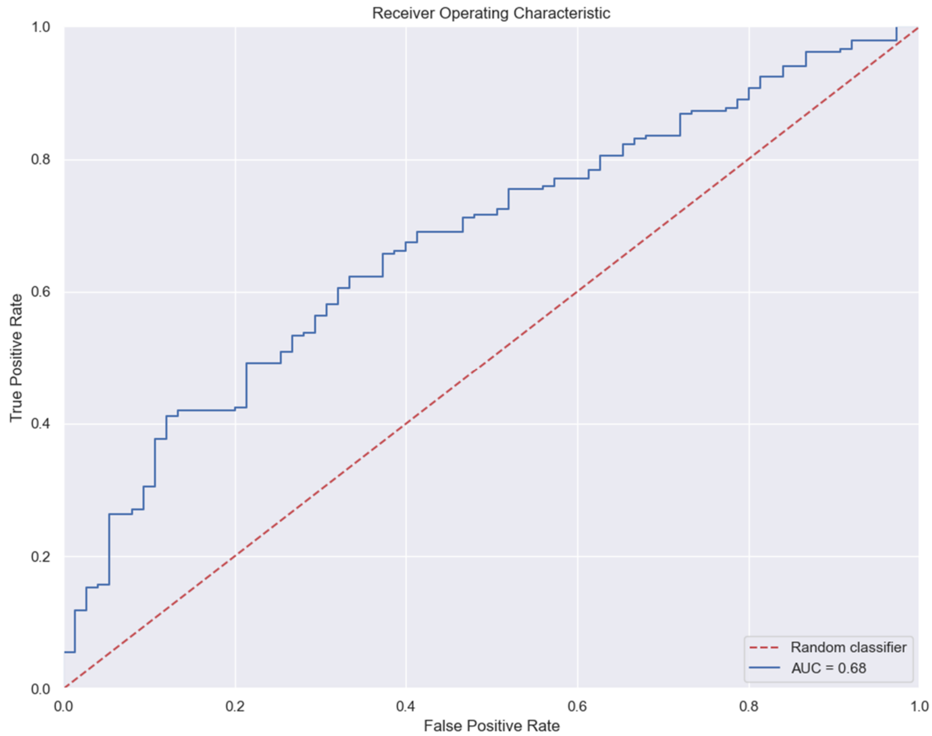
Normalized Confusion Matrix (Test)



This model behavior consistency is again appreciated on the ROC, which again features an AUC = 0.68, shown in Figure 7-41: the model (blue line) still performs better than a random classifier (red line), and while it still struggles to reach a larger AUC, it plateaus less.

Figure 7-41

Test ROC curve

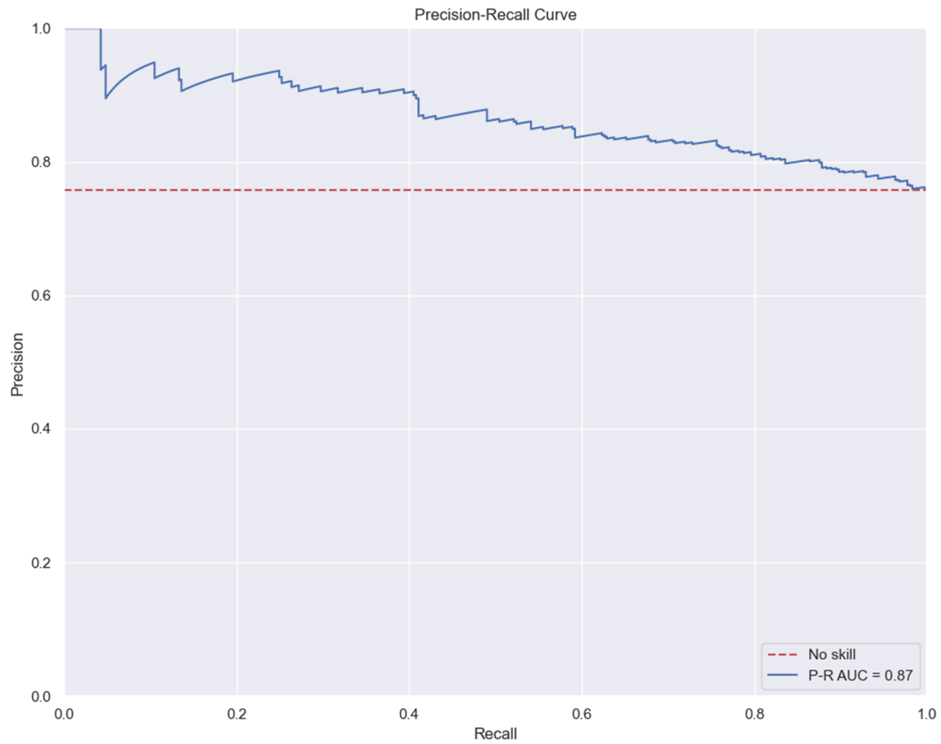


Note: ROC curve (AUC = 0.68) is depicted in blue *vs.* a hypothetical random classifier, in red.

The PRC (depicted in blue in Figure 7-38Figure 7-38) features an AUC = 0.87 just like the PRC issued from the evaluation set. It also constantly always remains above the no-skill threshold deduced from the test set (red line at ~ 0.76).

Figure 7-42

Test PRC Curve



Note: The PRC Curve (AUC = 0.87) is depicted in blue *vs*. a hypothetical no-skill model, in red.

Discussion and conclusion

Experiment 2 proposes one Machine Learning flow-detecting trained model that allows for affordable metrics (*cf.* Chapter 5(ii) above), fast (less than a few milliseconds per participant), of negligeable environmental impact (~0.00000237222222 g of CO_2eq per run), automatic (once properly setup no further intervention is needed by MOOC maintainers), and transparent (no post nor prior intervention demanded to MOOC participants) flow detection in a MOOC context, *i.e.*, online, distant learning settings.

The featured model is the result of several (~50) GridSearchCV tasks, most of them[159] computing assessing about 3 600 distinct Machine Learning trained models, looking for the best fit to the input data according to the selected metrics.

Also, it benefits from being assessed on both an evaluation dataset (like any other Machine Learning model) and a test dataset reserved to simulate a deployment (*a.k.a.* production) phase. Metrics of both assessments are favorably consistent, attesting to the solidity and reliability of the model (no overfit) when facing human-generated, unprocessed, unseen data.

Yet, while this model's metrics on the test dataset seem affordable at first sight and without proper context (F1 = 0.689, AUC ROC = 0.68, Accuracy = 0.605, Recall = 0.578), they become quite attractive (AUC PR = 0.87, Precision = 0.854) when considering the human-sourced nature of its training dataset and its inherited target data imbalance, already covered in Discussion and conclusion of Experiment 1 – Proof-of-Concept.

Certainly, while the Accuracy results of the 10-fold Cross Validation for the evaluation and test datasets (0.572 and 0.589 respectively, plus 0.587 and 0.601 for Balanced accuracy respectively) remain lower compared to those of an ulterior, similar experiment[160] (Accuracy ~ 0.718[161]) (Moon et al., 2022, tbl. 3), our heavily imbalanced datasets force our hand to rely on more than a single metric but a collection thereof. Instead, the behavior of our trained model is better characterized in the Classification reports for imbalanced datasets (Table 7-23 & Table 7-25), the Confusion matrices (Figure 7-36 & Figure 7-40), and the ROC and PR Curves (Figure 7-37, Figure 7-38, and Figure 7-41, Figure 7-42).

Indeed, the resulting flow-detecting Machine Learning model never performs under the threshold of a random classifier (ROC AUC in Figure 7-37 & Figure 7-41) nor of a no-skill

159 A reduced number of incompatible combination of hyperparameters leads to slightly less models being calculated.
160 10-fold Cross Validation results of a Logistic regression classifier employed to predict flow in a video game settings, trained on qualitative (survey) and quantitative data (game logs) (Moon et al., 2022).
161 Assuming the default scorer for a Logistic regression classifier was employed (Accuracy).

classifier (PR AUC in Figure 7-38 & Figure 7-42), neither for the evaluation, nor for the test datasets.

Additionally, ANOVA-based FeatureSelection analysis, illustrated in Figure 7-34, showed feature "number of sessions started on a Monday" consistently outweighed (*k* score = 45.15) all other variables when determining feature importance for flow detection, closely followed by "number of all logged sessions" (*k* score = 39.76), and "number of events" (*k* score = 32.87).

Conclusion & Perspectives

This doctoral dissertation is well established in both the Education and Training Sciences and Computer Sciences by developing a flow-detecting mechanism for a MOOC using Machine Learning techniques. More specifically, this research project precisely fits into the Learning Analytics for Automated feedback function category (Caspari-Sadeghi, 2023, p. 5), to facilitate learners' non-intrusive, ongoing stealth (DiCerbo et al., 2017) affective assessment. Among the many affective and psychological human states, we approached flow because of its positive correlation to numerous learning-favorable metrics.

Steps taken

To provide contributing elements to answer the research question, we reviewed the historically employed flow measurement approaches to elicit insights for flow detection in our own online, distant, educational context.

An extensive literature review on flow taught us that this human psychological state is inherently determined by the individual's own perception of the difficulty of the presented task, and thus, it cannot be characterized as a single, generalized experience. Also, we learnt that flow's fragility entails transparent measuring techniques and multidimensional characterization approaches.

Considering both our research context and flow measurement constraints, we recognized the suitability of Machine Learning techniques to discern flow multidimensionally in a MOOC, in a transparent and automatic fashion, from among many individuals' distinct flow characterizations.

Thus, we approached the research question by first, determining flow measurement instruments adequate to our research needs, second, applying them to as many MOOC learners as possible to collect a variety of flow's individually characterized data, and third, employing this data, along the learners' digital traces from the same MOOC to train a Machine Learning model to recognize flow non-intrusively.

A literature review of flow measurement instruments historically employed in educational contexts contributed to determine the two flow complementary measurement instruments to apply to learners in our chosen research terrain: the MOOC GdP. Likewise, the LogisticRegression classifier proved to be the most adapted Machine Learning technique for our research needs. We validated these choices in a first experiment, with encouraging results.

During a second experiment, we capitalized on the learners' MOOC's digital traces to train the selected classifier to discern flow in a multidimensional space, yielding our proposed Machine Learning flow-detecting trained model.

Contribution

This thesis study provides relevant elements to verify the hypothesis behind the research question – the flow psychological state is somewhat and/or somehow represented and carried out in the digital traces of MOOC learners.

Namely, this study proposes a Machine Learning trained model that detects flow affordably and consistently by exploiting exclusively[162] the MOOC's learners' logs in a transparent and automatic fashion. The pertinence of this proposal is supported by relevant metrics that assess this model's performance when identifying flow in unseen MOOC learners.

It works by simply feeding it 23 aggregated features from learners' digital traces. Furthermore, the detection can be triggered at any given moment during the MOOC duration and given the availability of the input features, detection happens in real time, at a negligeable level of CO_2 emissions.

[162] No academic data was employed as we do not focus on the relationship between flow and (academic) performance (*cf.* Section "A Brief Account").

This result rejoins other studies' attempts in different domains and research contexts to detect flow. Still, when compared to these attempts, our approach benefits of the following:

- It is based on a sound theoretical flow model proper to online, distant, learning contexts, *i.e.*, the EduFlow-2 model.
- It deepens into more complex Machine Learning techniques without sacrificing the model's local interpretation, *i.e.*, the *logit* function, pipelining, pre-processing, feature selection, grid search, cross validation, etc.
- It approaches detection of flow directly, via its compositional dimensions, as opposed to positive-emotion-to-flow, or engagement-to-flow, or deep-concentration-to-flow adjacent mechanisms.
- It detects flow consistently and non-intrusively, without neither the staff nor learners' intervention after it has been set up once, but most remarkably,
- It employed two real (*vs.* synthetic), learner-generated (*vs.* hired and/or voluntary participants), in-context (commercial MOOC *vs.* an *ad hoc* sandboxed MOOC), input datasets as training data, composed of participants' answers (*n*~9 500) to pin-point flow measurement tools, and of their MOOC digital traces (~80GB), spanning all in all a two-year long data collection period.

Thus, our proposed Machine Learning trained model can automatically and transparently account, within a certain degree of certitude, beyond pure luck, for the human psychological state of flow in MOOC participants via their digital traces.

Moreover, in parallel yet not expressed in this doctoral dissertation *per se* but rather found in the various publications, conferences, and presentations cited in the "Published articles" Section, we delivered:

- A method for supporting Scoping Review.
- Theoretical and practical elements to score the EduFlow-2 measurement instrument.

- A proposed Open Learner Model accounting for the flow state.

Limits

This thesis study puts in motion notions from seemingly diverging knowledge fields to approach the research question. Because of this multifield approach, we also had to face and resolve multifield issues, which present limits to acknowledge to our results.

First and foremost, we review the concerns surrounding flow detection, where research is very clear: it is a non-obvious endeavor that requires a multidimensional approach, subject to various criteria affecting the final measuring outcome.

For starters, general questionnaires' ecological validity has long been challenged (Massimini et al., 1988; Willems, 1969; cited in Larson & Csíkszentmihályi, 2014), especially during *a posteriori* data collection (Yarmey, 1979; cited in Larson & Csíkszentmihályi, 2014), with people having difficulties to express "complex dimensions of their own personality or of their experiences" (Larson & Csíkszentmihályi, 2014, p. 22) or simply "not used to putting the contents of their consciousness into words" (Massimini et al., 1988), with cultural elements playing non-negligeable roles (D'Andrade, 1973; cited in Larson & Csíkszentmihályi, 2014).

This is even more pronounced during flow measurement because flow's own immersion leaves no room in one's awareness for introspection, which hinders the state's *a posteriori* reporting (Rheinberg & Engeser, 2018, p. 602), or where there is the "possibility that different subjects interpret flow in different ways, creating measurement error" (Hoffman & Novak, 2009). Recent research seems to indicate this might the case for FlowQ, where students "devoted more cognitive effort to read the flow description, than to read and complete the flow items" (Obadă, 2021, p. 115).

We were aware such a situation might arise but as our own questionnaires were applied in an online, distant setting, we cannot faithfully confirm or deny this. However, we did attempt to reduce the cognitive load when answering our selected flow measurement instruments to the best of our abilities by selecting "short" scales, and by allowing at least two weeks to answer them.

On that same note, we face the issue of reduced data sample frequency, surveying learners twice per MOOC Session, *i.e.*, two data points for a nine-week period, or sometimes only one. This reduced sampling frequency has an immediate consequence on data resolution, and thus on the results' rough granularity. Our flow-detecting model does not specify the moment flow started nor when it ended, as it was not trained with that kind of information because it was not available to us. The model can tell if a learner, based on its historic MOOC log data up to that moment, reunited all the considered characteristics for having flow, as accounted by a similar learner. Further, because we considered only previous-to-the-survey log data, the resulting model does not predict future flow[163].

Likewise, this research study focuses on flow detection, which we approach via the selected flow theoretical model EduFlow-2. Such theoretical model (as many others) does not address the question of what is "on the other side" of flow, *i.e.*, EduFlow-2 does not cover what no-flow is, nor anti-flow, nor flow absence. Therefore, neither do we, nor does our resulting model. To boot, both selected flow measurements instruments envisage flow as an all-or-nothing state, an idea reflected on our own flow-detecting model.

Yet, while we acknowledge the validity of both debates; the question of if the absence of flow is the same as the anti-flow experience (Delle Fave et al., 2011, p. 61), and if flow is to be regarded as a continuum (Peifer & Engeser, 2021b), this research study does not intend to address, nor to settle neither (*cf.* Section "Considerations When Measuring Flow").

Second, concerning the Machine Learning methods employed; we appreciate them as one tool among many in the shed to make sense of massive, multidimensional data. Within the context of the present study, Machine Learning techniques draw ideas from statistics and

[163] Which would imply a flow measurement instrument to ask participants if they plan to be in flow in the future.

Computer Science to "place a premium on empirical results and intuition over the more formal treatments […]" (Mehta et al., 2019, p. 4).

This is not to say that proofs are unnecessary, a meaning explicitly conveyed with our white box approach to Machine Learning modeling: we attempt to keep the resulting model inspectable and explicable. Certainly, while we strive for a Level 3 XAI (Sanneman, 2023, p. 54), our resulting Machine Learning model can be classified as a Level 1 XAI, as it allows for complementary implementations of explanations of what the AI system did or is doing. Furthermore, we acknowledge and concede that, in Machine Learning, just like in the field of deep learning[164], "many of the advances of the last two decades [in … deep learning] do not have formal justifications" (Mehta et al., 2019, p. 5).

Furthermore, the Machine Learning model produced in this research work does not pretend to be an exhaustive model that explains (or predicts) the phenomenon in question. It works – and makes sense – within a particular set of conditions and assumptions. It helps to simplify the understanding of the reality being studied but it does not claim to represent the wholeness (nor holiness) of such reality.

Just like any other Machine Learning model, our resulting model exploits correlations between the target and the provided features, which are our most educated guesses on what specific data available to us might carry information on the flow phenomenon. This recalls one of most basic adages in statistics: "correlation is not causation".

This is because the event we are trying to detect cannot be directly observed through any of the correlated variables we feed into the model, *i.e.*, none of the input variables fully captures the essence of the phenomenon, and thus, no choice is left but to attempt to capture such essence through "proxies" (Leetaru, 2019), which might seem at first glance to be unrelated variables.

[164] Deep learning is another subset of Artificial Intelligence.

Likewise, issues in Machine Learning models can be varied, ranging from traditional bugs in the algorithm's implementation to a poor choice of hyperparameters, without forgetting data's quality (Isbell et al., 2023, p. 37). Under this light, the failure or success of any Machine Learning model depends so heavily on the conditions in which it was trained, that they might instead become entirely determinant on the model's performance. One such example is the photography context surrounding the numerous images employed for identifying skin cancer in a cancer study listed in "The Alignment Problem" (Christian, 2020): an example of a resulting Machine Learning model showed to be "[…] much more likely to classify any image with a ruler in it as cancerous" (Christian, 2020, p. 133; Perota et al., 2023, sc. 18'21") because "[…] medical images of malignancies are much more likely to contain a ruler for scale than images of healthy skin", *i.e.*, the presence of a ruler carried more weight than the features of the depicted lesion: the model had instead learnt that measuring rulers determined malignant skin cancer.

Indeed, even if any of the features is mathematically more prone to determine the target than others, it does not imply that such variable is indeed related to, or explains the phenomenon we are attempting to detect: "A pile of dog photographs cannot build a model to recognize their barks" (Leetaru, 2019), or "A Roomba[165] that was instructed not to bump into furniture learned to drive backward [*sic*], because there were no bumper sensors on the rear" (Isbell et al., 2023, p. 36).

That is why it is important to acknowledge that training conditions (which we exhaustively tried to describe) matter at least as much as the variables' design and construction. Understanding the conditions will help understand for how long the model might still be a valid model and its present and future relevance (Leetaru, 2019).

A major factor affecting training is the quality of the input dataset. Indeed, our most valued assets are the two real, learner-generated, in-context, datasets. Because of their human

[165] "Roomba is a series of autonomous robotic vacuum cleaners made by the company iRobot" ("Roomba," 2023).

origin, particular care was taken to ensure that no flawed data was being passed on to the next processing Machine Learning stage, even if it meant drastically reducing its amount. A strict filtering took place: any participant answering incorrectly any check questions was removed from the sample, without verifying the validity, coherence, or logical sequencing of their subsequent answers.

Despite these strict efforts, it is impossible to negate human bias in human-generated data. Our resulting Machine Learning flow-detecting trained model detects flow better than it does no-flow, presumably due to human bias two ways:

1. The intricate writing style employed in Flow-Q (an issue already covered above) might have contributed to cognitive load when answering the psychometric tests, in turn leading to self-identification effects, conducive to measurement error, or to simply leaving the item empty-answered, which would warrant elimination from the data sample.

2. Following "normal" MOOC dropout tendencies, participants more committed to the MOOC completion would answer questionnaires more accurately, hence a larger proportion of "committed" respondents, more likely to have experienced flow themselves, end up in a major representation in the final sample, when compared to those who a) did not experience flow and dropped the MOOC before reporting their lack of flow; or b) did not experience flow and thus, chose not to answer the surveys and end up not reporting their lack of flow.

Regrettably, we had no way to faithfully measure the impact of assumed human bias in our data.

The unforeseen consequences of human bias -along with heterogenous survey nomenclatures and the construction of flow indicators from digital traces- induced an exceptionally long and arduous but meticulous data-cleaning stage. This step took considerably more lines of code and time than the actual Machine Learning training stage, which could be first perceived as the heart of this research project. That is why this research project firmly and irrevocably insists on dedicating more than enough time and resources to the data cleaning stage: it is imperative to precisely stipulate the verifications to perform, the prevailing rules on field naming, on data values ranges, and data types.

Accordingly, just like in any other Machine Learning project, larger, well-structured input datasets[166] might provide additional information or a different insight into the phenomenon being detected. A clear example being that a different characterization of the flow state by researchers designing the flow measurement instruments, and/or by researchers translating these into Machine Learning features and transformations for the model in training, might deeply impact the metrics of the final model, as evidenced in Machine Learning for MOOC studies (Dalipi et al., 2018, sec. IV). Furthermore, it is of paramount importance to recall that Machine Learning predictions in this research project are performed on people's data and not solely on data, as an abstract notion (Chancellor et al., 2019, p. 87).

On that same note, we must reassure the reader of our strict adherence to all ethical guidelines available to us on the subject at hand, and yet to remind the reader that inherently, ethics always eludes computation because such social concepts "are not fixed or determinate in their precise meaning. To be applied they must be interpreted, and interpretations vary among individuals and groups, from context to context, and may change over time" (D. G. Johnson & Verdicchio, 2023, p. 33), *i.e.*, ethics cannot be added to Artificial Intelligence because ethics cannot be expressed nor understood in a computational manner.

Finally, it is impossible to regard this research work without acknowledging the world-changing health context in which it was performed and how it affected the educational aspect treated in this thesis. Indeed, the outbreak of the COVID-19 pandemic affected entire education systems worldwide when suddenly learners of all ages (1.5 billion in high education only) were made to attend online courses, like MOOCs (Kichu & Bhattacharya, 2021). Regarding this

[166] *Cf.* the "Data Bottleneck" (Shani et al., 2023).

research project, suffice it to say that such disruption greatly affected the delivery of the MOOC learners' digital traces, while causing psychological distress to those involved.

Perspectives

This research project aims to provide elements to answer the research question, materialized in a flow-detecting Machine Learning model.

Considering its satisfactory results, its primary short-term intended usage is deploying it as an independent API for the MOOC to make calls to. Its results would be sent back to the MOOC to be displayed on the trainers' dashboard. Knowing the learners' flow state would provide the pedagogical staff an insight into their psychological state, determinant to MOOC performance. Moreover, it would allow for informed, pedagogical decision-making, such as focusing on individual learners' needs, or reviewing potentially challenging materials.

In the long-term, it would be desirable to evaluate the resulting model's incidence on MOOC abandonment rates. This would require mobilizing not only its short-term implantation has a dashboard element previously described, but also a larger study comparing three scenarios on MOOC abandonment rates:

1. Without employing the model, considering the differences in seasonality and other demographics as a baseline.

2. Employing it in a dashboard for the pedagogical staff to manually consider its results, alongside other metrics, as they see fit (descriptive analytics, (Gartner, 2014)), and

3. Implementing it in the MOOC's Learner Model (Bodily et al., 2018; Corbett et al., 1995) for full personalization (M. Chen et al., 2021; Clerc et al., 2015; El Mawas et al., 2019; El Mawas, Ghergulescu, et al., 2018; Lefèvre et al., 2016; Sunar et al., 2015; Turner & Patrick, 2008).

Besides, as we mentioned in the Limits Section above, the resulting model could always be improved by re-training it under a different, enhanced approach.

Possible developments include first to consider the individual, elusive, unstable nature of the flow phenomena and attempt Time Series forecasting (Bhatnagar et al., 2021) to refine the granularity of flow detection with limited data points in time. This Machine Learning technique would attempt to compensate for our limited sampling on the time axis.

Further, increased resources to secure an increase in the number of participants and on the quality of their answers could go a long way during the training phase, although this would imply a compromise in the realness of the input dataset.

Thus, we conclude this research project in the hopes the resources and time invested in its realization contributed to cast a light on the specific matter at hand: flow detection in MOOC.

This project was supported by the French government through the Programme Investissement d'Avenir (I-SITE ULNE / ANR-16-IDEX-0004 ULNE) managed by the Agence Nationale de la Recherche.

References

Abbassi, M., Abbassi, W., Fenouillet, F., & Naceur, A. (2021). Validation of the Flow Scale Related to Physical Education in Arabic Language. *Advances in Physical Education*, *11*(02), 246–260. https://doi.org/10.4236/ape.2021.112020

Abuhamdeh, S. (2021a). Flow Theory and Cognitive Evaluation Theory: Two Sides of the Same Coin? In C. Peifer & S. Engeser (Eds.), *Advances in Flow Research* (pp. 137–153). Springer International Publishing. https://doi.org/10.1007/978-3-030-53468-4_5

Abuhamdeh, S. (2021b). On the Relationship Between Flow and Enjoyment. In C. Peifer & S. Engeser (Eds.), *Advances in Flow Research* (pp. 155–169). Springer International Publishing. https://doi.org/10.1007/978-3-030-53468-4_6

Abyaa, A., Khalidi Idrissi, M., & Bennani, S. (2019). Learner modelling: Systematic review of the literature from the last 5 years. *Educational Technology Research and Development*, *67*(5), 1105–1143. https://doi.org/10.1007/s11423-018-09644-1

Adams, D. (2005). *The hitchhiker's guide to the galaxy*. Ballantine Books.

Adlai-Gail, W. S. (1995). *Exploring the autotelic personality.* [Doctoral dissertation]. The University of Chicago.

Admiraal, W., Huisman, B., & Pilli, O. (2015). Assessment in massive open online courses. *Electronic Journal of E-Learning*, *13*(4), pp207-216.

Aix Marseille Université, F. des sciences médicales et paramédicales. (2023). *Habilitation to Conduct Research (HDR)* [University Website]. Faculty of Medical and Paramedical Sciences. https://smpm.univ-amu.fr/en/education/hdr

Allison, M. T., & Duncan, M. C. (1987). Women, work, and leisure: The days of our lives. *Leisure Sciences*, *9*(3), 143–161.

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software engineering for machine learning: A case study. *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, 291–300. https://doi.org/10.1109/ICSE-SEIP.2019.00042

Amruta, A., & Naik Ramgir, V. (2021). Adoption of Open Learning Systems and MOOCS during COVID-19 by Academic Libraries. *International Journal of Library and Information Studies 2021*, *11*(1), 56–64. https://papers.ssrn.com/abstract=3860766

Anderson, P., Backhouse, G., Curtis, D., Redding, S., & Wallom, D. (2009). Low Carbon Computing: A view to 2050 and beyond. *JISC Technology & Standards Watch (TechWatch).*

Arafat, Y., Chowdhury, H. R., Qusar, S., & Hafez, M. A. (2016). Cross-cultural adaptation and psychometric validation of research instruments: A methodological review. *Journal of Behavioral Health*, *5*(3), 129–136. https://doi.org/10.5455/jbh.20160615121755

Arora, A. K., & Srinivasan, R. (2020). Impact of Pandemic COVID-19 on the Teaching – Learning Process: A Study of Higher Education Teachers. *Prabandhan: Indian Journal of Management*, *13*(4), Article 4. https://doi.org/10.17010/pijom/2020/v13i4/151825

Asakawa, K. (2010). Flow experience, culture, and well-being: How do autotelic Japanese college students feel, behave, and think in their daily lives? *Journal of Happiness Studies*, *11*(2), 205–223. https://doi.org/10.1007/s10902-008-9132-3

Atkinson, J. W. (1958). *Motives in fantasy, action, and society: A method of assessment and study.*

Avenier, M.-J., & Thomas, C. (2015). Finding one's way around various methodological guidelines for doing rigorous case studies: A comparison of four epistemological frameworks. *Systèmes d'information et Management*, *20*(1), 61–98. https://doi.org/10.3917/sim.151.0061

Baak, M., Koopman, R., Snoek, H., & Klous, S. (2019). *A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics* (arXiv:1811.11440). arXiv. http://arxiv.org/abs/1811.11440

Bachelet, R. (2019). *LE MOOC GdP: Chiffres presse* [MOOC]. MOOC Gestion de Projet. https://gestiondeprojet.pm/mooc-gdp/#chiffres-presse

Bachen, C. M., Hernández-Ramos, P., Raphael, C., & Waldron, A. (2016). How do presence, flow, and character identification affect players' empathy and interest in learning from a serious computer game? *Computers in Human Behavior*, *64*, 77–87. https://doi.org/10.1016/j.chb.2016.06.043

Bakker, A. B. (2005). Flow among music teachers and their students: The crossover of peak experiences. *Journal of Vocational Behavior*, *66*(1), 26–44. https://doi.org/10.1016/j.jvb.2003.11.001

Bakker, A. B. (2008). The work-related flow inventory: Construction and initial validation of the WOLF. *Journal of Vocational Behavior*, *72*(3), 400–414. https://doi.org/10.1016/j.jvb.2007.11.007

Bakker, A. B., & Demerouti, E. (2008). Towards a model of work engagement. *Career Development International*, *13*(3), 209–223. https://doi.org/10.1108/13620430810870476

Bakker, A. B., & Leiter, M. P. (Eds.). (2010). *Work engagement: A handbook of essential theory and research*. Psychology Press.

Bakker, A. B., Ljubin Golub, T., & Rijavec, M. (2017). Validation of the Study-Related Flow Inventory (WOLF-S) / Validacija Inventara zanesenosti u studiranju (WOLF-S). *Croatian Journal of Education - Hrvatski Časopis Za Odgoj i Obrazovanje*, *19*(1). https://doi.org/10.15516/cje.v19i1.2194

Balachef, N. (2018). Les mots de la recherche sur les EIAH, enjeux et questions. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*, *25*(18), 63–95. https://doi.org/10.23709/STICEF.25.2.2

Bali, M. (2014). MOOC pedagogy: Gleaning good practice from existing MOOCs. *Journal of Online Learning and Teaching*, *10*(1), 44.

Bandura, A. (2001). Social Cognitive Theory: An Agentic Perspective. *Annual Review of Psychology*, *52*(1), 1–26. https://doi.org/10.1146/annurev.psych.52.1.1

Bandura, A., Freeman, W. H., & Lightsey, R. (1997). Self-Efficacy: The Exercise of Control. *Journal of Cognitive Psychotherapy*, *13*(2), 158–166. https://doi.org/10.1891/0889-8391.13.2.158

Bassi, M., & Fave, A. D. (2012). Optimal Experience Among Teachers: New Insights Into the Work Paradox. *The Journal of Psychology*, *146*(5), 533–557. https://doi.org/10.1080/00223980.2012.656156

Bassi, M., Steca, P., Fave, A. D., & Caprara, G. V. (2007). Academic Self-Efficacy Beliefs and Quality of Experience in Learning. *Journal of Youth and Adolescence*, *36*(3), 301–312. https://doi.org/10.1007/s10964-006-9069-y

Bassi, M., Steca, P., Monzani, D., Greco, A., & Delle Fave, A. (2014). Personality and optimal experience in adolescence: Implications for well-being and development. *Journal of Happiness Studies*, *15*(4), 829–843. https://doi.org/10.1007/s10902-013-9451-x

Bates, S., Hastie, T., & Tibshirani, R. (2022). *Cross-validation: What does it estimate and how well does it do it?* (arXiv:2104.00673). arXiv. https://doi.org/10.48550/arXiv.2104.00673

Baumann, N. (2012). Autotelic Personality. In S. Engeser (Ed.), *Advances in Flow Research* (pp. 165–186). Springer New York. https://doi.org/10.1007/978-1-4614-2359-1_9

Beel, J., & Gipp, B. (2010). Academic Search Engine Spam and Google Scholar's Resilience Against it. *The Journal of Electronic Publishing*, *13*(3). https://doi.org/10.3998/3336451.0013.305

Benjafield, H., & Moneta, G. B. (2023). Outline of a Model of Progressive Flow Absorption. *Archives in Neurology & Neuroscience*, *15*(4), 5. https://doi.org/10.33552/ANN.2023.14.000867

Bernard, R. (2009). Music Making, Transcendence, Flow, and Music Education. *International Journal of Education & the Arts*, *10*(14), 21. http://www.ijea.org/v10n14/

Berrar, D. (2019). Cross-Validation. In S. Ranganathan, M. Gribskov, K. Nakai, & C. Schönbach (Eds.), *Encyclopedia of Bioinformatics and Computational Biology* (pp. 542–545). Academic Press. https://doi.org/10.1016/B978-0-12-809633-8.20349-X

Bertereau, C., Marbot, E., & Chaudat, P. (2019). Positionnement épistémologique et orientation de la recherche: Un focus sur l'étude des stéréotypes: *RIMHE : Revue Interdisciplinaire Management, Homme & Entreprise*, *n° 34, 8*(1), 51–66. https://doi.org/10.3917/rimhe.034.0051

Bhatnagar, A., Kassianik, P., Liu, C., Lan, T., Yang, W., Cassius, R., Sahoo, D., Arpit, D., Subramanian, S., Woo, G., Saha, A., Jagota, A. K., Gopalakrishnan, G., Singh, M., Krithika, K. C., Maddineni, S., Cho, D., Zong, B., Zhou, Y., … Wang, H. (2021). Merlion: A Machine Learning Library for Time Series. *arXiv:2109.09265 [Cs, Stat]*. http://arxiv.org/abs/2109.09265

Biclar, W. J., Rabor, J., Asok, G., & Arcilla, F. (2019). "In The Zone": Lived Experiences of MVC Music Instrumentalists. *SMCC Higher Education Research Journal*, *6*(1). https://doi.org/10.18868/sherj6j.06.010119.04

Bisong, E. (2019a). *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*. Apress. https://doi.org/10.1007/978-1-4842-4470-8

Bisong, E. (2019b). Google Colaboratory. In E. Bisong (Ed.), *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners* (pp. 59–64). Apress. https://doi.org/10.1007/978-1-4842-4470-8_7

Blackmon, S., & Major, C. (2017). Wherefore art thou MOOC: Defining Massive Open Online Courses. *Online Learning Journal*, *21*(4), 195–221. https://www.learntechlib.org/p/183776/article_183776.pdf

Bloom, A. J., & Skutnick-Henley, P. (2005). Facilitating flow experiences among musicians. *The American Music Teacher*, *54*(5), 24.

Bodiam, M. (2017). Going with the flow: Autism and 'flow states'. In *Enhancing Lives – reducing restrictive practice*. kar.kent.ac.uk. https://kar.kent.ac.uk/63699/

Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., & Verbert, K. (2018). Open learner models and learning analytics dashboards: A systematic review. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 41–50. https://doi.org/10.1145/3170358.3170409

Boffi, M. (2012). *Flow as a measure of political engagement*. 6th European Conference on Positive Psychology, Moscow, Russia. https://hdl.handle.net/2434/231497

Booch, G., Rumbaugh, J., & Jacobson, I. (2005). *The unified modeling language user guide* (2nd ed). Addison-Wesley.

Breiman, L., & Spector, P. (1992). *Submodel selection and evaluation in regression. The X-random case* (197; pp. 291–319). University of California. https://digitalassets.lib.berkeley.edu/sdtr/ucb/text/197.pdf

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying Learning in the Worldwide Classroom Research into edX's First MOOC. *Research & Practice in Assessment*, *8*, 13–25. https://eric.ed.gov/?id=ej1062850

Brown, C. D., & Davis, H. T. (2006). Receiver operating characteristics curves and related decision measures: A tutorial. *Chemometrics and Intelligent Laboratory Systems*, *80*(1), 24–38. https://doi.org/10.1016/j.chemolab.2005.05.004

Brownlee, J. (2020a, August 14). A Tour of Machine Learning Algorithms. *Machine Learning Mastery*. https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

Brownlee, J. (2020b, August 14). What is the Difference Between Test and Validation Datasets? *MachineLearningMastery.Com*. https://machinelearningmastery.com/difference-test-validation-datasets/

Brownlee, J. (2020c, November 10). How to Identify Overfitting Machine Learning Models in Scikit-Learn. *MachineLearningMastery.Com*. https://machinelearningmastery.com/overfitting-machine-learning-models/

Brownlee, J. (2021, April 8). What Is Semi-Supervised Learning. *MachineLearningMastery.Com*. https://machinelearningmastery.com/what-is-semi-supervised-learning/

Budennyy, S., Lazarev, V., Zakharenko, N., Korovin, A., Plosskaya, O., Dimitrov, D., Akhripkin, V., Pavlov, I., Oseledets, I., Barsola, I., & others. (2023). Eco2ai: Carbon emissions tracking of machine learning models as the first step towards sustainable ai. *Doklady Mathematics*, 1–11.

Bulfin, S., Pangrazio, L., & Selwyn, N. (2014). Making 'MOOCs': The Construction of a New Digital Higher Education within News Media Discourse. *International Review of Research in Open and Distributed Learning*, *15*(5), 290–305. https://doi.org/10.19173/irrodl.v15i5.1856

Bull, S., & Kay, J. (2010). Open Learner Models. In R. Nkambou, J. Bourdeau, & R. Mizoguchi (Eds.), *Advances in Intelligent Tutoring Systems* (pp. 301–322). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-14363-2_15

Burt, I., & Gonzalez, T. (2021). Flow State as an Existential Tool to Increase Optimal Experience and Life Enjoyment. *The Journal of Humanistic Counseling*, *60*(3), 197–214. https://doi.org/10.1002/johc.12165

Butkovic, A., Ullén, F., & Mosing, M. A. (2015). Personality related traits as predictors of music practice: Underlying environmental and genetic influences. *Personality and Individual Differences*, *74*, 133–138. https://doi.org/10.1016/j.paid.2014.10.006

Carroll, W. D., Strenger, V., Eber, E., Porcaro, F., Cutrera, R., Fitzgerald, D. A., & Balfour-Lynn, I. M. (2020). European and United Kingdom COVID-19 pandemic experience: The same but different. *Paediatric Respiratory Reviews*, *35*, 50–56. https://doi.org/10.1016/j.prrv.2020.06.012

Carver, L., & Harrison, L. M. (2013). MOOCs and Democratic Education. *Liberal Education*, *99*(4), 20–25. https://eric.ed.gov/?id=EJ1094825

Caspari-Sadeghi, S. (2023). Learning assessment in the age of big data: Learning analytics in higher education. *Cogent Education*, *10*(1), 11. https://doi.org/10.1080/2331186X.2022.2162697

Cawley, G. C., & Talbot, N. L. C. (2010). On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *Journal of Machine Learning Research*, *11*, 29.

Ceja, L., & Navarro, J. (2009). Dynamics of flow: A nonlinear perspective. *Journal of Happiness Studies*, *10*(6), 665–684. https://doi.org/10.1007/s10902-008-9113-6

Ceja, L., & Navarro, J. (2011). Dynamic patterns of flow in the workplace: Characterizing within-individual variability using a complexity science approach. *Journal of Organizational Behavior*. https://onlinelibrary.wiley.com/doi/abs/10.1002/job.747

Ceja, L., & Navarro, J. (2012). 'Suddenly I get into the zone': Examining discontinuities and nonlinear changes in flow experiences at work. *Human Relations*, *65*(9), 1101–1127. https://doi.org/10.1177/0018726712447116

Center for Digital Education (CEDE), & École Polytechnique Fédérale de Lausanne (EPFL). (2016). *MOOCs Annual Report 2015* (p. 68). Center for Digital Education.

Centor, R. M. (1991). Signal Detectability: The Use of ROC Curves and Their Analyses. *Medical Decision Making*, *11*(2), 102–106. https://doi.org/10.1177/0272989X9101100205

Chakrabarti, S., Ester, M., Fayyad, U., Gehrke, J., Han, J., Morishita, S., Piatetsky-Shapiro, G., & Wang, W. (2006). *Data Mining Curriculum: A Proposal (version 1.0)*. ACM SIGKDD. https://www.kdd.org/curriculum/index.html

Chalghaf, N., Azaiez, C., Krakdiya, H., Guelmami, N., Re, T. S., Maldonado Briegas, J. J., Zerbetto, R., Del Puente, G., Garbarino, S., Bragazzi, N. L., & Azaiez, F. (2019). Trans-Cultural Validation of the "Academic Flow Scale" (Flow 4D 16) in Arabic Language: Insights for Occupational and Educational Psychology From an Exploratory Study. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02330

Chancellor, S. (2023). Toward Practices for Human-Centered Machine Learning. *Communications of the ACM*, *66*(3), 78–85. https://doi.org/10.1145/3530987

Chancellor, S., Birnbaum, M. L., Caine, E. D., Silenzio, V. M. B., & De Choudhury, M. (2019). A Taxonomy of Ethical Tensions in Inferring Mental Health States from Social Media. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 79–88. https://doi.org/10.1145/3287560.3287587

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, G.-S., & Lee, M.-F. (2012). Detecting emotion model in e-learning system. *2012 International Conference on Machine Learning and Cybernetics*, *5*, 1686–1691. https://doi.org/10.1109/ICMLC.2012.6359628

Chen, J. (2006). *Flow in Games* [Master dissertation, University of Southern California]. http://www.jenovachen.com/flowingames/Flow_in_games_final.pdf

Chen, L.-X., & Sun, C.-T. (2016). Self-regulation influence on game play flow state. *Computers in Human Behavior*, *54*, 341–350.

Chen, M., Wang, X., Wang, J., Zuo, C., Tian, J., & Cui, Y. (2021). Factors Affecting College Students' Continuous Intention to Use Online Course Platform. *SN Computer Science*, *2*(2), 114. https://doi.org/10.1007/s42979-021-00498-8

Chen, T., Peng, L., Yin, X., Rong, J., Yang, J., & Cong, G. (2020). Analysis of User Satisfaction with Online Education Platforms in China during the COVID-19 Pandemic. *Healthcare*, *8*(3), Article 3. https://doi.org/10.3390/healthcare8030200

Cherigny, F., El Kechai, H., Iksal, S., Lefevre, M., Labarthe, H., & Luengo, V. (2020). *L'analytique des apprentissages avec le numérique* (Groupes thématiques de la Direction du numérique pour l'Éducation (DNE -TN2)) [Report]. Direction du numérique pour l'éducation. https://hal.science/hal-02912386

Chermann, E. (2020, March 1). Enseignement en ligne: Les 1001 secrets d'un MOOC qui cartonne. *Le Monde*. https://www.lemonde.fr/economie/article/2020/03/01/enseignement-en-ligne-les-1001-secrets-d-un-mooc-qui-cartonne_6031425_3234.html

Cheron, G. (2016). How to Measure the Psychological "Flow"? A Neuroscience Perspective. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01823

Chien, A. A. (2023). GenAI: Giga$$$, TeraWatt-Hours, and GigaTons of CO2. *Communications of the ACM*, *66*(8), 5–5. https://doi.org/10.1145/3606254

Choi, D., & Kim, J. (2004). Why People Continue to Play Online Games: In Search of Critical Design Factors to Increase Customer Loyalty to Online Contents. *Cyberpsychology and Behavior*, *7*(1), 11–24. https://doi.org/10.1089/109493104322820066

Choi, D.-S., Kim, H.-Y., & Kim, J.-W. (2000). A cognitive and emotional strategy for computer game design. *Asia Pacific Journal of Information Systems*, *10*(1), 165–187.

Choquet, C. (2005). *DPULS Project-Design Patterns for recording and analysing Usage of Learning Systems* [Technical report]. Action of The European Network of Excellence Kaleidoscope.

Choquet, C., & Iksal, S. (2007). Modélisation et construction de traces d'utilisation d'une activité d'apprentissage: Une approche langage pour la réingénierie d'un EIAH. *Sciences et Technologies de l'Information et de La Communication Pour l'Éducation et La Formation*, *14*, 24. http://sticef.univ-lemans.fr/num/vol2007/14-choquet/sticef_2007_choquet_14.htm

Christian, B. (2020). *The alignment problem: Machine learning and human values* (First edition). W.W. Norton & Company. https://studylib.net/doc/25834546/alignment-problem

Chuang, I., & Ho, A. (2016). *HarvardX and MITx: Four Years of Open Online Courses -- Fall 2012-Summer 2016* (SSRN Scholarly Paper 2889436). https://doi.org/10.2139/ssrn.2889436

Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, *16*(1), 64–73. http://www.jstor.org/stable/3150876

Cisel, M. (2016). *Utilisations des MOOC: Éléments de typologie* [Doctoral dissertation, Université Paris-Saclay]. https://theses.hal.science/tel-01444125

Cisel, M. (2017). Une analyse de l'utilisation des vidéos pédagogiques des MOOC par les non-certifiés. *Sciences et Technologies de l'Information et de La Communication Pour l'Éducation et La Formation*, *24*(2). https://doi.org/10.23709/sticef.24.2.7

Clerc, F., Lefevre, M., Guin, N., & Marty, J.-C. (2015). Mise en place de la personnalisation dans le cadre des MOOCs. *7ème Conférence Sur Les Environnements Informatiques Pour l'Apprentissage Humain - EIAH'2015*, 291–300. https://hal.archives-ouvertes.fr/hal-01177839

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 134–138. https://doi.org/10.1145/2330601.2330636

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, *18*(6), 683–695. https://doi.org/10.1080/13562517.2013.827653

Conati, C., Porayska-Pomsta, K., & Mavrikis, M. (2018). AI in Education needs interpretable machine learning: Lessons from Open Learner Modelling. *arXiv:1807.00154 [Cs]*. http://arxiv.org/abs/1807.00154

Corbett, A., Anderson, J., & O'Brien, A. (1995). Chapter 2 -Student modeling in the ACT programming tutor. In P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively diagnostic assessment*. Lawrence Erlbaum Associates: Hillsdale, NJ.

Cormier, D. (2010, December 20). MOOCs, knowledge and the digital economy – a research project [Educational Blog]. *Dave's Educational Blog*. http://davecormier.com/edblog/2010/12/20/moocs-knowledge-and-the-digital-economy-a-research-project/

Creed-Dikeogu, G., & Clark, C. (2013). Are you MOOC-ing yet? A review for academic libraries. *Kansas Library Association College and University Libraries Section Proceedings*, *3*(1), 9–13. https://doi.org/10.4148/culs.v1i0.1830

Csíkszentmihályi, M. R. (1975a). *Beyond Boredom and Anxiety: The Experience of Play in Work and Games* (First Edition). San Francisco: Jossey Press. https://openlibrary.org/books/OL4879227M/Beyond_boredom_and_anxiety

Csíkszentmihályi, M. R. (1975b). Play and Intrinsic Rewards. *Journal of Humanistic Psychology*, *15*(3), 41–63. https://doi.org/10.1177/002216787501500306

Csíkszentmihályi, M. R. (1990a). *Flow: The Psychology of Optimal Experience*. Harper Perennial.

Csíkszentmihályi, M. R. (1990b). Literacy and Intrinsic Motivation. *Daedalus*, *119*(2), 115–140. https://www.jstor.org/stable/20025303

Csíkszentmihályi, M. R. (1993). *The evolving self: A psychology for the third millennium* (Vol. 5). HarperCollins Publishers New York.

Csíkszentmihályi, M. R. (1997). *Finding Flow—The Psychology of Engagement with Everyday Life*. Basic Books.

Csíkszentmihályi, M. R. (2014). *Flow and the foundations of positive psychology*. Springer. https://doi.org/10.1007/978-94-017-9088-8_14

Csíkszentmihályi, M. R., Abuhamdeh, S., & Nakamura, J. (2005). Flow. In A. . J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 598–608). The Guilford Press.

Csíkszentmihályi, M. R., & Csíkszentmihályi, I. S. (Eds.). (1988). *Optimal Experience: Psychological Studies of Flow in Consciousness* (First). Cambridge University Press.

Csíkszentmihályi, M. R., & Larson, R. (1984). *Being adolescent: Conflict and growth in the teenage years*. Basic Books.

Csíkszentmihályi, M. R., & Larson, R. (2014). Validity and Reliability of the Experience-Sampling Method. In M. R. Csíkszentmihályi (Ed.), *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi* (pp. 35–54). Springer Netherlands. https://doi.org/10.1007/978-94-017-9088-8_3

Csíkszentmihályi, M. R., & LeFèvre, J. (1989). Optimal experience in work and leisure. *Journal of Personality and Social Psychology*, *56*(5), 815. https://doi.org/10.1037/0022-3514.56.5.815

Csíkszentmihályi, M. R., & Rathunde, K. (1992). The Measurement of Flow in Everyday Life: Toward a Theory of Emergent Motivation. *Nebraska Symposium on Motivation. Nebraska Symposium on Motivation*, *40*, 57–97.

Culbertson, S. S., Fullagar, C. J., Simmons, M. J., & Zhu, M. (2015). Contagious Flow: Antecedents and Consequences of Optimal Experience in the Classroom. *Journal of Management Education*, *39*(3), 319–349. https://doi.org/10.1177/1052562914545336

Dalipi, F., Imran, A. S., & Kastrati, Z. (2018). MOOC Dropout Prediction Using Machine Learning Techniques: Review and Research Challenges. *2018 IEEE Global Engineering Education Conference (EDUCON)*, 1007–1014. https://doi.org/10.1109/EDUCON.2018.8363340

D'Andrade, R. G. (1973). Cultural constructions of reality. *Cultural Illness and Health*, 115–127.

Dangeti, P. (2017). *Statistics for machine learning*. Packt Publishing Ltd. https://univ-scholarvox-com.ressources-electroniques.univ-lille.fr/book/88842929

Daniel, S. J. (2020). Education and the COVID-19 pandemic. *Prospects*, *49*(1), 91–96. https://doi.org/10.1007/s11125-020-09464-3

Das, K., & Behera, R. N. (2017). A Survey on Machine Learning: Concept, Algorithms and Applications. *International Journal of Innovative Research in Computer and Communication Engineering*, *5*(2). https://doi.org/10.15680/IJIRCCE.2017. 0502001

Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning*, 233–240. https://doi.org/10.1145/1143844.1143874

de Moura Jr, P. J., & Porto Bellini, C. G. (2019). The measurement of flow and social flow at work: A 30-year systematic review of the literature. *Personnel Review*, *49*(2), 537–570. https://doi.org/10.1108/PR-07-2018-0240

de Souza Nascimento, E., Ahmed, I., Oliveira, E., Palheta, M. P., Steinmacher, I., & Conte, T. (2019). Understanding development process of machine learning systems: Challenges and solutions. *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 1–6. https://doi.org/10.1109/ESEM.2019.8870157

Deci, E. L. (1975). *Intrinsic Motivation* (First). Plenum Press. https://doi.org/10.1007/978-1-4613-4446-9

Deci, E. L., & Ryan, R. M. (1980). The empirical exploration of intrinsic motivational processes. In *Advances in experimental social psychology* (Vol. 13, pp. 39–80). Elsevier.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.

Deci, E. L., & Ryan, R. M. (2002). *Handbook of self-determination research*. University Rochester Press.

Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). *Mathematics for Machine Learning*. Cambridge University Press.

Delle Fave, A., & Massimini, F. (2003). Optimal experience in work and leisure among teachers and physicians: Individual and bio-cultural implications. *Leisure Studies*, *22*(4), 323–342. https://doi.org/10.1080/02614360310001594122

Delle Fave, A., Massimini, F., & Bassi, M. (2011). *Psychological Selection and Optimal Experience Across Cultures* (Vol. 2). Springer Netherlands. https://doi.org/10.1007/978-90-481-9876-4

Delozanne, E., Labat, J.-M., Le Calvez, F., & Merceron, A. (2005). *DPULS-D32. 6.1, a Structured Set of Design Patterns for the Usage Analysis* [Technical report]. Action of The European Network of Excellence Kaleidoscope.

Delozanne, E., Le Calvez, F., Merceron, A., & Labat, J.-M. (2007). Design Patterns en EIAH: Vers un langage de Patterns pour l'évaluation des apprenants. *Sciences et Technologies de l'Information et de La Communication Pour l'Éducation et La Formation*, *14*, 45–80. https://doi.org/10.3406/stice.2007.949

Delpeyroux, S., & Bachelet, R. (2015, June). Intégrer un MOOC dans un cursus de formation initiale. *Actes Du Colloque Questions de P´edagogie Dans l'Enseignement Sup´erieur (QPES) 2015*.

Colloque Questions de Pédagogie dans l'Enseignement Supérieur (QPES) 2015, Brest, France. https://shs.hal.science/halshs-01165975

Di Mitri, D., Scheffel, M., Drachsler, H., Börner, D., Ternier, S., & Specht, M. (2017). Learning pulse: A machine learning approach for predicting performance in self-regulated learning using multimodal data. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 188–197. https://doi.org/10.1145/3027385.3027447

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*, *34*(4), 338–349. https://doi.org/10.1111/jcal.12288

DiCerbo, K., Shute, V., & Kim, Y. J. (2017). The future of assessment in technology rich environments: Psychometric considerations. *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy*, 1–21.

Djoudi, M., Luengo, V., El Kechaï, H., Cerisier, J.-F., Maugard, E., Cherigny, F., Champalle, O., Iksal, S., & Beust, P. (2018). *Learning Analytics: Terminologie du Learning Analytics* [Research Report]. Direction du Numérique pour l'Éducation, Ministère de l'Enseignement Supérieur de la Recherche et de l'Innovation (DNE-MESRI). https://hal.archives-ouvertes.fr/hal-02464092

Doghonadze, N., Aliyev, A., Halawachy, H., Knodel, L., & Adedoyin, A. S. (2020). *The Degree of Readiness to Total Distance Learning in the Face of COVID-19 -. 5*(2), 41.

Doran, G. T. (1981). There's a SMART way to write management's goals and objectives. *Management Review*, *70*(11), 35–36.

Durik, A. M., & Matarazzo, K. L. (2009). Revved up or turned off? How domain knowledge changes the relationship between perceived task complexity and task interest. *Learning and Individual Differences*, *19*(1), 155–159.

Ebner, M., Schön, S., & Braun, C. (2020). More Than a MOOC—Seven Learning and Teaching Scenarios to Use MOOCs in Higher Education and Beyond. In S. Yu, M. Ally, & A.

Tsinakos (Eds.), *Emerging Technologies and Pedagogies in the Curriculum* (pp. 75–87). Springer. https://doi.org/10.1007/978-981-15-0618-5_5

ecma, I. (2017). *Standard ECMA-404, 2nd edition, December 2017*. ecma International. https://www.ecma-international.org/publications-and-standards/standards/ecma-404/

edX Inc. (2023). *2. Open edX Architecture* [Documentation]. Open edX Developer's Guide Documentation. https://edx.readthedocs.io/projects/edx-developer-guide/en/latest/architecture.html

Efklides, A., & Petkaki, C. (2005). Effects of mood on students' metacognitive experiences. *Learning and Instruction*, *15*(5), 415–431. https://doi.org/10.1016/j.learninstruc.2005.07.010

Efklides, A., & Volet, S. (Eds.). (2005). *Feelings and emotions in the learning process* (Vol. 15). Elsevier.

EFRN. (2014). *What is Flow ?* European Flow Researchers Network. https://efrn.eu/

El Mawas, N., Ghergulescu, I., Moldovan, A.-N., & Muntean, C. (2018). Pedagogical based Learner Model Characteristics. *Ireland International Conference on Education*, 138–142. http://www.newtonproject.eu/

El Mawas, N., Gilliot, J.-M., Garlatti, S., Euler, R., & Pascual, S. (2019). As One Size Doesn't Fit All, Personalized Massive Open Online Courses Are Required. In B. M. McLaren, R. Reilly, S. Zvacek, & J. Uhomoibhi (Eds.), *Computer Supported Education* (Vol. 1022, pp. 470–488). Springer International Publishing. https://doi.org/10.1007/978-3-030-21151-6_22

El Mawas, N., Gilliot, J.-M., Garlatti, S., Euler, R., & Pascual, S. (2018). Towards personalized content in massive open online courses. *10th International Conference on Computer Supported Education*. https://doi.org/10.5220/0006816703310339

El Mawas, N., & Heutte, J. (2019). A flow measurement instrument to test the students' motivation in a computer science course. *CSEDU 2019 - Proceedings of the 11th International Conference on Computer Supported Education*, *1*, 495–505. https://doi.org/10.5220/0007771504950505

Ellis, G. D., Voelkl, J. E., & Morris, C. (1994). Measurement and analysis issues with explanation of variance in daily experience using the flow model. *Journal of Leisure Research*, *26*(4), 337–356. https://doi.org/10.1080/00222216.1994.11969966

Emily Jones, M. (2020). *LibGuides: Systematic Reviews: Creating a PRISMA flow diagram*. https://www.prisma-statement.org//PRISMAStatement/FlowDiagram

Engeser, S. (Ed.). (2012). *Advances in Flow Research*. Springer New York. https://doi.org/10.1007/978-1-4614-2359-1

Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, *32*(3), 158–172. https://doi.org/10.1007/s11031-008-9102-4

Engeser, S., Rheinberg, F., Vollmeyer, R., & Bischoff, J. (2005). Motivation, flow-Erleben und Lernleistung in universitären Lernsettings. *Zeitschrift Für Pädagogische Psychologie*, *19*(3), 159–172. https://doi.org/10.1024/1010-0652.19.3.159

Engeser, S., & Schiepe-Tiska, A. (2012). Historical Lines and an Overview of Current Research on Flow. In S. Engeser (Ed.), *Advances in Flow Research* (pp. 1–22). Springer New York. https://doi.org/10.1007/978-1-4614-2359-1_1

Engeser, S., Schiepe-Tiska, A., & Peifer, C. (2021). Historical Lines and an Overview of Current Research on Flow. In C. Peifer & S. Engeser (Eds.), *Advances in Flow Research* (pp. 1–29). Springer International Publishing. https://doi.org/10.1007/978-3-030-53468-4_1

Eppler, M. A., & Harju, B. L. (1997). Achievement Motivation Goals in Relation to Academic Performance in Traditional and Nontraditional College Students. *Research in Higher Education*, *38*(5), 557–573. https://doi.org/10.1023/A:1024944429347

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS, 2021/0106/COD, CJ40/9/07994 (2021). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

European Union. (2022, September 29). *Regulatory framework proposal on artificial intelligence*. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Fang, J., Tang, L., Yang, J., & Peng, M. (2019). Social interaction in MOOCs: The mediating effects of immersive experience and psychological needs satisfaction. *Telematics and Informatics*, *39*, 75–91. https://doi.org/10.1016/j.tele.2019.01.006

Fave, A. D., & Massimini, F. (1988). Modernization and the Changing Contexts of Flow in Work and Leisure. In M. Csíkszentmihályi & I. S. Csíkszentmihályi (Eds.), *Optimal Experience: Psychological Studies of Flow in Consciousness*. Cambridge University Press.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, *17*(3), 37–37.

Feng, X.-L., Hu, X.-C., Fan, K.-Y., & Yu, T. (2020). A Brief Discussion About the Impact of Coronavirus Disease 2019 on Teaching in Colleges and Universities of China. *2020 International Conference on E-Commerce and Internet Technology (ECIT)*, 167–170. https://doi.org/10.1109/ECIT50008.2020.00044

Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from imbalanced data sets* (Vol. 10). Springer. https://doi.org/10.1007/978-3-319-98074-4

Ferreira Marques, J., & Bernardino, J. (2020). Analysis of Data Anonymization Techniques. *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 235–241. https://doi.org/10.5220/0010142302350241

Fratamico, L., Conati, C., Kardan, S., & Roll, I. (2017). Applying a Framework for Student Modeling in Exploratory Learning Environments: Comparing Data Representation Granularity to Handle Environment Complexity. *International Journal of Artificial Intelligence in Education*, *27*(2), 320–352. https://doi.org/10.1007/s40593-016-0131-y

Fu, F.-L., Su, R.-C., & Yu, S.-C. (2009). EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education*, *52*(1), 101–112. https://doi.org/10.1016/j.compedu.2008.07.004

Galindo, G. (2017). À la recherche d'un idéal-type pour caractériser la GRH de la start-up high-tech: *Revue de Gestion Des Ressources Humaines*, *N° 103*(1), 55–70. https://doi.org/10.3917/grhu.103.0055

García Calvo, T., Jiménez Castuera, R., Santos-Rosa Ruano, F. J., Reina Vaíllo, R., & Cervelló Gimeno, E. (2008). Psychometric Properties of the Spanish Version of the Flow State Scale. *The Spanish Journal of Psychology*, *11*(2), 660–669. https://doi.org/10.1017/S1138741600004662

Gartner. (2014, October 21). *Gartner Says Advanced Analytics Is a Top Business Priority* [Research & Advisory]. Gartner. https://www.gartner.com/en/newsroom/press-releases/2014-10-21-gartner-says-advanced-analytics-is-a-top-business-priority

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (Second edition). O'Reilly Media, Inc.

Ghaleb, E., Popa, M., Hortal, E., Asteriadis, S., & Weiss, G. (2018). Towards Affect Recognition through Interactions with Learning Materials. *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 372–379. https://doi.org/10.1109/ICMLA.2018.00062

Ghani, J. A., & Deshpande, S. P. (1994). Task characteristics and the experience of optimal flow in human—Computer interaction. *The Journal of Psychology*, *128*(4), 381–391. https://doi.org/10.1080/00223980.1994.9712742

Giannandrea, L., & Sansoni, M. (2013). A literature review on intelligent tutoring systems and on student profiling. *Learning & Teaching with Media & Technology*, *287*, 287–294.

Gilliot, J.-M., Garlatti, S., Rebai, I., & Belen-Sapia, M. (2013, May). Le concept de iMOOC pour une ouverture maîtrisée. *EIAH 2013: atelier thématique MOOC-Massive Open Online Course-État des lieux de la recherche francophone*. Environnements Informatiques pour l'Apprentissage Humain 2013, Toulouse, France.

Google Developers. (2022a, July 18). *Classification: Accuracy | Machine Learning | Google Developers* [MOOC]. https://developers.google.com/machine-learning/crash-course/classification/accuracy

Google Developers. (2022b, July 18). *Classification: Precision and Recall | Machine Learning | Google Developers* [MOOC]. https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall

Google Developers. (2022c, July 18). *Classification: ROC Curve and AUC | Machine Learning | Google Developers* [MOOC]. https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc

Google Developers. (2022d, July 18). *Classification: Thresholding | Machine Learning | Google Developers* [MOOC]. https://developers.google.com/machine-learning/crash-course/classification/thresholding

Google Developers. (2022e, July 18). *Classification: True vs. False and Positive vs. Negative | Machine Learning | Google Developers* [MOOC]. https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative

Google Developers. (2022f, July 18). *What is Machine Learning? | Google Developers* [MOOC]. https://developers.google.com/machine-learning/intro-to-ml/what-is-ml

GoogleDoodle. (2023, September 29). *Mihály Csíkszentmihályi's 89th Birthday! #GoogleDoodle*. Mihály Csíkszentmihályi's 89th Birthday. https://www.google.com/doodles/mihaly-csikszentmihalyis-89th-birthday

Goopio, J., & Cheung, C. (2021). The MOOC dropout phenomenon and retention strategies. *Journal of Teaching in Travel & Tourism*, *21*(2), 177–197. https://doi.org/10.1080/15313220.2020.1809050

Gottgtroy, M. P. B., Rodrigues, M. J., & de Sousa, M. T. G. (1970). Data mining agents. *WIT Transactions on Information and Communication Technologies*, *22*. https://doi.org/10.2495/DATA980111

Guo, Y. M., & Poole, M. S. (2009). Antecedents of flow in online shopping: A test of alternative models. *Information Systems Journal*, *19*(4), 369–390. https://doi.org/10.1111/j.1365-2575.2007.00292.x

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*(Mar), 1157–1182.

Han, J.-W., Kang, K. I., & Joung, J. (2020). Enhancing Happiness for Nursing Students through Positive Psychology Activities: A Mixed Methods Study. *International Journal of Environmental Research and Public Health*, *17*(24), Article 9274. https://doi.org/10.3390/ijerph17249274

Han, S. (1988). The relationship between life satisfaction and flow in elderly Korean immigrants. In I. S. Csíkszentmihályi & M. Csíkszentmihályi (Eds.), *Optimal Experience: Psychological Studies of Flow in Consciousness* (pp. 138–149). Cambridge University Press. https://doi.org/10.1017/CBO9780511621956.008

Harris, D. J., Allen, K. L., Vine, S. J., & Wilson, M. R. (2021). A systematic review and meta-analysis of the relationship between flow states and performance. *International Review of Sport and Exercise Psychology*, 1–29. https://doi.org/10.1080/1750984X.2021.1929402

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 1322–1328.

Hebb, D. O. (1955). Drives and the C. N. S. (conceptual nervous system). *Psychological Review*, *62*(4), 243–254. https://doi.org/10.1037/h0041823

Heutte, J. (2014). Persister dans la conception de son environnement personnel d'apprentissage: Contributions et complémentarités de trois théories du self (autodétermination, auto-efficacité, autotélisme-flow). *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*, *21*(1), 149–184. https://doi.org/10.3406/stice.2014.1095

Heutte, J. (2017a). Apports de la théorie de l'autotélisme-flow à la recherche fondamentale en sciences de l'éducation. *Le Journal Des Psychologues*, *4*, 42–47. https://doi.org/10.3917/jdp.346.0042

Heutte, J. (2017b). L'environnement optimal d'apprentissage: Contribution de la recherche empirique sur les déterminants psychologiques de l'expérience positive subjective aux sciences de l'éducation et de la formation des adultes. *Sciences & Bonheur*, *2*(2), 82–99. https://hal.archives-ouvertes.fr/hal-01597551

Heutte, J. (2017c). Motivations, volition et expérience du flow: Quelques références théoriques pour l'étude des communautés d'apprenance. In O. Las Vergnas, *Le e-learning informel?: Des apprentissages diffus, noyés dans la participation en ligne* (pp. 199–241). Archives contemporaines.

Heutte, J. (2019). *Les fondements de l'éducation positive: Perspective psychosociale et systémique de l'apprentissage*. Dunod.

Heutte, J. (2020). Psychologie positive et formation des adultes: Le flow ou le plaisir de comprendre tout au long de la vie. *Savoirs*, *54*(3), 17–61. https://doi.org/10.3917/savo.054.0017

Heutte, J. (2021). L'expérience autotélique dans les EIAH: Genèse socio-historique, épistémologique et critique de l'émergence des technologies positives pour l'apprentissage. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*, *28*(Spécial). http://sticef.org/num/vol2021/28.2.5.heutte/28.2.5.heutte.htm

Heutte, J. (2015). *L'environnement optimal d'apprentissage vidéo-ludique: Contribution de la psychologie positive à la définition d'une ingénierie ludo-éduquante autotélique.* [Séminaire]. CNAM-ENJIM "Bases cognitives, sociales et émotionnelles des jeux et médias interactifs numériques," Angoûleme, France.

Heutte, J., & Fenouillet, F. (2010, September). Propositions pour une mesure de l'expérience optimale (état de FLOW) en contexte éducatif. *Actes du congrès de l'Actualité de la recherche en éducation et en formation*. Actualité de la recherche en éducation et en formation, Geneva, Switzerland. https://plone.unige.ch/aref2010/communications-orales/premiers-auteurs-en-h/Propositions%20.pdf/view

Heutte, J., Fenouillet, F., Boniwell, I., Martin-Krumm, C., & Csíkszentmihályi, M. R. (2014, October 20). Optimal learning experience in digital environments: Theoretical concepts, measure and modelisation. *Symposium "Digital Learning in 21st Century Universities."* Digital Learning in 21st-Century Universities, Atlanta, USA. https://hal.archives-ouvertes.fr/hal-01470855

Heutte, J., Fenouillet, F., Kaplan, J., Martin-Krumm, C., & Bachelet, R. (2016). Chapter 9 The EduFlow model: A contribution toward the study of optimal learning environments. In *Flow experience* (pp. 127–143). Springer. https://doi.org/10.1007/978-3-319-28634-1_9

Heutte, J., Fenouillet, F., Martin-Krumm, C., Boniwell, I., & Csíkszentmihályi, M. R. (2016, June 29). Proposal for a conceptual evolution of the flow in education (EduFlow) model. *8th European Conference on Positive Psychology (ECPP 2016)*. 8th European Conference on Positive Psychology (ECPP 2016), Angers, France. https://hal.archives-ouvertes.fr/hal-01470857

Heutte, J., Fenouillet, F., Martin-Krumm, C., Gute, G., Raes, A., Gute, D., Bachelet, R., & Csíkszentmihályi, M. R. (2021). Optimal Experience in Adult Learning: Conception and Validation of the Flow in Education Scale (EduFlow-2). *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.828027

Heutte, J., Kaplan, J., Fenouillet, F., Caron, P.-A., & Rosselle, M. (2014). MOOC user persistence. Lessons from French Educational Policy Adoption and Deployment of a Pilot Course. *Learning Technology for Education in Cloud*, 13–24.

Heutte, J., Ramírez Luelmo, S. I., El Mawas, N., & LasVergnas, O. (2022, June 24). *Lexicometrical analysis method to support Scoping Review on social dimensions of flow* [Symposium]. 10th European Conference on Positive Psychology (ECPP 2022), Reykjavik, Iceland. https://hal.science/hal-02499145

Hignite, K. (2009). Low-Carbon Computing. *EDUCAUSE Review*, *44*(6), 34–50. https://eric.ed.gov/?id=EJ864634

Hirt, E. R., McDonald, H. E., & Melton, R. J. (1996). Processing goals and the affect–performance link: Mood as main effect or mood as input? *Journal of Personality and Social Psychology*, *71*(2), 245–261. https://doi.org/10.1037/0022-3514.71.2.245

Hoffman, D. L., & Novak, T. P. (1996). Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of Marketing*, *60*(3), 50–68.

Hoffman, D. L., & Novak, T. P. (2009). Flow Online: Lessons Learned and Future Prospects. *Journal of Interactive Marketing*, *23*, 23–34. https://doi.org/10.1016/J.INTMAR.2008.10.003

Hollands, F. M., & Tirthali, D. (2014). *MOOCs: Expectations and Reality. Full Report*. Teachers College, Columbia University. https://eric.ed.gov/?id=ED547237

Houghton, J. (2005). Global warming. *Reports on Progress in Physics*, *68*(6), 1343. https://doi.org/10.1088/0034-4885/68/6/R02

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Hussain, S., Monkaresi, H., & Calvo, R. A. (2012). Categorical vs. Dimensional Representations in Multimodal Affect Detection during Learning. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Intelligent Tutoring Systems* (pp. 78–83). Springer. https://doi.org/10.1007/978-3-642-30950-2_11

IBM. (2016). *What is Logistic regression?* https://www.ibm.com/topics/logistic-regression

IBM. (2020, December 18). *What is Machine Learning?* IBM Cloud Learn Hub. https://www.ibm.com/cloud/learn/machine-learning

IBM. (2022, November 30). *Explainable AI (XAI).* IBM Watson. https://www.ibm.com/watson/explainable-ai

Iksal, S. (2012). *Ingénierie de l'observation basée sur la prescription en EIAH* [Habilitation à Diriger des Recherches, Université du Maine]. https://theses.hal.science/tel-00991970

Isbell, C., Littman, M. L., & Norvig, P. (2023). Software Engineering of Machine Learning Systems. *Communications of the ACM*, *66*(2), 35–37. https://doi.org/10.1145/3539783

Jackman, P. C., Crust, L., & Swann, C. (2017). Systematically comparing methods used to study flow in sport: A longitudinal multiple-case study. *Psychology of Sport and Exercise*, *32*, 113–123. https://doi.org/10.1016/j.psychsport.2017.06.009

Jackson, S. A., & Eklund, R. C. (2002). Assessing Flow in Physical Activity: The Flow State Scale-2 and Dispositional Flow Scale-2. *Journal of Sport and Exercise Psychology*, *24*(2), 133–150. Academic Search Complete. https://doi.org/10.1123/jsep.24.2.133

Jackson, S. A., & Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: The Flow State Scale. *Journal of Sport and Exercise Psychology*, *18*(1), 17–35. https://doi.org/10.1123/jsep.18.1.17

Jackson, S. A., Martin, A., & Eklund, R. C. (2008). Long and Short Measures of Flow: The Construct Validity of the FSS-2, DFS-2, and New Brief Counterparts. *Journal of Sport & Exercise Psychology*, *30*, 561–587. https://doi.org/10.1123/jsep.30.5.561

Jacobs, A. J. (2013, April 20). Two Cheers for Web U! *The New York Times*. https://www.nytimes.com/2013/04/21/opinion/sunday/grading-the-mooc-university.html

Jacoby, J. (1978). Consumer Research: How valid and useful are all our consumer behavior research findings?: A State of the Art Review. *Journal of Marketing*, *42*(2), 87–96. https://doi.org/10.1177/002224297804200213

Jézégou, A. (2010). Le dispositif GEODE pour évaluer l'ouverture d'un environnement éducatif. *Journal of Distance Education/Revue de l'Education à Distance*, *24*(2), 83–108.

Jézégou, A. (2012). La présence en e-learning: Modèle théorique et perspectives pour la recherche. *Journal of Distance Éducation/Revue de l'éducation à Distance*, *26*(1).

Johnson, D. G., & Verdicchio, M. (2023). Ethical AI is Not about AI. *Communications of the ACM*, *66*(2), 32–34. https://doi.org/10.1145/3576932

Johnson, J. A., Keiser, H. N., Skarin, E. M., & Ross, S. R. (2014). The dispositional flow scale–2 as a measure of autotelic personality: An examination of criterion-related validity. *Journal of Personality Assessment*, *96*(4), 465–470. https://doi.org/10.1080/00223891.2014.891524

Jordan, K. (2014). Initial trends in enrolment and completion of massive open online courses. *The International Review of Research in Open and Distributed Learning*, *15*(1). https://doi.org/10.19173/irrodl.v15i1.1651

Jung, Y., & Lee, J. (2018). Learning Engagement and Persistence in Massive Open Online Courses (MOOCS). *Computers & Education*, *122*, 9–22. https://doi.org/10.1016/j.compedu.2018.02.013

Just Dance 2024 Edition [@justdancegame]. (2022, November 28). *Mihaly's name is a reference to Mihály Csíkszentmihályi, the psychologist who discussed the concept of Flow!* [Tweet]. X (Formerly Twitter). https://twitter.com/justdancegame/status/1597287275553247232

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science*, *306*(5702), 1776–1780. https://doi.org/10.1126/science.1103572

Kaplan, J. (2014). Co-regulation in Technology Enhanced Learning Environments. In L. Uden, J. Sinclair, Y.-H. Tao, & D. Liberona (Eds.), *Learning Technology for Education in Cloud. MOOC and Big Data* (pp. 72–81). Springer International Publishing. https://doi.org/10.1007/978-3-319-10671-7_7

Kawabata, M., & Mallett, C. J. (2011). Flow experience in physical activity: Examination of the internal structure of flow from a process-related perspective. *Motivation and Emotion*, *35*(4), 393–402. https://doi.org/10.1007/s11031-011-9221-1

Kay, J., & Kummerfeld, B. (2019). From data to personal user models for life-long, life-wide learners. *British Journal of Educational Technology*, *50*(6), 2871–2884.

Keller, J., & Bless, H. (2008). Flow and Regulatory Compatibility: An Experimental Approach to the Flow Model of Intrinsic Motivation. *Personality and Social Psychology Bulletin*, *34*(2), 196–209. https://doi.org/10.1177/0146167207310026

Keller, J., & Landhäußer, A. (2012). The Flow Model Revisited. In S. Engeser (Ed.), *Advances in Flow Research* (pp. 51–64). Springer New York. https://doi.org/10.1007/978-1-4614-2359-1_3

Keller, J., Ringelhan, S., & Blomann, F. (2011). Does skills–demands compatibility result in intrinsic motivation? Experimental test of a basic notion proposed in the theory of flow-experiences. *The Journal of Positive Psychology*, *6*(5), 408–417. https://doi.org/10.1080/17439760.2011.604041

Kellogg, S., Booth, S., & Oliver, K. (2014). A social network perspective on peer supported learning in MOOCs for educators. *International Review of Research in Open and Distributed Learning*, *15*(5), 263–289. https://doi.org/10.19173/irrodl.v15i5.1852

Kerr, R. A. (2007). Global warming is changing the world. *Science*, *316*(5822), 188–190. https://doi.org/10.1126/science.316.5822.188

Kichu, M., & Bhattacharya, M. (2021). COVID-19 Pandemic impels surge in MOOC learning and the New Normal: A Literature Review. *International Journal of Innovative Research in Technology*, *7*(10), 282–285. https://doi.org/10.6084/m9.figshare.14350622

Kiili, K., & Lainema, T. (2008). Foundation for measuring engagement in educational games. *Journal of Interactive Learning Research*, *19*(3), 469–488.

Kim, A.-Y., Tack, H., & Lee, C.-H. (2010). The development and validation of a learning flow scale for adults. *The Korean Journal of Educational Psychology*, *24*(1), 39–59.

Kim, Y.-Y., Kim, S.-W., & Koo, K.-H. (2017). Development and Validation of the Learning Flow Scale for High School Students. *Journal of Curriculum and Evaluation*, *20*(4), 95–119. https://doi.org/10.29221/jce.2017.20.4.95

Kirkpatrick, K. (2023). The Carbon Footprint of Artificial Intelligence. *Communications of the ACM*, *66*(8), 17–19. https://doi.org/10.1145/3603746

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering—Version 2.3* (EBSE Technical Report 2.3). Keele University & University of Durham.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 170–179. https://doi.org/10.1145/2460296.2460330

Kompare, B., Bratko, I., Steinman, F., & Džeroski, S. (1994). Using machine learning techniques in the construction of models I. Introduction. *Ecological Modelling*, *75–76*, 617–628. https://doi.org/10.1016/0304-3800(94)90054-X

Kop, R. (2011). The challenges to connectivist learning on open online networks: Learning experiences during a massive open online course. *International Review of Research in Open and Distributed Learning*, *12*(3), 19–38. https://doi.org/10.19173/irrodl.v12i3.882

Koutropoulos, A., Gallagher, M. S., Abajian, S. C., de Waard, I., Hogue, R. J., Keskin, N. O., & Rodriguez, C. O. (2012). Emotive Vocabulary in MOOCs: Context & Participant Retention. *European Journal of Open, Distance and E-Learning.*

Kühne, K., Bartsch, N., Tate, R. D., Higson, J., & Habet, A. (2022). "Carbon Bombs"—Mapping key fossil fuel projects. *Energy Policy*, *166*, 112950. https://doi.org/10.1016/j.enpol.2022.112950

Kulkarni, A., Chong, D., & Batarseh, F. A. (2020). Foundations of data imbalance and solutions for a data democracy. In *Data Democracy* (pp. 83–106). Elsevier. https://doi.org/10.1016/B978-0-12-818366-3.00005-8

Kumar, S., & Chong, I. (2018). Correlation Analysis to Identify the Effective Data in Machine Learning: Prediction of Depressive Disorder and Emotion States. *International Journal of*

Environmental Research and Public Health, *15*(12), 2907. https://doi.org/10.3390/ijerph15122907

Laird, K. T., Vergeer, I., Hennelly, S., & Siddarth, P. (2021). Conscious dance: Perceived benefits and psychological well-being of participants. *Complementary Therapies in Clinical Practice*, *44*. https://doi.org/10.1016/j.ctcp.2021.101440

Lane, L. N. (2012, August 15). Three Kinds of MOOCs [Personal Blog]. *Lisahistory*. https://lisahistory.net/wordpress/2012/08/three-kinds-of-moocs/

Larson, R., & Csíkszentmihályi, M. R. (2014). The Experience Sampling Method. In M. Csikszentmihalyi, *Flow and the Foundations of Positive Psychology* (pp. 21–34). Springer Netherlands. https://doi.org/10.1007/978-94-017-9088-8_2

Lecomte, J. (2006). *Donner un sens à sa vie*. Odile Jacob.

Lecomte, J. (2009). La théorie du flux. Comment la motivation intrinsèque donne du sens à notre vie. In *Traité de psychologie de la motivation* (pp. 107–124). Dunod. https://doi.org/10.3917/dunod.carre.2009.01.0107

Leetaru, K. (2019, January 15). A Reminder That Machine Learning Is About Correlations Not Causation. *Forbes*. https://www.forbes.com/sites/kalevleetaru/2019/01/15/a-reminder-that-machine-learning-is-about-correlations-not-causation/

Lefèvre, M., Guin, N., Marty, J.-C., & Clerc, F. (2016). Supporting Teaching Teams in Personalizing MOOCs Course Paths. In K. Verbert, M. Sharples, & T. Klobučar (Eds.), *Adaptive and Adaptable Learning* (pp. 605–609). Springer International Publishing. https://doi.org/10.1007/978-3-319-45153-4_73

Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research*, *18*(17), 1–5. http://jmlr.org/papers/v18/16-365.html

Leyens, J.-P., Yzerbyt, V., & Schadron, G. (1996). *Stéréotypes et cognition sociale* (Vol. 214). Editions Mardaga.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLoS Medicine*, *6*(7), Article e1000100. https://doi.org/10.1371/journal.pmed.1000100

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, J., Gotzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *British Medical Journal*, *339*(July 21st), Article BMJ 2009;339:b2700. https://doi.org/10.1136/bmj.b2700

Liu, L., & Özsu, M. T. (Eds.). (2020). *Encyclopedia of Database Systems*. Springer New York. https://doi.org/10.1007/978-1-4899-7993-3

Liyanagunawardena, T. R., Adams, A. A., & Williams, S. A. (2013). MOOCs: A systematic study of the published literature 2008-2012. *International Review of Research in Open and Distributed Learning*, *14*(3), 202–227. https://doi.org/10.19173/irrodl.v14i3.1455

Lonczak, H. S. (2019, August 28). *How to Measure Flow with Scales and Questionnaires*. PositivePsychology.Com. https://positivepsychology.com/how-to-measure-flow-scales-questionnaires/

Lorenzoni, G., Alencar, P., Nascimento, N., & Cowan, D. (2021). *Machine Learning Model Development from a Software Engineering Perspective: A Systematic Literature Review* (Software Engineering (Cs.SE); Artificial Intelligence (Cs.AI); Machine Learning (Cs.LG) arXiv:2102.07574 [cs.SE]). arXiv. https://doi.org/10.48550/arXiv.2102.07574

Lottick, K., Susai, S., Friedler, S. A., & Wilson, J. P. (2019). *Energy Usage Reports: Environmental awareness as part of algorithmic accountability* (arXiv:1911.08354). arXiv. https://doi.org/10.48550/arXiv.1911.08354

Lusted, L. B. (1971). Signal Detectability and Medical Decision-Making: Signal detectability studies help radiologists evaluate equipment systems and performance of assistants. *Science*, *171*(3977), 1217–1219. https://doi.org/10.1126/science.171.3977.1217

Mack, K., Das, M., Jain, D., Bragg, D., Tang, J., Begel, A., Beneteau, E., Davis, J. U., Glasser, A., Park, J. S., & Potluri, V. (2023). Mixed Abilities and Varied Experiences: A Group Autoethnography of a Virtual Summer Internship. *Communications of the ACM*, *66*(8), 105–113. https://doi.org/10.1145/3604622

Magyaródi, T., Nagy, H., Soltész, P., Mózes, T., & Oláh, A. (2013). Psychometric properties of a newly established flow state questionnaire. *The Journal of Happiness & Well-Being*, *1*(2), 85–96.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, *5*(4), 333–369. https://doi.org/10.1207/s15516709cog0504_2

Mandran, N., & Dupuy-Chessa, S. (2017, September). THEDRE: A Traceable Process for High Quality in Human Centred Computer Science Research. *26th International Conference on Information Systems Development, ISD 2017*. Information Systems Development (ISD), Larnaca, Cyprus. https://hal.science/hal-01996110

Martin, A. J., & Jackson, S. A. (2008). Brief approaches to assessing task absorption and enhanced subjective experience: Examining 'short' and 'core' flow in diverse performance domains. *Motivation and Emotion*, *32*(3), 141–157. https://doi.org/10.1007/s11031-008-9094-0

Maslow, A. H. (1965). Humanistic science and transcendent experiences. *Journal of Humanistic Psychology*, *5*(2), 219–227. https://doi.org/10.1177/002216786500500211

Massimini, F., & Carli, M. (1986). La selezione psicologica umana tra biologia e cultura. *L'esperienza Quotidiana. Milan: Franco Angeli.*

Massimini, F., & Carli, M. (1988). 16. The systematic assessment of flow in daily experience. In *Optimal experience: Psychological studies of flow in consciousness* (1st ed., pp. 266–287). Cambridge University Press. https://doi.org/10.1017/CBO9780511621956.016

Massimini, F., Csíkszentmihályi, M. R., & Fave, A. D. (1988). *4—Flow and biocultural evolution* (pp. 60–82). Cambridge University Press. https://doi.org/10.1017/CBO9780511621956.004

MAUT. (2015, November 16). *A Brief History of MOOCs*. McGill Association of University Teachers. https://www.mcgill.ca/maut/news-current-affairs/moocs/history

Mavrikis, M. (2010). Modelling student interactions in intelligent learning environments: Constructing bayesian networks from data. *International Journal on Artificial Intelligence Tools*, *19*(06), 733–753. https://doi.org/10.1142/S0218213010000406

May, M. (2009). *Using tracking data as reflexive tools to support tutors and learners in distance learning situations: An application to Computer-Mediated Communications* [Doctoral dissertation]. INSA de Lyon.

Mayers, P. L. (1978). *Flow in Adolescence and its Relation to School Experience.* [Unpublished Doctoral dissertation, University of Chicago]. https://www.proquest.com/openview/b97f3bddcbcf3b41951f5b5d7580d006/1.pdf?pq-origsite=gscholar&cbl=18750&diss=y

McAuley, A., Stewart, B., Siemens, G., & Cormier, D. (2010). *The MOOC model for digital practice* (pp. 1–64). University of Prince Edward Island. https://www.oerknowledgecloud.org/record500

Medina-Medina, N., & García-Cabrera, L. (2016). A taxonomy for user models in adaptive systems: Special considerations for learning environments. *The Knowledge Engineering Review*, *31*(2), 124–141. https://doi.org/10.1017/S0269888916000035

Mehta, P., Bukov, M., Wang, C.-H., Day, A. G., Richardson, C., Fisher, C. K., & Schwab, D. J. (2019). A high-bias, low-variance introduction to machine learning for physicists. *Physics Reports*, *810*, 1–124. https://doi.org/10.1016/j.physrep.2019.03.001

Merzeau, L. (2013). Traces numériques et recrutement: Du symptôme au cheminement. *Traces Numériques: De La Production à l'interprétation*, 35–53.

Mesurado, B. (2008a). Elaboración y análisis discriminativo de los items del Cuestionario de Experiencia Óptima para niños y adolescentes. *Memorias de las XV Jornadas de Investigación y*

Cuarto Encuentro de Investigadores en Psicología del Mercosur, *2*, 478–480. https://www.aacademica.org/000-032/652

Mesurado, B. (2008b). Validez Factorial y Fiabilidad del Cuestionario de Experiencia Óptima (Flow) para niños y adolescentes. *Revista Iberoamericana de Diagnóstico y Evaluación-e Avaliação Psicológica*, *1*(25), 159–178.

Mesurado, B., Cristina Richaud, M., & José Mateo, N. (2016). Engagement, Flow, Self-Efficacy, and Eustress of University Students: A Cross-National Comparison Between the Philippines and Argentina. *The Journal of Psychology*, *150*(3), 281–299. https://doi.org/10.1080/00223980.2015.1024595

Michel, C. (2015). *Analyse des usages des plateformes de construction de connaissances par des méthodes mixtes et réflexives pour l'amélioration de l'appropriation et de la structuration de l'information* [Habilitation à Diriger des Recherches]. Université Lyon 1.

Mille, A. (2013). De la trace à la connaissance à l'ère du Web. Introduction au dossier. *Intellectica. Revue de l'Association Pour La Recherche Cognitive*, *59*(1), 7–28. https://doi.org/10.3406/intel.2013.1083

Milton, D. (2019, June). *Building Connections with Autistic People* [Conference or workshop item (Lecture)]. Asia Pacific Autism Conference 2019, Resorts World Convention Centre, Singapore. https://kar.kent.ac.uk/id/eprint/75301

MIT. (n.d.). *A brief history of MIT*. MIT Admissions. Retrieved November 3, 2023, from https://mitadmissions.org/discover/about-mit/a-brief-history-of-mit/

Mitchell, R. G. (1988). Sociological implications of the flow experience. In M. R. Csíkszentmihályi & I. S. Csíkszentmihályi (Eds.), *Optimal experience: Psychological studies of flow in consciousness* (pp. 36–59). Cambridge University Press New York.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine*, *6*(7). https://doi.org/10.1371/journal.pmed.1000097

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of Machine Learning, second edition*. MIT Press.

Moneta, G. B. (2004). The Flow Experience Across Cultures. *Journal of Happiness Studies*, *5*(2), 115–121. https://doi.org/10.1023/B:JOHS.0000035913.65762.b5

Moneta, G. B. (2012a). On the measurement and conceptualization of flow. In S. Engeser (Ed.), *Advances in flow research* (pp. 23–50). Springer.

Moneta, G. B. (2012b). Opportunity for creativity in the job as a moderator of the relation between trait intrinsic motivation and flow in work. *Motivation and Emotion*, *36*(4), 491–503. https://doi.org/10.1007/s11031-012-9278-5

Moneta, G. B. (2021). On the Conceptualization and Measurement of Flow. In C. Peifer & S. Engeser (Eds.), *Advances in Flow Research* (pp. 31–69). Springer International Publishing. https://doi.org/10.1007/978-3-030-53468-4_2

Moneta, G. B., & Csíkszentmihályi, M. R. (1996). The Effect of Perceived Challenges and Skills on the Quality of Subjective Experience. *Journal of Personality*, *64*(2), 275–310. https://doi.org/10.1111/j.1467-6494.1996.tb00512.x

Monkaresi, H., Bosch, N., Calvo, R. A., & D'Mello, S. K. (2016). Automated detection of engagement using video-based estimation of facial expressions and heart rate. *IEEE Transactions on Affective Computing*, *8*(1), 15–28. https://doi.org/10.1109/TAFFC.2016.2515084

Moon, J., Choi, Y., Park, T., Choi, J., Hong, J.-H., & Kim, K.-J. (2022). Diversifying dynamic difficulty adjustment agent by integrating player state models into Monte-Carlo tree search. *Expert Systems with Applications*, *205*, 117677. https://doi.org/10.1016/j.eswa.2022.117677

Morozov, E. (2013). *To save everything, click here: Technology, solutionism and the urge to fix problems that don't exist*. Allen Lane. https://archive.org/details/tosaveeverything0000moro_q6o4

Mota, R., & Scott, D. (2014). *Education for innovation and independent learning*. Elsevier.

Motlagh, S. E., Amrai, K., Yazdani, M. J., Abderahim, H. altaib, & Souri, H. (2011). The relationship between self-efficacy and academic achievement in high school students. *Procedia - Social and Behavioral Sciences*, *15*, 765–768. https://doi.org/10.1016/j.sbspro.2011.03.180

Muhamedyev, R. (2015). Machine learning methods: An overview. *Computer Modelling & New Technologies*, *19*(6), 14–29.

Muramatsu, P. K., Oliveira, W., Hamari, J., & Oyibo, K. (2023). Does Mouse Click Frequency Predict Students' Flow Experience? *Proceedings of the 56th Hawaii International Conference on System Sciences*, 1281–1290. https://hdl.handle.net/10125/102788

Nabukeera, M. (2020). The COVID-19 and online education during emergencies in higher education: *Archives of Business Research*, *8*(5), Article 5. https://doi.org/10.14738/abr.85.8130

Nakamura, J., & Csíkszentmihályi, M. R. (2002). 7 The concept of flow. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of Positive Psychology* (pp. 89–105). Oxford University Press.

Nakamura, J., & Csíkszentmihályi, M. R. (2009). Chapter 18 Flow theory and research. In S. J. Lopez & C. R. Snyder (Eds.), *The Oxford Handbook of Positive Psychology* (Second Edition, pp. 194–206). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195187243.013.0018

Nakić, J., Granić, A., & Glavinić, V. (2015). Anatomy of student models in adaptive learning systems: A systematic literature review of individual differences from 2001 to 2013. *Journal of Educational Computing Research*, *51*(4), 459–489.

Neunhoeffer, M., & Sternberg, S. (2019). How Cross-Validation Can Go Wrong and What to Do About It. *Political Analysis*, *27*(1), 101–106. https://doi.org/10.1017/pan.2018.39

Ngoc, D. P. T. (2011). *Specification and design of usage analysis services for a tel system* [Doctoral dissertation, Université du Maine]. https://theses.hal.science/tel-00689025

Ngoc, D. P. T., Iksal, S., Choquet, C., & Klinger, E. (2009). UTL-CL: A Declarative Calculation Language Proposal for a Learning Tracks Analysis Process. *2009 Ninth IEEE International Conference on Advanced Learning Technologies*, 681–685. https://doi.org/10.1109/ICALT.2009.37

Nguyen, L., & Do, P. (2008). Learner model in adaptive learning. *World Academy of Science, Engineering and Technology*, *45*(70), 395–400.

Novak, T. P., Hoffman, D. L., & Yung, Y. F. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science*, *19*(1), 22–42. https://doi.org/10.1287/mksc.19.1.22.15184

Obadă, D.-R. (2021). Pretesting Flow Questionnaire Design Using Eye-Tracking: An Exploratory Study. *Argumentum. Journal of the Seminar of Discursive Logic, Argumentation Theory and Rhetoric*, *1*(19).

Oertig, D., Schüler, J., Brandstätter, V., & Augustine, A. A. (2014). The influence of avoidance temperament and avoidance-based achievement goals on flow. *Journal of Personality*, *82*(3), 171–181. https://doi.org/10.1111/jopy.12043

Oláh, A. (2005). *Az optimális élmény mérésének lehetőségei: Egy új szituáció-specifikus Flow Kérdőív tesztkönyve*. HI Press.

Onyema, E. M., Eucheria, N. C., Obafemi, F. A., Sen, S., Atonye, F. G., Sharma, A., & Alsayed, A. O. (2020). Impact of Coronavirus pandemic on education. *Journal of Education and Practice*, *11*(13), 108–121.

Papadopoulos, F., Corrigan, L. J., Jones, A., & Castellano, G. (2013). Learner modelling and automatic engagement recognition with robotic tutors. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 740–744. https://doi.org/10.1109/ACII.2013.137

Pappano, L. (2012, November 2). The Year of the MOOC. *The New York Times*. https://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html

Parker, K., Horowitz, J. M., & Minkin, R. (2020). How the coronavirus outbreak has–and hasn't–changed the way Americans work. *Pew Research Center*, *9*. https://www.pewresearch.org/social-trends/wp-content/uploads/sites/3/2020/12/psdt_12.09.20_covid.work_fullreport.pdf

Parks, B. Kirshner. (1996). *"Flow", boredom, and anxiety in therapeutic work: A study of psychotherapists' intrinsic motivation and professional development* [Doctoral dissertation, University of Chicago]. http://pi.lib.uchicago.edu/1001/cat/bib/2567351

Parnas, D. L., Shore, J. E., & Weiss, D. (1976). Abstract types defined as classes of variables. *ACM SIGPLAN Notices*, *11*(SI), 149–154. https://doi.org/10.1145/942574.807133

Payne, B. R., Jackson, J. J., Noh, S. R., & Stine-Morrow, E. A. (2011). In the zone: Flow state and cognition in older adults. *Psychology and Aging*, *26*(3), 738–743. https://doi.org/10.1037/a0022359

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

Peifer, C. (2012). Psychophysiological Correlates of Flow-Experience. In S. Engeser (Ed.), *Advances in Flow Research* (pp. 139–164). Springer New York, NY.

Peifer, C., & Engeser, S. (Eds.). (2021a). *Advances in Flow Research* (2nd ed.). Springer Cham. https://link.springer.com/book/10.1007/978-3-030-53468-4

Peifer, C., & Engeser, S. (2021b). Theoretical Integration and Future Lines of Flow Research. In C. Peifer & S. Engeser (Eds.), *Advances in Flow Research* (pp. 417–439). Springer International Publishing. https://doi.org/10.1007/978-3-030-53468-4_16

Peifer, C., & Tan, J. (2021). The Psychophysiology of Flow Experience. In C. Peifer & S. Engeser (Eds.), *Advances in Flow Research* (pp. 191–230). Springer International Publishing. https://doi.org/10.1007/978-3-030-53468-4_8

Peifer, C., Wolters, G., Harmat', L., Heutte, J., Tan, J., Freire, T., Tavares, D., Fonte, C., Andersen, F. O., Hout, J. van den, Pola, L., Ceja, L., & Triberti, S. (2022). A Scoping Review of Flow Research. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2022.815665

Pels, F., Kleinert, J., & Mennigen, F. (2018). Group flow: A scoping review of definitions, theoretical approaches, measures and findings. *PLoS ONE*, *13*(12). https://doi.org/10.1371/journal.pone.0210117

Peraya, D., & Luengo, V. (2019). Les Learning Analytics vus par Vanda Luengo: Entretien. *Distances et médiations des savoirs*, *27*. https://doi.org/10.4000/dms.4096

Pérez-Sanagustín, M., Hernández-Correa, J., Gelmi, C., Hilliger, I., & Rodriguez, M. F. (2016). Does Taking a MOOC as a Complement for Remedial Courses Have an Effect on My Learning Outcomes? A Pilot Study on Calculus. *Proceedings of the 11th European Conference on Technology Enhanced Learning, EC-TEL 2016*, 221–233. https://doi.org/10.1007/978-3-319-45153-4_17

Perota, J., Werner, C., Hoskinson, J., Pennolino, P., & Leddy, B. (Directors). (2023, February 26). Artificial Intelligence: Last Week Tonight with John Oliver (271) [YouTube]. In *Last Week Tonight*. HBO. https://youtu.be/Sqa8Zo2XWc4

Pevneva, I., & Edmunds, P. (2020). Online Learning vs. Extreme Learning in Mining Higher Education under COVID. *E3S Web of Conferences*, *174*, 1–6. https://doi.org/10.1051/e3sconf/202017404001

Pfister, R. (2002). *Flow im Alltag: Untersuchungen zum Quadrantenmodell des Flow-Erlebens und zum Konzept der autotelischen Persönlichkeit mit der experience sampling method (ESM)*. Peter Lang.

Pierrot, L. (2018). *Social circulation of juvenile digital practices and instrumental genesis* [Doctoral dissertation, Université de Poitiers]. https://hal.science/tel-02566693

Pierrot, L., Cerisier, J.-F., El-Kechaï, H., Ramirez, S., & Pottier, L. (2017). Using a Mixed Analysis Process to Identify the Students' Digital Practices. In É. Lavoué, H. Drachsler, K. Verbert, J. Broisin, & M. Pérez-Sanagustín (Eds.), *Data Driven Approaches in Digital Education* (pp. 448–453). Springer International Publishing. https://doi.org/10.1007/978-3-319-66610-5_41

Pilli, O., & Admiraal, W. (2016). A Taxonomy of Massive Open Online Courses. *Contemporary Educational Technology*, *7*(3). https://doi.org/10.30935/cedtech/6174

Piniel, K., & Albert, Á. (2020). Motivation and flow. In M. Lamb, K. Csizér, H. Alastair, & S. Ryan (Eds.), *The Palgrave Handbook of Motivation for Language Learning* (pp. 579–597). Palgrave Macmillan Cham. https://doi.org/10.1007/978-3-030-28380-3_28

Poellhuber, B., Roy, N., & Bouchoucha, I. (2019). Understanding participant's behaviour in massively open online courses. *International Review of Research in Open and Distributed Learning*, *20*(1). https://doi.org/10.19173/irrodl.v20i1.3709

Pokhrel, S., & Chhetri, R. (2021). A literature review on impact of COVID-19 pandemic on teaching and learning. *Higher Education for the Future*, *8*(1), 133–141. https://doi.org/10.1177/2347631120983481

Polyzotis, N., Roy, S., Whang, S. E., & Zinkevich, M. (2017). Data management challenges in production machine learning. *Proceedings of the 2017 ACM International Conference on Management of Data*, 1723–1726. https://doi.org/10.1145/3035918.3054782

Raffel, C. (2023). Building Machine Learning Models Like Open Source Software. *Communications of the ACM*, *66*(2), 38–40. https://doi.org/10.1145/3545111

Ramírez Luelmo, S. I. (2020). Vers une modélisation de l'expérience optimale d'apprentissage via les Learning Analytics. In A. Yessad, S. Jolivet, & C. Michel (Eds.), *8èmes RJC EIAH 2020: Environnements Informatiques pour l'Apprentissage Humain* (pp. 144–149). ATIEF. https://rjceiah20.conference.univ-poitiers.fr/wp-content/uploads/sites/406/2020/05/RAMIREZ_texte_poster.pdf

Ramírez Luelmo, S. I. (2022). Vers une prédiction semi-automatique du flow dans un MOOC. In C. Bonnat & R. Venant (Eds.), *Actes des neuvièmes rencontres jeunes chercheur·e·s en EIAH* (pp. 126–133). ATIEF. https://rjc-eiah-2022.univ-lille.fr/fileadmin/user_upload/rjc-eiah-2022/documents/acteRJCEIAH2022_v1.1.pdf

Ramírez Luelmo, S. I., El Mawas, N., Bachelet, R., & Heutte, J. (2022). Towards a Machine Learning Flow-predicting Model in a MOOC Context. *Proceedings of the 14th International Conference on Computer Supported Education*, 124–134. https://doi.org/10.5220/0011070300003182

Ramírez Luelmo, S. I., El Mawas, N., & Heutte, J. (2020a). A literature review on Learner Models for MOOC to support Lifelong Learning. *Proceedings of the 12th International Conference on Computer Supported Education (CSEDU 2020)*, *1*, 527–539. https://doi.org/10.5220/0009782005270539

Ramírez Luelmo, S. I., El Mawas, N., & Heutte, J. (2020b). Towards Open Learner Models Including the Flow State. *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 305–310. https://doi.org/10.1145/3386392.3399295

Ramírez Luelmo, S. I., El Mawas, N., & Heutte, J. (2021a). Learner Models for MOOC in a Lifelong Learning Context: A Systematic Literature Review. In H. C. Lane, S. Zvacek, & J. Uhomoibhi (Eds.), *Computer Supported Education* (Vol. 1473, pp. 392–415). Springer International Publishing. https://doi.org/10.1007/978-3-030-86439-2_20

Ramírez Luelmo, S. I., El Mawas, N., & Heutte, J. (2021b). Les modèles apprenant pour soutenir l'apprentissage tout au long de la vie: Revue de littérature. In M. Lefevre, C. Michel, T.

Geoffre, M. Rodi, L. Alvarez, & A. Karoui (Eds.), *Actes de la 10e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain* (pp. 200–211). ATIEF. https://hal.archives-ouvertes.fr/hal-03292891

Ramírez Luelmo, S. I., El Mawas, N., & Heutte, J. (2021c). Machine Learning Techniques for Knowledge Tracing: A Systematic Literature Review. *Proceedings of the 13th International Conference on Computer Supported Education*, *1*, 60–70. https://doi.org/10.5220/0010515500600070

Ramírez Luelmo, S. I., El Mawas, N., & Heutte, J. (2022). Existing Machine Learning Techniques for Knowledge Tracing: A Review Using the PRISMA Guidelines. In B. Csapó & J. Uhomoibhi (Eds.), *International Conference on Computer Supported Education* (Vol. 1624, pp. 73–94). Springer International Publishing. https://doi.org/10.1007/978-3-031-14756-2_5

Rao, R. B., Fung, G., & Rosales, R. (2008). On the Dangers of Cross-Validation. An Experimental Evaluation. *Proceedings of the 2008 SIAM International Conference on Data Mining*, 588–596. https://doi.org/10.1137/1.9781611972788.54

Raschka, S., & Mirjalili, V. (2019). *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2* (Third edition). Packt.

Ratner, B. (2017). *Statistical and machine-learning data mining: Techniques for better predictive modeling and analysis of big data* (Third Edition). CRC Press, Taylor & Francis Group.

Redaelli, C., & Riva, G. (2011). Flow for Presence Questionnaire. In L. Canetta, C. Redaelli, & M. Flores (Eds.), *Digital Factory for Human-oriented Production Systems* (1st ed., pp. 3–22). Springer London. https://doi.org/10.1007/978-1-84996-172-1_1

Rheinberg, F. (2020). Intrinsic motivation and flow. *Legacies in Motivation Science*, *6*(3), 199–200. https://doi.org/10.1037/mot0000165

Rheinberg, F., & Engeser, S. (2011). Motivational competence: The joint effect of implicit and explicit motives on self-regulation and flow experience. *Motivation, Consciousness, and Self-Regulation*, 79–87.

Rheinberg, F., & Engeser, S. (2018). Intrinsic Motivation and Flow. In J. Heckhausen & H. Heckhausen (Eds.), *Motivation and Action* (pp. 579–622). Springer Cham. https://doi.org/10.1007/978-3-319-65094-4_14

Rheinberg, F., Vollmeyer, R., & Engeser, S. (2003). Kapitel 14 Die Erfassung des Flow-Erlebens. In J. Stiensmeier-Pelster & F. Rheinberg (Eds.), *Diagnostik von Motivation und Selbstkonzept* (Vol. 2). Hogrefe Verlag GmbH & Company KG. https://doi.org/10.23668/psycharchives.8590

Rissler, R., Nadj, M., Li, M. X., Knierim, M. T., & Maedche, A. (2018). Got flow? Using machine learning on physiological data to classify flow. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–6.

Rissler, R., Nadj, M., Li, M. X., Loewe, N., Knierim, M. T., & Maedche, A. (2020). To be or not to be in flow at work: Physiological classification of flow using machine learning. *IEEE Transactions on Affective Computing*. https://doi.org/10.1109/TAFFC.2020.3045269

Riva, E., Freire, T., & Bassi, M. (2016). The Flow Experience in Clinical Settings: Applications in Psychotherapy and Mental Health Rehabilitation. In L. Harmat, F. Ørsted Andersen, F. Ullén, J. Wright, & G. Sadlo (Eds.), *Flow Experience* (pp. 309–326). Springer International Publishing. https://doi.org/10.1007/978-3-319-28634-1_19

Riva, G., Waterworth, J. A., Waterworth, E. L., & Mantovani, F. (2011). From intention to action: The role of presence. *New Ideas in Psychology*, *29*(1), 24–37. https://doi.org/10.1016/j.newideapsych.2009.11.002

Rodríguez-Sánchez, A., Salanova, M., Cifre, E., & Schaufeli, W. B. (2011). When good is good: A virtuous circle of self-efficacy and flow at work among teachers. *International Journal of Social Psychology*, *26*(3), 427–441. https://doi.org/10.1174/021347411797361257

Roomba. (2023). In *Wikipedia*. https://en.wikipedia.org/wiki/Roomba

Rosas, D. A., Padilla-Zea, N., & Burgos, D. (2023). Validated Questionnaires in Flow Theory: A Systematic Review. *Electronics*, *12*(13), Article 2769. https://doi.org/10.3390/electronics12132769

Ross, J., Sinclair, C., Knox, J., Bayne, S., & Macleod, H. (2014). Teacher experiences and academic identity: The missing components of MOOC pedagogy. *MERLOT Journal of Online Learning and Teaching*, *10*(1), 57–69.

Rosselle, M., Caron, P.-A., & Heutte, J. (2014). A typology and dimensions of a description framework for MOOCs. *European MOOCs Stakeholders Summit 2014, eMOOCs 2014*, 130–139.

Ruck, B. M., Walley, W. J., & Hawkes, H. A. (1993). Biological classification of river water quality using neural networks. *WIT Transactions on Information and Communication Technologies*, *1*, 12. https://doi.org/10.2495/AIENG930251

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Rufi, S., Javaloy, F., Batista-Foguet, J. M., Solanas, A., & Páez, D. (2014). Flow dimensions on daily activities with the Spanish version of the flow scale (DFS). *Spanish Journal of Psychology*, *17*(2), 1–11. https://doi.org/10.1017/sjp.2014.34

Ruiz-Martínez, R., Kuschel, K., & Pastor, I. (2021). Craftswomen entrepreneurs in flow: No boundaries between business and leisure. *Community, Work & Family*, 1–20. https://doi.org/10.1080/13668803.2021.1873106

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Sah, S. (2020). *Machine Learning: A Review of Learning Types* (Preprints 2020, 2020070230). https://doi.org/10.20944/preprints202007.0230.v1

Sahid, D. S. S., Efendi, R., & Putra, E. H. (2020). Rough set and machine learning approach for identifying flow experience in e-learning. *IOP Conference Series: Materials Science and Engineering*, *732*(1), Article 012047. https://doi.org/10.1088/1757-899X/732/1/012047

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS One*, *10*(3), Article e0118432. https://doi.org/10.1371/journal.pone.0118432

Sajno, E., Beretta, A., Novielli, N., & Riva, G. (2022). Follow the Flow: A Prospective on the On-Line Detection of Flow Mental State through Machine Learning. *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, 217–222. https://doi.org/10.1109/MetroXRAINE54828.2022.9967605

Salanova, M., Bakker, A. B., & Llorens, S. (2006). Flow at work: Evidence for an upward spiral of personal and organizational resources. *Journal of Happiness Studies*, *7*(1), 1–22.

Salanova, M., Rodríguez-Sánchez, A. M., Schaufeli, W. B., & Cifre, E. (2014). Flowing Together: A Longitudinal Study of Collective Efficacy and Collective Flow Among Workgroups. *The Journal of Psychology*, *148*(4), 435–455. https://doi.org/10.1080/00223980.2013.806290

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

Samuel, A. L. (1963). Some studies in machine learning using the game of checkers. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and Thought* (pp. 71–105). McGraw-Hill.

Samuel, A. L. (1967). Some Studies in Machine Learning Using the Game of Checkers. II—Recent Progress. *IBM Journal of Research and Development*, *11*(6), 601–617. https://doi.org/10.1147/rd.116.0601

Samuel, A. L. (1969). Some studies in machine learning using the game of checkers. II—Recent progress. *Annual Review in Automatic Programming*, *6*, 1–36. https://doi.org/10.1016/0066-4138(69)90004-4

Sanneman, L. (2023). Understanding Our Robots With the Help of Human-Centered Explainable AI. *XRDS: Crossroads, The ACM Magazine for Students*, *30*(1), 52–57. https://doi.org/10.1145/3611686

Schattke, K. P. (2011). *Flow Experience as Consequence and Self-Determination as Antecedence of Congruence between Implicit and Explicit Motives* [Doctoral dissertation, Technische Universität München]. https://mediatum.ub.tum.de/1078244

Schaufeli, W. B., & Bakker, A. B. (2010). The conceptualization and measurement of work engagement. Defining and measuring work engagement: Bringing clarity to the concept. In A. B. Bakker & M. P. Leiter (Eds.), *Work engagement: A handbook of essential theory and research* (pp. 10–24). Psychology Press.

Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006a). The Measurement of Work Engagement With a Short Questionnaire: A Cross-National Study. *Educational and Psychological Measurement*, *66*(4), 701–716. https://doi.org/10.1177/0013164405282471

Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006b). *Utrecht Work Engagement Scale-9 (UWES-9)*. https://doi.org/10.1037/t05561-000

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Schmidt, V., Goyal, K., Joshi, A., Feld, B., Conell, L., Laskaris, N., Blank, D., Wilson, J., Friedler, S., & Luccioni, S. (2021). *CodeCarbon: Estimate and Track Carbon Emissions from Machine Learning Computing*. https://doi.org/10.5281/zenodo.4658424

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green ai. *Communications of the ACM*, *63*(12), 54–63. https://doi.org/10.1145/3381831

SciencesPo. (2018, January 26). *Accreditation to supervise research (HDR)*. School of Research. https://www.sciencespo.fr/ecole-doctorale/en/content/accreditation-supervise-research-hdr.html

Sein-Echaluce, M. L., Fidalgo-Blanco, Á., García-Peñalvo, F. J., & Conde, M. Á. (2016). iMOOC Platform: Adaptive MOOCs. In P. Zaphiris & A. Ioannou (Eds.), *Learning and Collaboration Technologies* (Vol. 9753, pp. 380–390). Springer Cham. https://doi.org/10.1007/978-3-319-39483-1_35

Seligman, M. E. P., & Csíkszentmihályi, M. R. (2000). Positive psychology: An introduction. *American Psychologist*, *55*(1), 5–14. https://doi.org/10.1037/0003-066X.55.1.5

Seppälä, P., Mauno, S., Feldt, T., Hakanen, J., Kinnunen, U., Tolvanen, A., & Schaufeli, W. (2009). The construct validity of the Utrecht Work Engagement Scale: Multisample and longitudinal evidence. *Journal of Happiness Studies*, *10*(4), 459–481. https://doi.org/10.1007/s10902-008-9100-y

Sevilla, J., Ho, A., & Besiroglu, T. (2023). Please Report Your Compute. *Communications of the ACM*, *66*(5), 30–32. https://doi.org/10.1145/3563035

Shah, D. (2015, December 21). By The Numbers: MOOCS in 2015. *The Report by Class Central*. https://www.classcentral.com/report/moocs-2015-stats/

Shah, D. (2016a, December 22). 6 Biggest MOOC Trends of 2016. *The Report by Class Central*. https://www.classcentral.com/report/biggest-mooc-trends-2016/

Shah, D. (2016b, December 26). By The Numbers: MOOCS in 2016. *The Report by Class Central*. https://www.classcentral.com/report/mooc-stats-2016/

Shah, D. (2017). By The Numbers: MOOCS in 2017. *The Report by Class Central*. https://www.classcentral.com/report/mooc-stats-2017/

Shah, D. (2018, December 11). By The Numbers: MOOCs in 2018. *The Report by Class Central*. https://www.classcentral.com/report/mooc-stats-2018/

Shah, D. (2019, December 3). By The Numbers: MOOCs in 2019. *The Report by Class Central*. https://www.classcentral.com/report/mooc-stats-2019/

Shah, D. (2020a, May 2). How Different MOOC Providers are Responding to the Pandemic. *The Report by Class Central*. https://www.classcentral.com/report/mooc-providers-response-to-the-pandemic/

Shah, D. (2020b, November 30). By The Numbers: MOOCs in 2020. *The Report by Class Central*. https://www.classcentral.com/report/mooc-stats-2020/

Shah, D. (2020c, December 14). The Second Year of The MOOC: A Review of MOOC Stats and Trends in 2020. *The Report by Class Central*. https://www.classcentral.com/report/the-second-year-of-the-mooc/

Shah, D. (2021a, November 16). Coursera's Monetization Journey: From Zero to IPO. *The Report by Class Central*. https://www.classcentral.com/report/coursera-monetization-revenues/

Shah, D. (2021b, December 1). By The Numbers: MOOCs in 2021. *The Report by Class Central*. https://www.classcentral.com/report/mooc-stats-2021/

Shah, D. (2021c, December 14). A Decade of MOOCs: A Review of MOOC Stats and Trends in 2021. *The Report by Class Central*. https://www.classcentral.com/report/moocs-stats-and-trends-2021/

Shani, C., Zarecki, J., & Shahaf, D. (2023). The Lean Data Scientist: Recent Advances Toward Overcoming the Data Bottleneck. *Communications of the ACM*, *66*(2), 92–102. https://doi.org/10.1145/3551635

Sharif, M. S. A. M., & Ramakrisnan, P. (2023). Log Data Indicators for Identifying Learner Engagement in MOOCs. *International Journal of Advanced Research in Education and Society*, *5*(1), 35–51. https://doi.org/10.55057/ijares.2023.5.1.5

Shehata, M., Cheng, M., Leung, A., Tsuchiya, N., Wu, D. A., Tseng, C. H., Nakauchi, S., & Shimojo, S. (2021). Team Flow Is a Unique Brain State Associated with Enhanced Information Integration and Interbrain Synchrony. *eNeuro*, *8*(5), Article ENEURO.0133-21.2021. https://doi.org/10.1523/ENEURO.0133-21.2021

Shernoff, D. J., & Csíkszentmihályi, M. R. (2009). Cultivating engaged learners and optimal learning environments. *Handbook of Positive Psychology in Schools*, *131*, 145.

Shin, N. (2006). Online learner's 'flow' experience: An empirical study. *British Journal of Educational Technology*, *37*(5), 705–720. https://doi.org/10.1111/j.1467-8535.2006.00641.x

Shirmohammadi, S., & Al Osman, H. (2021). Machine Learning in Measurement Part 1: Error Contribution and Terminology Confusion. *IEEE Instrumentation & Measurement Magazine*, *24*(2), 84–92. https://doi.org/10.1109/MIM.2021.9400955

Siemens, G. (2005). Connectivism: A Learning Theory for the Digital Age. *International Journal of Instructional Technology and Distance Learning*, *2*(1). https://www.itdl.org/Journal/Jan_05/article01.htm

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Siemens, G. (2015). Foreword 1: The Role of MOOCs in the Future of Education. In C. J. Bonk, M. M. Lee, T. C. Reeves, & T. H. Reynolds (Eds.), *MOOCs and Open Education Around the World* (pp. xiii–xvii). Routledge.

Siemens, G. (2011, February 27). Message from the LAK 2011 General & Program Chairs. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*. Learning Analytics and Knowledge, Banff, Canada. https://doi.org/10.1145/2090116

Simon, R. (2007). 8—Resampling Strategies for Model Assessment and Selection. In W. Dubitzky, M. Granzow, & D. P. Berrar (Eds.), *Fundamentals of Data Mining in Genomics and Proteomics* (1st ed., pp. 173–186). Springer Science & Business Media. https://doi.org/10.1007/978-0-387-47509-7_8

Skadberg, Y. X., & Kimmel, J. R. (2004). Visitors' flow experience while browsing a Web site: Its measurement, contributing factors and consequences. *Computers in Human Behavior*, *20*(3), 403–422. https://doi.org/10.1016/S0747-5632(03)00050-5

Slouma, M., Ramírez, S., & Kaldmäe, K. (2019, January). *Analyse des parcours des apprenants du MOOC "La classe inversée à l'ère du numérique"* [Conference presentation]. Colloque International Éducation 4.1 !, Poitiers, France. https://hal.archives-ouvertes.fr/hal-02568645

Soffer, T., & Cohen, A. (2015). Implementation of Tel Aviv University MOOCs in academic curriculum: A pilot study. *International Review of Research in Open and Distributed Learning*, *16*(1), 80–97. https://doi.org/10.19173/irrodl.v16i1.2031

Stack Overflow. (2020). *Stack Overflow Developer Survey 2020* [Knowledge market; Question and answer]. Stack Overflow. https://insights.stackoverflow.com/survey/2020/?utm_source=social-share&utm_medium=social&utm_campaign=dev-survey-2020

Stack Overflow Developer Survey 2019. (2019). Stack Overflow. https://insights.stackoverflow.com/survey/2019/?utm_source=social-share&utm_medium=social&utm_campaign=dev-survey-2019

Stack Overflow Developer Survey 2021. (2021). Stack Overflow. https://insights.stackoverflow.com/survey/2021/?utm_source=social-share&utm_medium=social&utm_campaign=dev-survey-2021

Stack Overflow Developer Survey 2022. (2022). Stack Overflow. https://survey.stackoverflow.co/2022/?utm_source=social-share&utm_medium=social&utm_campaign=dev-survey-2022

Stack Overflow Developer Survey 2023. (2023). Stack Overflow. https://survey.stackoverflow.co/2023/?utm_source=social-share&utm_medium=social&utm_campaign=dev-survey-2023

Stamatelopoulou, F., Pezirkianidis, C., Karakasidou, E., Lakioti, A., & Stalikas, A. (2018). "Being in the Zone": A Systematic Review on the Relationship of Psychological Correlates and the Occurrence of Flow Experiences in Sports' Performance. *Psychology*, *09*(08), 2011–2030. https://doi.org/10.4236/psych.2018.98115

Stamper, J. C., Koedinger, K. R., Baker, R. S. d, Skogsholm, A., Leber, B., Demi, S., Yu, S., & Spencer, D. (2011). Managing the educational dataset lifecycle with datashop. *Artificial Intelligence in Education: 15th International Conference, AIED 2011, Auckland, New Zealand, June 28–July 2011 15*, 557–559.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, *42*(4), 73–93. https://doi.org/10.1111/j.1460-2466.1992.tb00812.x

Stoll, O., & Ufer, M. (2021). Flow in Sports and Exercise: A Historical Overview. *Advances in Flow Research*, 351–375. https://doi.org/10.1007/978-3-030-53468-4_13

Stracke, C. M., Downes, S., Conole, G., Burgos, D., & Nascimbeni, F. (2019). Are MOOCs Open Educational Resources? A Literature Review on History, Definitions and Typologies of OER and MOOCs. *Open Praxis*, *11*(4), 331–341. https://doi.org/10.5944/openpraxis.11.4.1010

Strubell, E., Ganesh, A., & McCallum, A. (2019). *Energy and Policy Considerations for Deep Learning in NLP* (Computation and Language (Cs.CL) arXiv:1906.02243 [cs.CL]). arXiv. https://doi.org/10.48550/arXiv.1906.02243

Subasi, A. (2020). Machine learning techniques. In *Practical Machine Learning for Data Analysis Using Python* (pp. 91–202). Elsevier. https://doi.org/10.1016/B978-0-12-821379-7.00003-5

Sugimura, P., & Hartl, F. (2018). *Building a Reproducible Machine Learning Pipeline* (arXiv:1810.04570). arXiv. https://doi.org/10.48550/arXiv.1810.04570

Suk, I., & Kang, E. (2007). Development and Validation of the Learning Flow Scale. *Journal of Educational Technology*, *23*(1), 121–154. https://doi.org/10.17232/KSET.23.1.121

Sun, Y., Ni, L., Zhao, Y., Shen, X.-L., & Wang, N. (2019). Understanding students' engagement in MOOCs: An integration of self-determination theory and theory of relationship quality. *British Journal of Educational Technology*, *50*(6), Article e0001. https://doi.org/10.1111/bjet.12724

Sunar, A. S., Abdullah, N. A., White, S., & Davis, H. C. (2015). Personalisation of MOOCs: The state of the art. *Proceedings of the 7th International Conference on Computer Supported Education - Volume 1: CSEDU*, *1*, 88–97. https://doi.org/10.5220/0005445200880097

Tadesse, S., & Muluye, W. (2020). The impact of COVID-19 pandemic on education system in developing countries: A review. *Open Journal of Social Sciences*, *8*(10), 159–170. https://doi.org/10.4236/jss.2020.810011

Tan, L., & Sin, H. X. (2021). Flow research in music contexts: A systematic literature review. *Musicae Scientiae*, *25*(4), 399–428. https://doi.org/10.1177/1029864919877564

The Royal Society. (2019). *Explainable AI: The basics*. https://royalsociety.org/topics-policy/projects/explainable-ai/

Thinker of the Year—2000: Mihaly Csikszentmihalyi. (2000). Brain Channels. https://www.brainchannels.com/thinker/mihaly.html

Thomas, L. B., Mastorides, S. M., Viswanadhan, N. A., Jakey, C. E., & Borkowski, A. A. (2021). Artificial Intelligence: Review of Current and Future Applications in Medicine. *Federal Practitioner: For the Health Care Professionals of the VA, DoD, and PHS*, *38*(11), 527—538. https://doi.org/10.12788/fp.0174

Torre, M. V., Tan, E., & Hauff, C. (2020). edX log data analysis made easy: Introducing ELAT: An open-source, privacy-aware and browser-based edX log data analysis tool. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 502–511. https://doi.org/10.1145/3375462.3375510

Tran, H. (2019). *Survey of Machine Learning and Data Mining Techniques used in Multimedia System* [Computer Science]. https://doi.org/10.13140/RG.2.2.20395.49446/1

Trevino, L. K., & Webster, J. (1992). Flow in computer-mediated communication: Electronic mail and voice mail evaluation and impacts. *Communication Research*, *19*(5), 539–573. https://doi.org/10.1177/009365092019005001

Tse, D. C., Nakamura, J., & Csíkszentmihályi, M. R. (2022). Flow experiences across adulthood: Preliminary findings on the continuity hypothesis. *Journal of Happiness Studies*, 1–24.

Tse, T. H. (1986). Integrating the structured analysis and design models: An initial algebra approach. *Australian Computer Journal*, *18*(3), 121–127.

Tse, T. H., & Pong, L. (1989). Towards a formal foundation for DeMarco data flow diagrams. *The Computer Journal*, *32*(1), 1–12. https://doi.org/10.1093/comjnl/32.1.1

Turan, N. (2019). Akış Deneyimi Üzerine Genel Bir Literatür Taraması. *Pamukkale University Journal of Social Sciences Institute*, *2019*(37), 181–199. https://doi.org/10.30794/pausbed.562564

Turner, J. C., & Patrick, H. (2008). How does motivation develop and why does it change? Reframing motivation research. *Educational Psychologist*, *43*(3), 119–131. https://doi.org/10.1080/00461520802178441

UNESCO. (2016a). *Education 2030: Incheon Declaration and Framework for Action for the implementation of Sustainable Development Goal 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000245656

UNESCO. (2016b). *Making sense of MOOCS: A guide for policy makers in developing countries* (M. Patru & V. Balaji, Eds.). UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000245122

University of Michigan, & Coursera, Inc. (2012). *Contract Between Coursera and University of Michigan*. https://www.documentcloud.org/documents/400864-coursera-fully-executed-agreement.html#document/p40

Valeye, F. (2022, May 16). Tracarbon—Track your device's carbon footprint: Measure your energy consumption and track your carbon emissions using your location. *Medium*. https://medium.com/@florian.valeye/tracarbon-track-your-devices-carbon-footprint-fb051fcc9009

van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine Learning*, *109*(2), 373–440. https://doi.org/10.1007/s10994-019-05855-6

Van Rossum, G., & Drake Jr, F. L. (1995). *Python Tutorial—Release 3.8.1* (Vol. 620). Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands.

van Veen, F., & Leijnen, S. (2019). *The Neural Network Zoo*. The Asimov Institute. https://www.asimovinstitute.org/neural-network-zoo/

Venn, J. (1881). Symbolic logic. *Nature*, *24*(613), 284–285. https://doi.org/10.1038/024284e0

Verdejo, M., & Celorrio, C. (2005). *DPULS-D32. 5.1, The design pattern language* [Technical report]. Action of The European Network of Excellence Kaleidoscope.

Verzat, C., & Bachelet, R. (2020). Le 1er Mooc français: Les coulisses d'une aventure intrapreneuriale. *Entreprendre & Innover*, *47*(4), 51–61. https://doi.org/10.3917/entin.047.0051

Voigt, P., & Von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR): A Practical Guide* (1st ed., Vol. 10). https://doi.org/10.1007/978-3-319-57959-7

Vrillon, É. (2017). Une typologie de MOOC de France Université Numérique (FUN). *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*, *24*(2), 1–27.

Wang, Y., & Baker, R. (2018). Grit and Intention: Why Do Learners Complete MOOCs? *The International Review of Research in Open and Distributed Learning*, *19*(3). https://doi.org/10.19173/irrodl.v19i3.3393

Waskom, M. L. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

Watted, A., & Barak, M. (2018). Motivating factors of MOOC completers: Comparing between university-affiliated students and general participants. *The Internet and Higher Education*, *37*, 11–20. https://doi.org/10.1016/j.iheduc.2017.12.001

Weber, M. (1922). *ÉCONOMIE ET SOCIÉTÉ 1. LES CATÉGORIES DE LA SOCIOLOGIE: Vol. I* (J. Chavy & É. de Dampierre, Eds.; J. Freund, P. Kamnitzer, P. Bertrand, É. de Dampierre, J. Chavy, & J. Maillard, Trans.; 1995th ed.). Pocket.

What is a MOOC? (2010, December 9). [YouTube]. https://youtu.be/eW3gMGqcZQc

Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality*, *59*(3), 339–354. https://doi.org/10.1111/j.1467-6494.1991.tb00252.x

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, *66*(5), 297–333. https://doi.org/10.1037/h0040934

Willems, E. P. (1969). Planning a rationale for naturalistic research. *Naturalistic Viewpoints in Psychological Research*, 44–71.

Xiong, Y., Ling, Q., & Li, X. (2021). Ubiquitous e-Teaching and e-Learning: China's Massive Adoption of Online Education and Launching MOOCs Internationally during the COVID-19 Outbreak. *Wireless Communications and Mobile Computing*, *2021*, Article e6358976. https://doi.org/10.1155/2021/6358976

Yarmey, A. D. (1979). *The psychology of eyewitness testimony*. Free Press New York.

Yousef, A. M. F., Chatti, M. A., Schroeder, U., & Wosnitza, M. (2014). What Drives a Successful MOOC? An Empirical Examination of Criteria to Assure Design Quality of MOOCs. *2014 IEEE 14th International Conference on Advanced Learning Technologies*, 44–48. https://doi.org/10.1109/ICALT.2014.23

Yousef, A. M. F., Chatti, M. A., Schroeder, U., Wosnitza, M., & Jakobs, H. (2014). MOOCs—A Review of the State-of-the-Art. *Proceedings of the 6th International Conference on Computer Supported Education*, 9–20. https://doi.org/10.5220/0004791400090020

Yuan, L., & Powell, S. J. (2013). *MOOCs and open education: Implications for higher education* [Report]. Cetis. https://www.cetis.org.uk/

Zhang, Y. (2013). Benefiting from MOOC. *EdMedia+ Innovate Learning*, 1372–1377.

Zhu, M., Sari, A. R., & Lee, M. M. (2020). A comprehensive systematic review of MOOC research: Research techniques, topics, and trends from 2009 to 2019. *Educational Technology Research and Development*, *68*(4), 1685–1710. https://doi.org/10.1007/s11423-020-09798-x

Footnotes

1 December 2019 – July 2020 in France but as of January 2023 still ongoing in some regions, *e.g.*, China.
2 Page numbering facilitates interested readers to locate and thus, corroborate the releveant passage.
3 Not to be confused nor associated with "in trance".
4 *e.g.*, doodling, humming, chewing gum, hair smoothing, finger tapping, smoking, etc. (Csíkszentmihályi, 1975a, p. 108).
5 Author uses dancing as an example: it can be started and stopped at any moment at will (without "grave" consequences), and songs are "usually short".
6 Pun intended.
7 Paradoxically, in the later findings of Csíkszentmihályi & LeFevre (1989), flow experiences were reported "when working, not when in leisure" although motivation was higher "in leisure than in work".
8 *vs.* exotelic activities: "activities done for external reasons only" (Csíkszentmihályi, 1990a, p. 67)
9 Also named "participant" or "person" by Csíkszentmihályi.
10 The models of autotelic personality (Baumann, 2012; Nakamura & Csíkszentmihályi, 2002; D. C. Tse et al., 2022) do not concern this thesis.
11 Flow researchers named it "conceptual flow model" or "theoretical flow model" but mostly simply employ the term "flow model" (Csíkszentmihályi, 1975a, 1975b) or "Flow Model" (initial uppercase). In this thesis, we follow the flow researchers' convention and, otherwise noted *e.g.*, Machine Learning flow-detecting model, we imply the term "conceptual flow model according to its authors" when referring to a "flow […] model". Also, we respect the authors' original uppercase writing when citing primary sources, *e.g.*, "Model of the Flow State".
12 *A.k.a.* Flow Theory, a subject out of the scope of this thesis.
13 *A.k.a.* elements, components, characteristics, dimensions, features, traits, …
14 "Negative entropy or a state of order"(Csíkszentmihályi & Csíkszentmihályi, 1988).
15 Although such definition is not found in that specific citation, the view of flow as an "optimal situation" did appear in that book, and it is described as "when the challenges match skills" (Csíkszentmihályi, 1975a, p. 66).
16 This first, quick definition, followed by the already-familiar definition, might obey to the non-academic nature (according to the author) and vulgarization intent behind the book.
17 In this context, a more appropriate term would be 'resolution'.
18 The "Cusp catastrophe model of flow" is not shown here because of its complex representation.
19 The minor difference between both definitions is underlined.
20 https://efrn.eu/
21 While the definition correctly attributes flow fatherhood to Csíkszentmihályi, Hungarian diacritics are missing in the original 2014 source quoted here.
22 Leaving room for psychological and physiological studies of flow.
23 Replacing initial recurrent terms such as 'enjoyment' and 'pleasure', *cf.* Section "Antecedents of Flow".
24 At least firstly.
25 Also, "demands".
26 Other terms found in the literature are "dimension" or "feature".
27 *Cf.* Abuhamdeh discusses on the relationship between flow and enjoyment (2021b).
28 Creative endeavors, like painting or music composing, rely on the individual's own sense of intention to set and recognize goals and feedback indicators among the initial vague goals and gauges of feedback.
29 *Cf.* R. G. Mitchell (1988) for a more comprehensive literature review on these two concepts and how they relate to flow.
30 Risk-taking model of Atkinson (1958; cited by Engeser & Rheinberg, 2008).
31 The Experience Sampling Method relies on the participant answering a form at random moments of the day prompted by a "beeper".
32 *Cf.* (de Moura Jr & Porto Bellini, 2019; Pels et al., 2018; Tan & Sin, 2021) for field-specific literature reviews of flow measurement in music, work, and social flow, and/or (Rosas et al., 2023) for a compiled list of 34 validated flow measurement instruments in English (Non-ESM scales).
33 Also surveyed by Nakamura & Csíkszentmihályi (2009, p. 198)
34 This contrasts the contents of the Peifer & Engeser (2021a) publication, where the subject is not addressed by any of the contributors.

[35] These percentages are relative to those in education (94) and not to the total (256).
[36] Other 4% might include mixed domain articles, *i.e.*, on something else **and** education, rounding error < 2%
[37] An extended definition on "Learner Model" is proposed in a litterature review by this thesis author (Ramírez Luelmo et al., 2020a, sec. 2.1): Learner Models represent the system's beliefs about the learner's specific characteristics, relevant to the educational practice (Giannandrea & Sansoni, 2013), they are usually enriched by data collection techniques (Nguyen & Do, 2008) and they aim to encode individual learners using a specific set of dimensions (Nakić et al., 2015).
[38] Some authors consider motivation and engagement as separate psychological states as well (Abyaa et al., 2019).
[39] [In both studies this assumption was confirmed. Even when controlling for performance-relevant competence factors, flow experience during the learning phase predicted later learning performance. […] What is remarkable about the replicated finding is that although the studies were both conducted in university contexts, they involved quite different settings. […] Despite this difference, both studies showed the expected relationships between motivation, flow, and learning performance].
[40] Presence and Telepresence are "the sense of being in an environment, generated by natural or mediated means, respectively" (Steuer, 1992, p. 3).
[41] Some studies also highlight the importance of collective (or social) flow in this process.
[42] Rudely oversimplified here, as the Flow experience encompasses proximal conditions and the Flow state. Indeed, Fulfillment of Proximal conditions lead to the Flow state.
[43] Linked to Social Cognitive Theory (Bandura, 2001).
[44] Please notice that most of them were not originally designed with that specific context in mind but were nonetheless employed for that purpose.
[45] Cronbach's α coefficient.
[46] In Spain.
[47] Originally called "Signaling device" (Csíkszentmihályi & Larson, 2014, p. 37 originally published in 1987).
[48] Although not specifically designed nor for flow detection nor measurement but historically employed as such.
[49] At the time of writing this thesis, the article with the most updated version of this scale (2017, High School students) accounts only one citation in the Google Scholar search engine, which might be an indication it has been abandoned.
[50] Nor Spanish, French, Portuguese, or German; the languages the author of this thesis is most comfortable understanding.
[51] That is the "paradox of [flow at] work" (Heutte, 2020), which is resumed as: « *on fait pas ce que l'on veut au travail, pourtant, si l'on réussi, ça deviant plaisant* », [at work, we don't do what we want although if we succeed it becomes enjoyable].
[52] Based on the Flow in Education theoretical model quickly overviewed at the end of Section "Flow in Educational Contexts".
[53] EduFlow is known to have been employed as a basis for the creation of a flow scale in physical education in Arabic as well (Abbassi et al., 2021).
[54] More recently, EduFlow is available in Arabic as well (Chalghaf et al., 2019).
[55] Other than Internet access and the participant's own interest.
[56] *Environnements Informatiques pour l'Apprentissage Humain* (EIAH) (Balachef, 2018, p. 66).
[57] Originally [http://ltc.umanitoba.ca/connectivism/] but now a defunct link.
[58] https://moodle.org/
[59] https://www.udacity.com/course/cs271
[60] Not accounting for China, whose metrics might be unreliable (Shah, 2021c).
[61] Based on Alexa Ranking, a now-defunct web service since May 1st, 2022.
[62] *Cf.* Relevant recent work on MOOC taxonomies (Blackmon & Major, 2017; Pilli & Admiraal, 2016; Rosselle et al., 2014; Stracke et al., 2019) were consulted while researching for this Chapter.
[63] Massachusetts Institute of Technology; a private research university in Cambridge, U.S.A. (MIT, n.d.)
[64] Four pairs of complementary forms determined by their positive *vs.* negative values linked on the corresponding dimension.
[65] Authors clarify that variables length and certification do not primarily contribute to their proposal.
[66] Meaning that a hypothetical second dog, next to the first, equally competent, and subject to the same sampling, might not create the same identical smell profile.
[67] Male or female, old or young, healthy, or unhealthy, happy, sour, acrid, sweaty, or rancid, etc. for a lack of a better way to convey to the reader the specter of options available to the canine to detect and choose from.

68 The dog has not met all the possible types of humans on earth as to know what characteristics define a human smell profile in its entirety. However, the dog knows what characteristics define a human smell for **most** of the humans the dog has smelled before.
69 Or the police, in this analogy.
70 More specifically, a mini-Neural Network ("highly organized network"), limited at the time per the "data-handling ability and […]computational speed" (Samuel, 1959, p. 211).
71 This definition is widely attributed (over a thousand results on a strict text query in Google Scholar) to Arthur Lee Samuel as part of his 1959 paper (Samuel, 1959). However, we could not find this quote nor in the original 1959 paper (Samuel, 1959), nor in its official 1963 reprint (Samuel, 1963), nor in its second part ("II-Recent Progress") published a decade later (Samuel, 1969).
72 https://chat.openai.com/chat ; no citation knowingly exists at the time of querying the model.
73 Flawed input data will produce flawed or nonsensical output.
74 Undersampling, Oversampling, and Hybrid methods (Kulkarni et al., 2020, p. 8).
75 Other standard tasks include anomaly detection, structured annotation, translation, density estimation, etc.
76 Authors specifically point out "Generative AI capabilities and applications" (Chien, 2023, p. 5).
77 https://github.com/responsibleproblemsolving/energy-usage
78 https://github.com/sb-ai-lab/Eco2AI
79 https://github.com/fvaleye/tracarbon
80 Different world regions generate electricity in different enegy mixes, with wildly different proportion of fossil-fuels-based primary sources.
81 https://github.com/mlco2/codecarbon
82 "[…] ML systems are often used for the most complex and ill-defined tasks—if the tasks were easy, we would not need an ML solution" (Isbell et al., 2023, p. 37).
83 *Cf.* https://scikit-learn.org/stable/glossary.html for a comprehensive glossary of terms and API elements.
84 Very recent research suggest the creation of "tools and research advances that will allow pretrained models to be built in the same way that we build open source software" to approach this phenomenon (Raffel, 2023, p. 38).
85 Mostly in terms of computational resource allocation.
86 *i.e.*, the probability of an event occurring.
87 A quite complete and updated chart of existing Neural Networks was made available by van Veen & Leijnen (2019).
88 Neither of which we cover in this thesis.
89 A perfect classifier is considered a theoretical construct and thus, its appearance in real-life scenarios is likely an indication of a modeling issue.
90 These could be alternatively reconsidered as caveats in specific model cases where invariance is necessary.
91 Please notice that the ROC AUC is to be in the upper-LEFT-hand whereas the PR AUC is to be in the upper-RIGHT-hand corner.
92 Repeated *k*-fold, Leave One Out, Leave P Out, Shuffle & Split, Group *k*-fold, Leave One Group Out, etc. (Pedregosa et al., 2011, sec. 3.1).
93 https://thedre.imag.fr/
94 RHCCS aims to consider people and their IT environment, whether it is for work (in the information system domain), in a learning environment (in the learning domain), or simply to account for interactions with a given machines (in a human-computer interaction context) (Mandran & Dupuy-Chessa, 2017, p. 1).
95 Please notice that the term "indicator" within the THEDRE method is vastly different to the meaning employed in Chapter 6 above, "Trace analysis in a MOOC".
96 The University's academic access performs a security ban on accounts attempting to mass download articles within a short period of time. A few hours' wait time to lift the temporary ban ensues.
97 https://www.limesurvey.org/
98 http://iramuteq.org/
99 https://www.r-project.org/
100 https://www.python.org/
101 https://www.ibm.com/products/spss-statistics
102 https://www.tableau.com/products/desktop
103 https://harzing.com/resources/publish-or-perish/
104 https://www.libreoffice.org/discover/calc/
105 https://www.apple.com/numbers/
106 https://www.microsoft.com/en-us/microsoft-365/excel

107 https://www.mendeley.com/download-reference-manager
108 https://www.zotero.org/
109 https://www.microsoft.com/en-us/microsoft-365/word
110 *Cf.* The EU GDPR Practical Guide (Voigt & Von dem Bussche, 2017).
111 *Cf.* Practices of Human-Centered Machine Learning (Chancellor, 2023, p. 81).
112 *Cf.* The Regulatory framework proposal on artificial intelligence (Artificial Intelligence Act, 2021; European Union, 2022).
113 https://edx.readthedocs.io/projects/edx-installing-configuring-and-running/en/latest/installation/index.html
114 *« L'analyse des traces d'apprentissages »* (Cherigny et al., 2020).
115 Unnamed in the original source (2013, p. 38).
116 The "*Habilitation à Diriger des Recherches*" (HDR), "Habilitation to Conduct Research" or "Accreditation to Supervise Research" is the highest French education qualification diploma, obtained after a PhD (Aix Marseille Université, 2023; SciencesPo, 2018).
117 In our specific context, it is not to be confused with the most common notion (dating back to the mid 1950's) in Computer Science of data type: "a small list of possible types supported by the [programming] language" (Parnas et al., 1976, p. 149), implying a collection of possible data values, their allowed operations and expected behavior, and their most basic, internal machine representation (primitive). In such definition, a variable would be a primitive, and data types would be "equivalence classes of variables" (1976, p. 149).
118 In a UML class diagram, arrows connecting elements (B →A) are to be read as "B inherits from A", and not as a sequence.
119 *Cf.* also "Multimodal learning" (Shani et al., 2023).
120 Maybe even illegal activities.
121 https://ecole.centralelille.fr/
122 https://mooc.gestiondeprojet.pm/
123 http://bit.ly/2LozEFI (URL shortening provided and maintained by the MOOC GdP staff)
124 https://www.study.eu/article/what-is-the-ects-european-credit-transfer-and-accumulation-system
125 https://systemcheck.rpexams.com/
126 https://www.youtube.com/
127 https://moocit.fr/
128 https://openedx.org/
129 https://aws.amazon.com/fr
130 https://docs.openedx.org/en/latest/
131 https://edx.readthedocs.io/projects/devdata/en/latest/index.html
132 https://edx.readthedocs.io/projects/devdata/en/latest/internal_data_formats/tracking_logs/index.html
133 https://edx.readthedocs.io/projects/edx-developer-guide/en/latest/analytics.html
134 https://www.mysql.com/
135 https://www.json.org/json-en.html
136 As in range, spectrum.
137 The existing and validated data of the flow measurement instruments during the Proof-of-Concept subprocess is employed again.
138 https://www.libarchive.org/
139 https://www.gnu.org/software/gzip/
140 https://www.mongodb.com/try/download/community
141 https://www.mongodb.com/docs/manual/core/document/
142 DFD notation preferred over flowcharts or UML activity diagrams because of their simplicity.
143 https://pypi.org/project/pandas-profiling/
144 https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression & https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
145 Python has shown to be one of the two most employed languages in data science (Stack Overflow, 2020), constantly raising in usage from 2015 to 2020, in 2019 being the "fastest-growing major language" (*Stack Overflow Developer Survey 2019*, 2019), all-in-all positiong itself among the top three most commonly-used programming languages (*Stack Overflow Developer Survey 2021*, 2021; *Stack Overflow Developer Survey 2022*, 2022; *Stack Overflow Developer Survey 2023*, 2023).
146 https://joblib.readthedocs.io/
147 Please notice that differentiate in this Section the meaning of "session", as the time period internally determined by the Open edX platform that starts with a user login, from a GdP "Session" (initial uppercase) employed so far.
148 https://edx.readthedocs.io/projects/devdata/en/stable/internal_data_formats/tracking_logs.html

[149] Seconds elapsed since 00:00:00 UTC on 1 January 1970, the beginning of the Unix epoch.
[150] « *épaisseur* », in the original text.
[151] A violin plot shows the distribution of quantitative data across several levels of one (or more) categorical variables. The violin plot features a kernel density estimation of the underlying distribution (Hunter, 2007; Waskom, 2021).
[152] All CONCAT operations require an appropriate separator ("_" or "/" or "*", depending on the content of the original concatenated strings.
[153] *Cf.* the curse of dimensionality.
[154] While accounting for target data imbalance.
[155] Most imbalanced datasets commonly showcase reduced examples of the positive class, *e.g.*, during fraud detection, fraudulent transactions tend to be orders of magnitude smaller than usual transactions.
[156] "Answer to the Ultimate Question of Life, The Universe, and Everything" (Adams, 2005).
[157] Accessing, retrieving, and aggregating the necessary data for features' creation takes orders of magnitude longer than the milliseconds the model requires to yield a flow label.
[158] Besides the mandatory stratified data 70/30 split.
[159] A reduced number of incompatible combination of hyperparameters leads to slightly less models being calculated.
[160] 10-fold Cross Validation results of a Logistic regression classifier employed to predict flow in a video game settings, trained on qualitative (survey) and quantitative data (game logs) (Moon et al., 2022).
[161] Assuming the default scorer for a Logistic regression classifier was employed (Accuracy).
[162] No academic data was employed as we do not focus on the relationship between flow and (academic) performance (*cf.* Section "A Brief Account").
[163] Which would imply a flow measurement instrument to ask participants if they plan to be in flow in the future.
[164] Deep learning is another subset of Artificial Intelligence.
[165] "Roomba is a series of autonomous robotic vacuum cleaners made by the company iRobot" ("Roomba," 2023).
[166] *Cf.* the "Data Bottleneck" (Shani et al., 2023).

Published articles

Here, we present the accepted and published articles at national and international conferences. Two papers were among the finalists for the "Best doctoral paper" award at the CSEdu Conference, in Prague, Czech Republic, of which the paper titled "Towards a Machine Learning Flow-detecting Model in a MOOC Context" obtained it in 2022. Furthermore, a few of them were considered noteworthy by the conferences' organizers to become part of a larger publication and thus their extended version saw the light as book chapters, listed at the end of this Chapter.

Communications in International Conferences

Ramírez Luelmo, S. I., El Mawas, N., Bachelet, R., & Heutte, J. (2022). Towards a Machine Learning Flow-predicting Model in a MOOC Context. *Proceedings of the 14th International Conference on Computer Supported Education*, 124–134. https://doi.org/10.5220/0011070300003182

Ramírez Luelmo, S. I., El Mawas, N., & Heutte, J. (2021). Machine Learning Techniques for Knowledge Tracing: A Systematic Literature Review. *Proceedings of the 13th International Conference on Computer Supported Education*, *1*, 60–70. https://doi.org/10.5220/0010515500600070

Ramírez Luelmo, S. I., El Mawas, N., & Heutte, J. (2020). A literature review on Learner Models for MOOC to support Lifelong Learning. *Proceedings of the 12th International Conference on Computer Supported Education (CSEDU 2020)*, *1*, 527–539. https://doi.org/10.5220/0009782005270539

Ramírez Luelmo, S. I., El Mawas, N., & Heutte, J. (2020). Towards Open Learner Models Including the Flow State. *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 305–310. https://doi.org/10.1145/3386392.3399295

Communications in National Conferences

Ramírez Luelmo, S. I. (2022). Vers une prédiction semi-automatique du flow dans un MOOC. In C. Bonnat & R. Venant (Eds.), *Actes des neuvièmes rencontres jeunes chercheur·e·s en EIAH* (pp. 126–133). Lille, France: ATIEF. Retrieved from https://rjc-eiah-2022.univ-lille.fr/fileadmin/user_upload/rjc-eiah-2022/documents/acteRJCEIAH2022_v1.1.pdf

Ramírez Luelmo, S. I., El Mawas, N., & Heutte, J. (2021). Les modèles apprenant pour soutenir l'apprentissage tout au long de la vie: Revue de littérature. In M. Lefevre, C. Michel, T. Geoffre, M. Rodi, L. Alvarez, & A. Karoui (Eds.), *10e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain* (pp. 200–211). Fribourg, Switzerland: ATIEF. Retrieved from https://hal.archives-ouvertes.fr/hal-03292891

Ramírez Luelmo, S. I. (2020). Vers une modélisation de l'expérience optimale d'apprentissage via les Learning Analytics. In A. Yessad, S. Jolivet, & C. Michel (Eds.), *8èmes RJC EIAH 2020: Environnements Informatiques pour l'Apprentissage Humain* (pp. 144–149). Poitiers, France: ATIEF. Retrieved from https://rjceiah20.conference.univ-poitiers.fr/wp-content/uploads/sites/406/2020/05/RAMIREZ_texte_poster.pdf

Book chapters

Ramírez Luelmo, S. I., El Mawas, N., & Heutte, J. (2022). Existing Machine Learning Techniques for Knowledge Tracing: A Review Using the PRISMA Guidelines. In B. Csapó & J. Uhomoibhi (Eds.), *International Conference on Computer Supported Education* (Vol. 1624, pp. 73–94). Springer International Publishing. https://doi.org/10.1007/978-3-031-14756-2_5

Ramírez Luelmo, S. I., El Mawas, N., & Heutte, J. (2021). Learner Models for MOOC in a Lifelong Learning Context: A Systematic Literature Review. In H. C. Lane, S. Zvacek, & J. Uhomoibhi (Eds.), *Computer Supported Education* (Vol. 1473, pp. 392–415). Springer International Publishing. https://doi.org/10.1007/978-3-030-86439-2_20

Submitted book chapter, decision pending

Ramírez Luelmo, S. I., El Mawas, N., Bachelet, R., & Heutte, J. (2022). Detection and asynchronous prediction of flow in MOOC. SN Computer Science

Other communications

Ramírez, S., El Mawas, N., & Heutte, J. (2022). Optimal experience modelling: Detection via Learning Analytics in a Lifelong Learning context. *10th European Conference on Positive Psychology (ECPP 2022).* Presented at the 10th European Conference on Positive Psychology (ECPP 2022), Reykjavik, Iceland. https://hal.science/hal-02499272

Heutte, J., Ramírez Luelmo, S. I., El Mawas, N., & LasVergnas, O. (2022). Lexicometrical analysis method to support Scoping Review on social dimensions of flow. *10th European Conference on Positive Psychology (ECPP 2022).* Presented at the 10th European Conference on Positive Psychology (ECPP 2022), Reykjavik, Iceland. https://hal.science/hal-02499145

Ramírez Luelmo, S.I., El Mawas, N., Bachelet, R., Fenouillet, F., & Heutte, J. (2022). Prédire l'Expérience Autotélique des participants à un MOOC. Vers une implémentation dans un Tableau de Bord. *9e Colloque international en éducation,* Montréal, Québec, Canada.

Ramírez Luelmo, S.I., El Mawas, N., & Heutte, J. (2021). Machine Learning for Knowledge Tracing in Learner Models. *8e Colloque international en éducation,* Montréal, Québec, Canada.

Ramírez Luelmo, S.I., El Mawas, N., & Heutte, J. (2020). Learner Models in MOOCs in a Lifelong Learning perspective. *7e Colloque international en éducation,* Montréal, Québec, Canada.

Ramírez Luelmo, S. I. (2021, November). *Flow & Learning Analytics.* Poster presented at the "From learning to classroom analytics": 1st international workshop of the TELS chair, Villeneuve-d'Ascq, France. Villeneuve-d'Ascq, France. http://www.isite-ulne.fr/index.php/en/2021/10/15/from-learning-to-classroom-analytics-1st-international-workshop-of-the-tels-chair/

Ramírez, S., El Mawas, N., & Heutte, J. (2022). Advances on a Machine Learning flow prediction model in a MOOC. *11th Annual European Flow Researchers Network Meeting (EFRN 2022)*. Presented at the 11th Annual European Flow Researchers Network Meeting (EFRN 2022), Lille, France.

Ramírez, S., El Mawas, N., & Heutte, J. (2021). Towards a Flow prediction model in an educational, online context. *10th Annual European Flow Researchers Network Meeting (EFRN 2021)*. Presented at the 10th Annual European Flow Researchers Network Meeting (EFRN 2021), Lübeck, Germany.

Ramírez, S., El Mawas, N., & Heutte, J. (2021). Flow detection in a MOOC environment. *10th Annual European Flow Researchers Network Meeting (EFRN 2021)*. Presented at the 10th Annual European Flow Researchers Network Meeting (EFRN 2021), Lübeck, Germany.

Ramírez, S., El Mawas, N., & Heutte, J. (2020). Flow Learning Modeling via Learning Analytics. *9th Annual European Flow Researchers Network Meeting (EFRN 2020)*. Presented at the 9th Annual European Flow Researchers Network Meeting (EFRN 2020), Milano, Italia.

Finally, other publications are in the works, of which one detailing a method for the score calculation for the EduFlow-2 measure instrument is eagerly expected, but also on Social Flow, and on Machine Learning for Learning Analytics.

Table index

Figures index

Appendices

Appendix 1. – The Experience Sampling Form

Date: ________ Time Beeped: __________ am/pm Time Filled Out__ am/pm

As you were beeped... What were you thinking about?

____

Where were you?

____

What was the MAIN thing you were doing?

____

What other things were you doing?

____

WHY were you doing this particular activity?

( ) I had to ( ) I wanted to do it ( ) I had nothing else to do

	Not at all	**Some what**	**Quite**	**Very**
How well were you concentrating?	0	1 2 3	4 5 6	7 8 9
Was it hard to concentrate?	0	1 2 3	4 5 6	7 8 9
How self-conscious were you?	0	1 2 3	4 5 6	7 8 9
Did you feel good about yourself?	0	1 2 3	4 5 6	7 8 9
Were you in control of the situation?	0	1 2 3	4 5 6	7 8 9
Were you living up to your own expectations?	0	1 2 3	4 5 6	7 8 9
Were you living up to expectations of others?	0	1 2 3	4 5 6	7 8 9

Describe your mood as you were beeped:

	Very	**Quite**	**Some**	**Neither**	**Some**	**Quite**	**Very**	
Alert	0	o	.	-	.	o	0	Drowsy
Happy	0	o	.	-	.	o	0	Sad
Irritable	0	o	.	-	.	o	0	Cheerful
Strong	0	o	.	-	.	o	0	Weak
Active	0	o	.	-	.	o	0	Passive
Lonely	0	o	.	-	.	o	0	Sociable
Ashamed	0	o	.	-	.	o	0	Proud
Involved	0	o	.	-	.	o	0	Detached
Excited	0	o	.	-	.	o	0	Bored
Closer	0	o	.	-	.	o	0	Open

Clear	0	o	.	-	.	o	0	Confused
Tense	0	o	.	-	.	o	0	Relaxed
Competitive	0	o	.	-	.	o	0	Cooperative

Did you feel any physical discomfort as you were beeped:

Overall pain or discomfort	**none**				**slight**				**bothersome**				**severe**
	0	1	2	3	4	5	6	7	8	9			

Please

specify: ____

Who were you with?

( ) Alone	( ) Friend(s)	How many?
( ) Mother		_______________
	Female ( )	
( ) Father	( ) Strangers	
( ) Sister(s) of brother(s)	( ) Other _______________	

Indicate how you felt about your activity:

	low		**high**
Challenges of the activity	0	1 2 3 4 5 6 7 8 9	
Your skills in the activity	0	1 2 3 4 5 6 7 8 9	
	not at all		very much
Was this activity important to you?	0	1 2 3 4 5 6 7 8 9	
Was this activity important to others?	0	1 2 3 4 5 6 7 8 9	
Were you succeeding at what you were doing?	0	1 2 3 4 5 6 7 8 9	
Do you wish you had been doing something else?	0	1 2 3 4 5 6 7 8 9	
Were you satisfied with how you were doing?	0	1 2 3 4 5 6 7 8 9	
How important was this activity in relation to your overall goals?	0	1 2 3 4 5 6 7 8 9	
If you had a choice			
Who would you be with? ____			
What would you be doing? ____			

Since you were last beeped has anything happened or have you done anything which could have affected the way you feel? Nasty cracks, comments, etc.

(Csíkszentmihályi, 2014, pp. 50–51 originally published in 1987)

Appendix 2. – The FlowQuestionnaire

(Csíkszentmihályi, 1975a; Csíkszentmihályi & Csíkszentmihályi, 1988; S. Han, 1988)

1. My mind isn't wandering. I am not thinking of something else. I am totally involved in what I am doing. My body feels good. I don't seem to hear anything. The world seems to be cut off from me. I am less aware of myself and my problems.

2. My concentration is like breathing. I never think of it. I am really quite oblivious to my surroundings after I really get going. I think that the phone could ring, and the doorbell could ring, or the house burn down or something like that. When I start, I really do shut out the whole world. Once I stop, I can let it back in again.

3. I am so involved in what I am doing. I don't see myself as separate from what I am doing.

(Csíkszentmihályi & Csíkszentmihályi, 1988; translated by Heutte, 2015)

1. Merci de bien vouloir lire les 3 citations suivantes :

    c. Mon esprit ne vagabonde pas. Je ne pense à rien d'autre. Je suis totalement impliqué•e dans ce que je fais. Je me sens bien dans mon corps. J'ai l'impression de ne plus rien entendre. J'ai l'impression d'être coupé•e du monde. Je n'ai plus vraiment conscience de moi-même ni de mes problèmes.
    d. Ma concentration est comme ma respiration, je n'ai pas besoin d'y penser. Une fois que j'ai commencé je ne suis plus vraiment tout à fait conscient•e de mon environnement. Je ne pense pas que le téléphone pourrait sonner, que quelqu'un viendrait sonner à la porte, que la maison pourrait brûler ou quelque chose comme ça. Quand je commence, je me coupe vraiment du monde. Ce n'est que lorsque je m'arrête que je peux tout laisser revenir.
    e. Je suis tellement impliqué•e que je ne fais qu'un avec ce que je fais.

2. Avez-vous déjà vécu une expérience qui corresponde à une ou plusieurs de ces citations ?

3. Dans l'affirmative, quelles activités étiez-vous en train de pratiquer lorsque vous avez ressenti cette sensation ?

4. Merci d'indiquer le nom de l'activité – parmi celles que vous avez cité précédemment, le cas échéant – qui correspond le mieux aux sensations décrites dans les 3 citations, c'est-à-dire celle où l'expérience a été la plus intense.

(Csíkszentmihályi & Csíkszentmihályi, 1988, p. 195; employed by Moneta, 2021, p. 3)

1. Please read the following quotes:
2. My mind isn't wandering. I am not thinking of something else. I am totally involved in what I am doing. My body feels good. I don't seem to hear anything. The world seems to be cut off from me. I am less aware of myself and my problems.
3. My concentration is like breathing I never think of it. When I start, I really do shut out the world. I am really quite oblivious to my surroundings after I really get going. I think that the phone could ring, and the doorbell could ring or the house burn down or something like that. When I start I really do shut out the world. Once I stop I can let it back in again.
4. I am so involved in what I am doing. I don't see myself as separate from what I am doing.
5. Have you ever felt similar experiences?
6. If yes, what activities were you engaged in when you had such experiences?
7. Please write here the name of the activity - among those you quoted, if any - which best represents the experience described in the three quotations, *i.e.* the activity where you feel this experience with the highest intensity.
8. On the next pages there are a number of items referring to the ways people could feel while doing an activity (e.g. ratings on the activity quoted in section 4, work or study, or spending time with the family). For each item please tell us how you feel doing each of these activities.

(Asakawa, 2010, app. Flow Quotations)

1. Do you ever do something where your concentration is so intense, your attention so undivided and wrapped up in what you are doing that you sometimes become unaware of things you normally notice (for instance, other people talking, loud noises, the passage of time, being hungry or tired, having an appointment, having some physical discomfort)?

2. Do you ever do something where your skills have become so "second nature" that sometimes everything seems to come to you "naturally" or "effortlessly," and where you feel confident that you will be ready to meet any new challenges?

3. Do you ever do something where you feel that the activity is worth doing in itself? In other words, even if there were no other benefits associated with it (for instance, financial reward, improved skills, recognition from others, and so on), you would still do it?

Appendix 3. – Die Flow-Kurzskala

(Rheinberg et al., 2003)

Flow-Kurzfragebogen entwickelt von Falko Rheinberg & Regina Vollmeyer, Universität Potsdam

Alter: ☐ Jahre

Geschlecht: ☐ weiblich
☐ männlich

Ich mache gerade ____ Uhrzeit _____________
(Beziehen Sie sich bitte auf die eben unterbrochene Tätigkeit.)

	Trifft nicht zu	teils-teils	Trifft zu
• **Ich fühle mich optimal beansprucht.**			
• **Meine Gedanken bzw. Aktivitäten laufen flüssig und glatt.**			
• **Ich merke gar nicht, wie die Zeit vergeht.**			
• **Ich habe keine Mühe, mich zu konzentrieren.**			
• **Mein Kopf ist völlig klar.**			
• **Ich bin ganz vertieft in das, was ich gerade mache.**			
• **Die richtigen Gedanken/Bewegungen kommen wie von selbst.**			
• **Ich weiß bei jedem Schritt, was ich zu tun habe.**			
• **Ich habe das Gefühl, den Ablauf unter Kontrolle zu haben.**			
• **Ich bin völlig selbstvergessen.**			
• **Es steht etwas für mich Wichtiges auf dem Spiel.**			
• **Ich darf jetzt keine Fehler machen.**			
• **Ich mache mir Sorgen über einen Misserfolg.**			

• **Verglichen mit allen anderen Tätigkeiten, die ich sonst mache, ist die jetzige Tätigkeit...**	leicht		schwer
• **Ich denke, meine Fähigkeiten auf diesem Gebiet sind...**	niedrig		hoch
• **Für mich persönlich sind die jetzigen Anforderungen...**	zu gering	gerade richtig	zu hoch

(Engeser & Rheinberg, 2008, app. Flow Short Scale)

	Not at all			Partly			Very much
I feel just the right amount of challenge.	○	○	○	○	○	○	○
My thoughts/activities run fluidly and smoothly.	○	○	○	○	○	○	○
I don't notice time passing.	○	○	○	○	○	○	○
I have no difficulty concentrating.	○	○	○	○	○	○	○
My mind is completely clear.	○	○	○	○	○	○	○
I am totally absorbed in what I am doing.	○	○	○	○	○	○	○
The right thoughts/movements occur of their own accord.	○	○	○	○	○	○	○
I know what I have to do each step of the way.	○	○	○	○	○	○	○
I feel that I have everything under control.	○	○	○	○	○	○	○
I am completely lost in thought.	○	○	○	○	○	○	○

	easy								**difficult**
Compared to all other activities which I partake in, this one is ...	○	○	○	○	○	○	○	○	○
	low						high		
I think that my competence in this area is ...	○	○	○	○	○	○	○	○	○
	too low			just right			too high		
For me personally, the current demands are ...	○	○	○	○	○	○	○	○	○

(Rheinberg & Engeser, 2018, p. 607)

	disagree						**agree**
I feel just the right amount of challenge.							
My thoughts/activities run fluidly and smoothly.							
I don't notice time passing.							
I have no difficulty concentrating.							
My mind is completely clear.							
I am totally absorbed in what I am doing.							
The right thoughts/movements occur of their own accord.							
I know what I have to do each step of the way.							
I feel that I have everything under control.							
I am completely lost in thought.							

	disagree								**agree**
Something important to me is at stake.									
I mustn't make any mistakes.									
I am worried about failure.									

Appendix 4. – The Flow State Scale (FSS) and the Dispositional Flow Scale (DFS)

(Jackson & Marsh, 1996, app. Flow State Scale)

Please answer the following questions in relation to your experience in the event you have just completed. These questions relate to the thoughts and feelings you may have experienced during the event. There are no right or wrong answers. Think about how you felt during the event and answer the questions using the rating scale below. Circle the number that best matches your experience from the options to the right of each question.

Rating Scale.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1	2	3	4	5

		1	2	3	4	5
1.	I was challenged, but I believed my skills would allow me to meet the challenge.					
2.	I made the correct movements without thinking about trying to do so.					
3.	I knew clearly what I wanted to do.					
4.	It was really clear to me that I was doing well.					
5.	My attention was focused entirely on what I was doing.					
6.	I felt in total control of what I was doing.					
7.	I was not concerned with what others may have been thinking of me.					
8.	Time seemed to alter (either slowed down or speeded up).					
9.	I really enjoyed the experience.					
10.	My abilities matched the high challenge of the situation.					
11.	Things just seemed to be happening automatically.					
12.	I had a strong sense of what I wanted to do.					
13.	I was aware of how well I was performing.					
14.	It was no effort to keep my mind on what was happening.					
15.	I felt like I could control what I was doing.					
16.	I was not worried about my performance during the event.					
17.	The way time passed seemed to be different from normal.					
18.	I loved the feeling of that performance and want to capture it again.					
19.	I felt I was competent enough to meet the high demands of the situation.					
20.	I performed automatically.					
21.	I knew what I wanted to achieve.					
22.	I had a good idea while I was performing about how well I was doing.					
23.	I had total concentration.					
24.	I had a feeling of total control.					

25.	I was not concerned with how I was presenting myself.
26.	It felt like time stopped while I was performing.
27.	The experience left me feeling great.
28.	The challenge and my skills were at an equally high level.
29.	I did things spontaneously and automatically without having to think.
30.	My goals were clearly defined.
31.	I could tell by the way I was performing how well I was doing.
32.	I was completely focused on the task at hand.
33.	I felt in total control of my body.
34.	I was not worried about what others may have been thinking of me.
35.	At times, it almost seemed like things were happening in slow motion.
36.	I found the experience extremely rewarding.

(García Calvo et al., 2008, p. 665; Rufi et al., 2014, p. 6)

		1	2	3	4	5
1.	Sabía que mi capacidad me permitiría hacer frente al desafío que se me planteaba.					
2.	Hice los gestos correctos sin pensar, de forma automática.					
3.	Conocía claramente lo que quería hacer.					
4.	Tenía realmente claro que lo estaba haciendo bien.					
5.	Mi atención estaba completamente centrada en lo que estaba haciendo.					
6.	Sentía un control total de lo que estaba haciendo.					
7.	No me importaba lo que los otros podían haber estado pensando de mí.					
8.	El tiempo parecía diferente a otras veces (ni lento, ni rápido).					
9.	Realmente me divertía lo que estaba haciendo					
10.	Mi habilidad estaba al mismo nivel de lo que me exigía la situación.					
11.	Parecía que las cosas estaban sucediendo automáticamente.					
12.	Estaba seguro de lo que quería hacer.					
13.	Sabía lo bien que lo estaba haciendo.					
14.	No me costaba mantener mi mente en lo que estaba sucediendo.					
15.	Sentía que podía controlar lo que estaba haciendo.					
16.	No estaba preocupado por mi ejecución					
17.	El paso del tiempo parecía ser diferente al normal.					
18.	Me gustaba lo que estaba experimentando en ese momento y me gustaría sentirlo de nuevo.					
19.	Sentía que era lo suficientemente bueno para hacer frente a la dificultad de la situación.					
20.	Ejecutaba automáticamente.					
21.	Sabía lo que quería conseguir.					
22.	Tenía buenos pensamientos acerca de lo bien que lo estaba haciendo mientras estaba practicando.					
23.	Tenía una total concentración.					
24.	Tenía un sentimiento de control total.					
25.	No estaba preocupado por la imagen que daba a los demás.					
26.	Sentía como si el tiempo se parase mientras estaba practicando.					
27.	La experiencia me dejó un buen sabor de boca (buena impresión).					
28.	Las dificultades y mis habilidades para superarlas, estaban a un mismo nivel.					
29.	Hacía las cosas espontánea y automáticamente.					
30.	Mis objetivos estaban claramente definidos.					
31.	Estaba seguro de que en ese momento, lo estaba haciendo muy bien.					
32.	Estaba totalmente centrado en lo que estaba haciendo.					
33.	Sentía un control total de mi cuerpo.					
34.	No me preocupaba lo que otros pudieran estar pensando de mí.					
35.	A veces parecía que las cosas estaban sucediendo como a cámara lenta.					
36.	Encontré la experiencia muy valiosa y reconfortante.					

Appendix 5. – Flow in Human-Computer Interaction

The following questions ask about your feeling while using computers. Please describe a typical (work-related) computer session by placing check marks on the scales given below.

Enjoyment			
	Interesting	____ : ____ : ____ : ____ : ____ : ____ : ____ :	Uninteresting
	Fun	____ : ____ : ____ : ____ : ____ : ____ : ____ :	Not fun
	Exciting	____ : ____ : ____ : ____ : ____ : ____ : ____ :	Dull
	Enjoyable	____ : ____ : ____ : ____ : ____ : ____ : ____ :	Not enjoyable
Concentration			
	Am deeply engrossed in activity	____ : ____ : ____ : ____ : ____ : ____ : ____ :	Not deeply engrossed
	Am absorbed intensely in activity	____ : ____ : ____ : ____ : ____ : ____ : ____ :	Not absorbed intensely
	Attention is focused on activity	____ : ____ : ____ : ____ : ____ : ____ : ____ :	Attention not focused
	Concentrate fully on activity	____ : ____ : ____ : ____ : ____ : ____ : ____ :	Do not fully concentrate
Control			
	Clearly know the right things to do	____ : ____ : ____ : ____ : ____ : ____ : ____ :	Feel confused about what to do
	Feel calm	____ : ____ : ____ : ____ : ____ : ____ : ____ :	feel agitated
	Feel in control	____ : ____ : ____ : ____ : ____ : ____ : ____ :	Do not feel in control
Exploratory use			
	Experiment with commands	____ : ____ : ____ : ____ : ____ : ____ : ____ :	Do not experiment
	Try out new commands	____ : ____ : ____ : ____ : ____ : ____ : ____ :	Do not try new commands
	Experiment with output formats	____ : ____ : ____ : ____ : ____ : ____ : ____ :	Do not experiment with output

Challenge

Overall how challenging do you find the use of computers

Low 0 1 2 3 4 5 6 7 8 9 *High*

(Ghani & Deshpande, 1994, sec. Appendix)

Appendix 6. – The Work-Related Flow Inventory (WOLF)

The following statements refer to the way in which you experienced your work during the last two weeks. Please indicate how often you experienced each of the statements. (1 = never, 2 = almost never, 3 = sometimes, 4 = regularly, 5 = often, 6 = very often, 7 = always).

Absorption

1. When I am working, I think about nothing else
2. I get carried away by my work
3. When I am working, I forget everything else around me
4. I am totally immersed in my work

Work Enjoyment

5. My work gives me a good feeling
6. I do my work with a lot of enjoyment
7. I feel happy during my work
8. I feel cheerful when I am working

Intrinsic Work Motivation

9. I would still do this work, even if I received less pay
10. I find that I also want to work in my free time
11. I work because I enjoy it
12. When I am working on something, I am doing it for myself
13. I get my motivation from the work itself, and not from the reward for it

(Bakker, 2008, app. 1)

Appendix 7. – The Study-Related Flow Inventory (WOLF-S)

(Bakker et al., 2017, app. 1)

The English version of the Study-Related Flow Inventory (WOLF-S)

The following statements refer to the way in which you experienced your academic work during the last two weeks. Please indicate how often you experienced each of the statements. (1=never, 2=almost never, 3=sometimes, 4=regularly, 5=often, 6=very often, 7=always).

1. When I am learning, I think about nothing else
2. I get carried away when I am learning
3. When I am learning, I forget everything else around me
4. I am totally immersed in my studying
5. My studying gives me a good feeling
6. I do my study obligations with a lot of enjoyment
7. I feel happy during my learning
8. I feel cheerful when I am learning
9. I would still learn even if I did not have to
10. I find that I also want to learn in my free time
11. I study because I enjoy it
12. I am learning for my own sake
13. I get my motivation from the learning itself, and not from the grades

(Bakker et al., 2017, app. 1)

Hrvatska verzija Inventara zanesenosti u studiranju (WOLF-S)

Niže navedene tvrdnje odnose se na vaše iskustvo tijekom akademskih aktivnosti u protekla dva tjedna. Molimo vas označite koliko često ste doživjeli što tvrdnja opisuje

(1=nikad, 2=gotovo nikad, 3=ponekad, 4=redovito, 5=često, 6=vrlo često, 7=uvijek)

1. Kada učim, ne mislim ni na što drugo.
2. Učenje me ponese.
3. Kada učim, zaboravim na sve drugo oko mene.
4. Posve sam udubljen/a u studiranje.
5. Moj studij mi daje dobar osjećaj.
6. S puno uživanja obavljam svoje studentske obveze.
7. Osjećam se sretno dok učim.
8. Osjećam se radosno dok učim.
9. Čak i kada ne bih morao/la, učio/la bih i dalje.
10. Shvatio/la sam da želim učiti i u svoje slobodno vrijeme.
11. Učim jer u tome uživam.
12. Učim zbog samog sebe.
13. Motivaciju za učenje nalazim u samom učenju, a ne u ocjenama.

Appendix 8. – The EduFlow & EduFlow-2 measure instruments

(Heutte, Fenouillet, et al., 2014; Heutte, Fenouillet, Kaplan, et al., 2016)

EduFlow	
Dimension	Item
FlowD1a	Je me sens capable de faire face aux exigences élevées de la situation. [I feel I am able to meet the high demands of the situation.]
FlowD1b	Je sens que je contrôle parfaitement mes actions. [I feel that what I do is under my control.]
FlowD1c	À chaque étape, je sais ce que je dois faire. [I know what I have to do at every step of the task.]
FlowD2a	Le temps semble s'écouler de façon différente que d'habitude. [Time seems to flow by in a different way than ever before.]
FlowD2b	J'ai l'impression que le temps passe rapidement. [I feel like the time is flying very fast.]
FlowD2c	Je ne vois pas le temps passer. [I don't notice the time passing.]
FlowD3a	Je ne suis pas préoccupé par ce que les autres pourraient penser de moi. [I didn't care about what the others could think of me.]
FlowD3b	Je ne suis pas préoccupé par le jugement des autres. [I don't fear the judgment of others.]
FlowD3c	Je ne suis pas inquiet de ce que les autres peuvent penser de moi. [I was not worrying about what the others think of me.]
FlowD4a	J'ai l'impression de vivre un moment enthousiasmant. [I have the feeling of living a moment of excitement.]
FlowD4b	Cette activité me procure beaucoup de bien-être. [This activity makes me happy.]
FlowD4c	Quand j'évoque cette activité, je ressens une émotion que j'ai envie de partager. [When I talk about this activity, I feel a strong emotion and I want to share it.]

(Heutte et al., 2021; Heutte, Fenouillet, Martin-Krumm, et al., 2016)

EduFlow-2

Dimension	Item
FlowD1a	Je me sens capable de faire face aux exigences élevées de la situation. [I trust my ability to meet the high demands of the situation.]
FlowD1b	Je sens que je contrôle parfaitement mes actions. [I feel completely in control of my actions.]
FlowD1c	À chaque étape, je sais ce que je dois faire. [At each step, I know exactly what I have to do.]
FlowIMa	Je suis totalement absorbé par ce que je fais. [I am wholly absorbed in what I am doing.]
FlowIMb	Je suis profondément concentré(e) sur ce que je fais. [I am deeply focused on what I am doing.]
FlowD2c	Je ne vois pas le temps passer. [I am losing track of time.]
FlowD3a	Je ne suis pas préoccupé par ce que les autres pourraient penser de moi. [I don't care about what others may think of me.]
FlowD3b	Je ne suis pas préoccupé par le jugement des autres. [I am not concerned about the judgment of others.]
FlowD3c	Je ne suis pas inquiet de ce que les autres peuvent penser de moi. [I am not worried about what others might think of me.]
FlowD4a	J'ai l'impression de vivre un moment enthousiasmant. [I have the feeling I am living a very exciting experience.]
FlowD4b	Cette activité me procure beaucoup de bien-être. [This activity brings me a sense of well-being.]
FlowD4c	Quand j'évoque cette activité, je ressens une émotion que j'ai envie de partager. [When I talk about this activity, I feel such a deep emotion that I want to share it.]

Appendix 9. – Correlations on multidimensional input sample for Experiment 1

Figure 0-1

Spearman's ρ Correlation Heatmap for Flow-Q target y and EduFlow-2 Dimensions

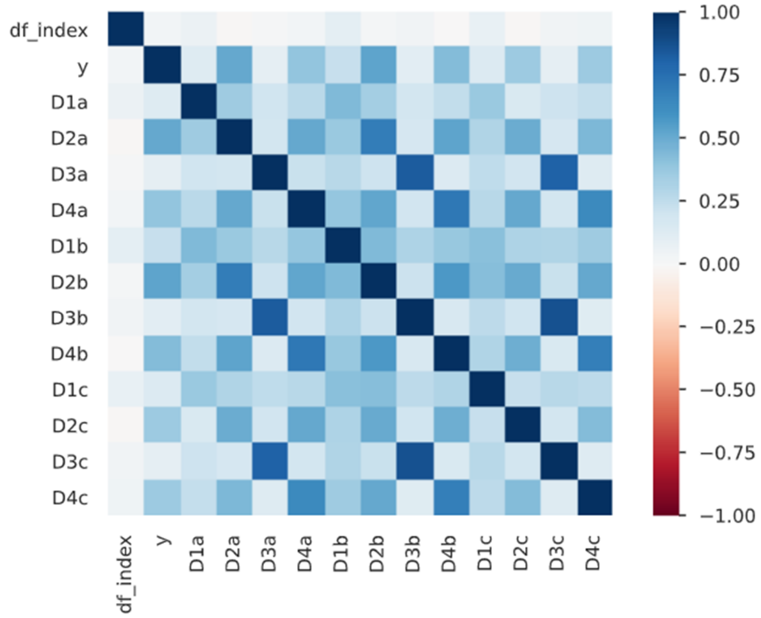


Figure 0-2

Pearson's r Correlation Heatmap for Flow-Q target y and EduFlow-2 Dimensions

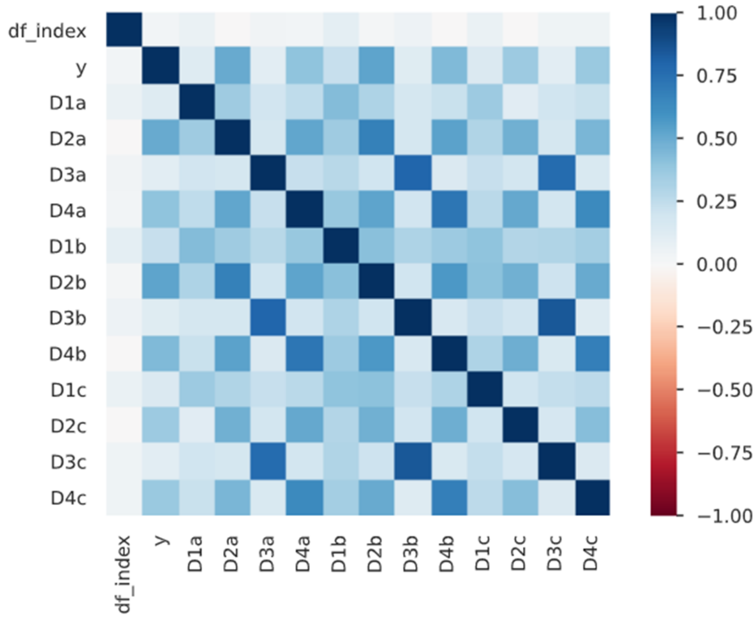
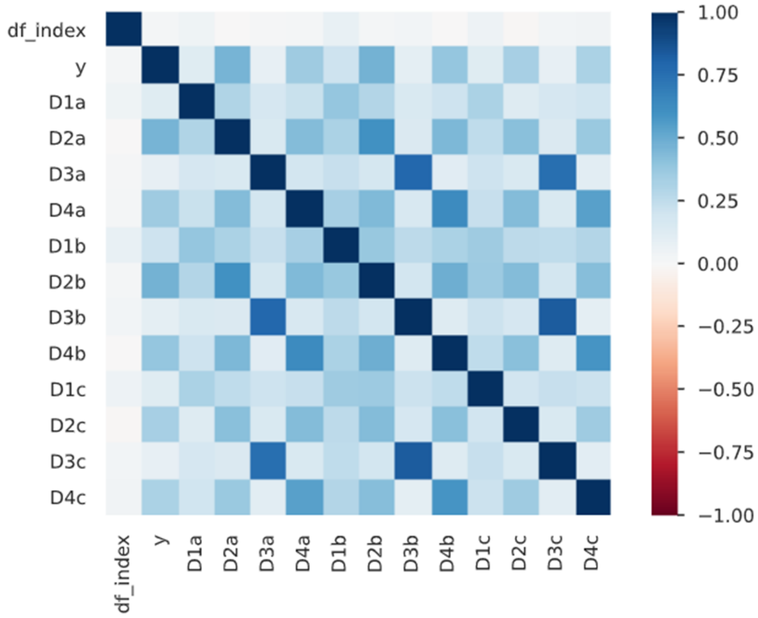


Figure 0-3

Kendall's τ Correlation Heatmap for Flow-Q target y and EduFlow-2 Dimensions



Note: An interactive, stand-alone HTML Pandas Profiling Report download link (*cf.* footnote [143] above) is available at:

https://drive.google.com/file/d/1B0E2uHLawAmb4b7VJprjRiiixzUaEr8R/view?usp=sharing and

https://nextcloud.univ-lille.fr/index.php/s/EH4XpjSn2N4kw8Y

Appendix 10. – Scrollable lists value pairs employed in the questionnaires

List of countries (in French)

Value	Description
1	*Afghanistan*
2	*Afrique du Sud*
3	*Albanie*
4	*Algérie*
5	*Allemagne*
6	*Angola*
7	*Antigua-et-Barbuda*
8	*Arabie saoudite*
9	*Argentine*
10	*Arménie*
11	*Australie*
12	*Autriche*
13	*Azerbaïdjan*
14	*Bahamas*
15	*Bahreïn*
16	*Bangladesh*
17	*Barbade*
18	*Belau*
19	*Belgique*
20	*Belize*
21	*Bénin*
22	*Bhoutan*
23	*Biélorussie*
24	*Birmanie*
25	*Bolivie*
26	*Bosnie-Herzégovine*
27	*Botswana*
28	*Brésil*
29	*Brunei*
30	*Bulgarie*
31	*Burkina*
32	*Burundi*
33	*Cambodge*
34	*Cameroun*
35	*Canada*
36	*Cap-Vert*

37	*Chili*
38	*Chine*
39	*Chypre*
40	*Colombie*
41	*Comores*
42	*Congo*
43	*Corée du Nord*
44	*Corée du Sud*
45	*Costa Rica*
46	*Côte d'Ivoire*
47	*Croatie*
48	*Cuba*
49	*Danemark*
50	*Djibouti*
51	*Dominique*
52	*Écosse*
53	*Égypte*
54	*Émirats arabes unis*
55	*Équateur*
56	*Érythrée*
57	*Espagne*
58	*Estonie*
59	*États-Unis*
60	*Éthiopie*
61	*Fidji*
62	*Finlande*
63	*France*
64	*Gabon*
65	*Gambie*
66	*Géorgie*
67	*Ghana*
68	*Grèce*
69	*Grenade*
70	*Guatemala*
71	*Guinée*
72	*Guinée-Bissao*
73	*Guinée équatoriale*
74	*Guyana*
75	*Haïti*
76	*Honduras*
77	*Hongrie*
78	*Inde*
79	*Indonésie*
80	*Iran*

81	*Irak*
82	*Irlande*
83	*Islande*
84	*Italie*
85	*Jamaïque*
86	*Japon*
87	*Jordanie*
88	*Kazakhstan*
89	*Kenya*
90	*Kirghizistan*
91	*Kiribati*
92	*Koweït*
93	*Laos*
94	*Lesotho*
95	*Lettonie*
96	*Liban*
97	*Liberia*
98	*Libye*
99	*Liechtenstein*
100	*Lituanie*
101	*Luxembourg*
102	*Macédoine*
103	*Madagascar*
104	*Malaisie*
105	*Malawi*
106	*Maldives*
107	*Mali*
108	*Malte*
109	*Maroc*
110	*Marshall*
111	*Maurice*
112	*Mauritanie*
113	*Mexique*
114	*Micronésie*
115	*Moldavie*
116	*Monaco*
117	*Mongolie*
118	*Mozambique*
119	*Namibie*
120	*Nauru*
121	*Népal*
122	*Nicaragua*
123	*Niger*
124	*Nigeria*

125	*Niue*
126	*Norvège*
127	*Nouvelle-Zélande*
128	*Oman*
129	*Ouganda*
130	*Ouzbékistan*
131	*Pakistan*
132	*Palestine*
133	*Panama*
134	*Papouasie - Nouvelle Guinée*
135	*Paraguay*
136	*Pays-Bas*
137	*Pérou*
138	*Philippines*
139	*Pologne*
140	*Portugal*
141	*Qatar*
142	*République centrafricaine*
143	*République démocratique du Congo*
144	*République dominicaine*
145	*République tchèque*
146	*Roumanie*
147	*Royaume-Uni*
148	*Russie*
149	*Rwanda*
150	*Sainte-Lucie*
151	*Saint-Marin*
152	*Saint-Siège*
153	*Saint-Vincent-et-les-Grenadine*
154	*Salomon*
155	*Salvador*
156	*Samoa occidentales*
157	*Sao Tomé-et-Principe*
158	*Sénégal*
159	*Seychelles*
160	*Sierra Leone*
161	*Singapour*
162	*Slovaquie*
163	*Slovénie*
164	*Somalie*
165	*Soudan*
166	*Sri Lanka*
167	*Suède*
168	*Suisse*

169	*Suriname*
170	*Swaziland*
171	*Syrie*
172	*Tadjikistan*
173	*Tanzanie*
174	*Tchad*
175	*Thaïlande*
176	*Togo*
177	*Tonga*
178	*Trinité-et-Tobago*
179	*Tunisie*
180	*Turkménistan*
181	*Turquie*
182	*Tuvalu*
183	*Ukraine*
184	*Uruguay*
185	*Vanuatu*
186	*Venezuela*
187	*Viêt Nam*
188	*Yémen*
189	*Yougoslavie*
190	*Zambie*
191	*Zimbabwe*

List of statuses (in French)

Value	Description
1	Agriculteur-exploitant
2	Artisan, commerçant, chef d'entreprise
3	Cadre et professions intellectuelles (enseignant etc..)
4	Profession intermédiaire (technicien, agent de maîtrise..)
5	Employé
6	Ouvrier
7	Retraité
8	Étudiant
9	En recherche d'emploi
10	Inactif (autre que étudiant, retraité, ou en recherche d'emploi)

List of previous trainings (in French)

Value	Description
1	Non, je n'ai aucune formation dans le domaine.
2	Oui, j'ai une brève formation dans le domaine.
3	Oui, j'ai fait une spécialisation en gestion de projet.

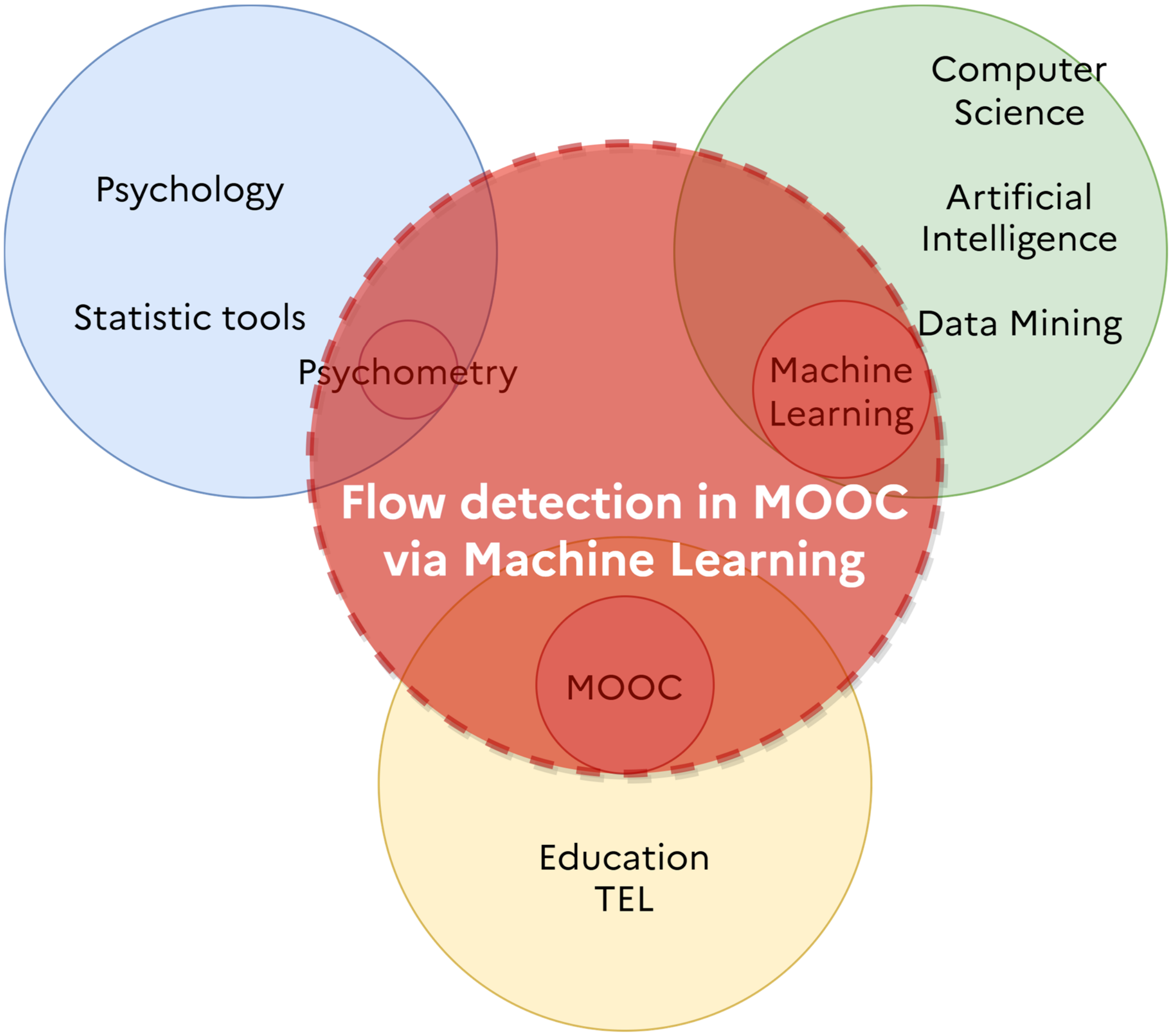
List of ensuing trainings (in French)

Value	Description
1	Des formations régulières à la gestion de projet.
2	Des formations ponctuelles à la gestion de projet.
3	Aucune formation.
4	Je n'ai pas terminé mes études.

Appendix 11. – Doctoral dissertation scope

Figure 0-4

Venn diagram of notions at play for Flow detection in MOOC via Machine Learning

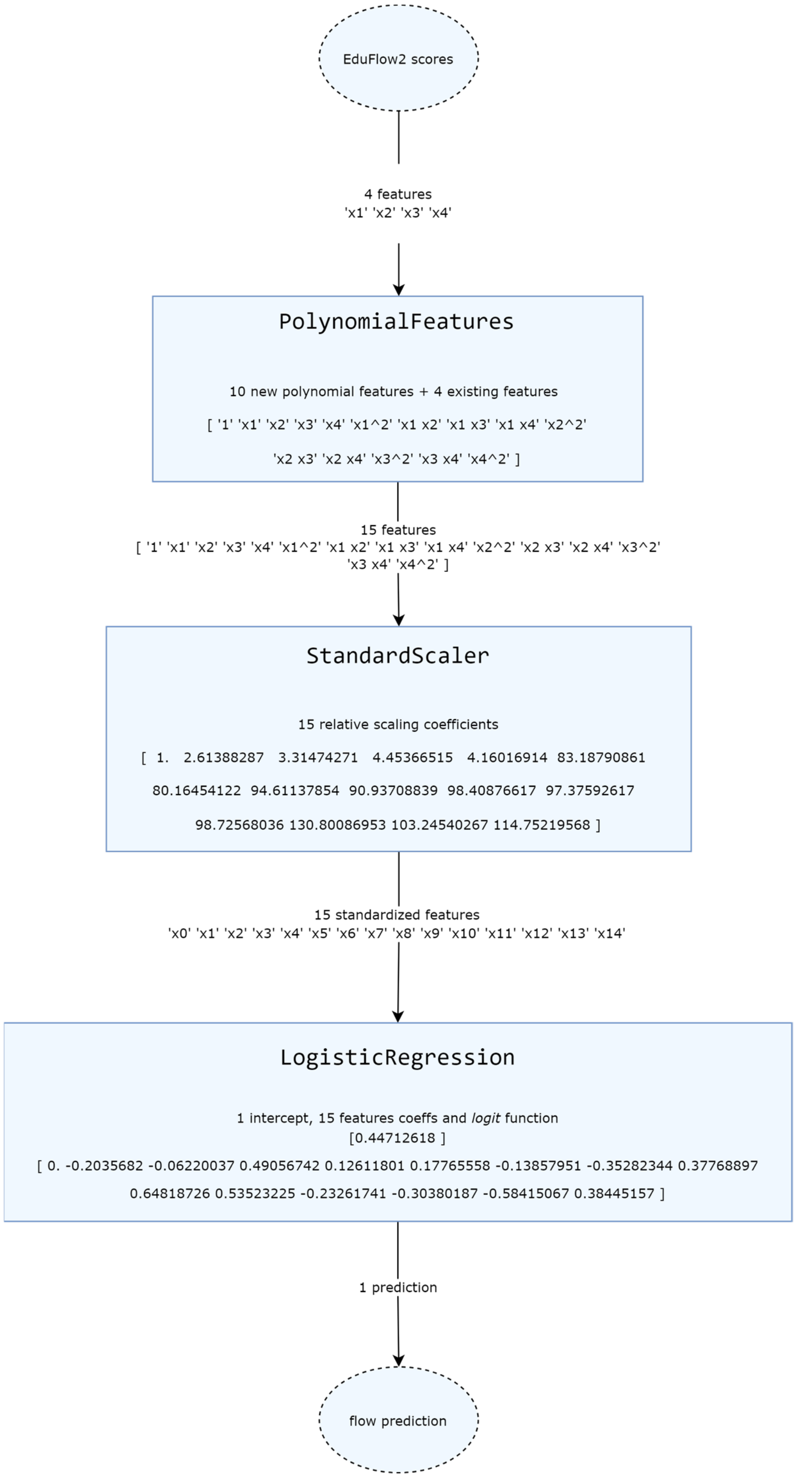


Note: A Venn diagram shows the logical relation between sets (Venn, 1881, Chapter V). Circles' colors and absolute sizes do not carry any meaning but only serve to distinguish sets. The dashed, central, red circle represents the scope of this doctoral dissertation, taking elements from Psychology, Computer Science, and Education.

Appendix 12. – Proof-of-Concept scikit-learn pipeline

Figure 0-5

scikit-learn pipeline of the PoC flow-detecting model



Appendix 13. – Categorization of Open edX events

Table 0-1

Proposed Categorization of Open edX Student Events

Category name	edX events
cat_forum	"edx.forum.response.voted",
	"edx.forum.searched",
	"edx.forum.thread.viewed",
	"edx.forum.thread.voted"
cat_forum_post_comments	"edx.forum.comment.created",
	"edx.forum.thread.created",
	"edx.forum.response.created"
cat_video	"edx.video.closed_captions.hidden",
	"edx.video.closed_captions.shown",
	"edx.video.language_menu.hidden",
	"edx.video.language_menu.shown",
	"hide_transcript",
	"load_video",
	"pause_video",
	"play_video",
	"seek_video",
	"show_transcript",
	"speed_change_video",
	"stop_video"
cat_admin	"edx.instructor.report.downloaded",
	"edx.certificate.configuration.activated",
	"edx.certificate.configuration.created",
	"edx.certificate.configuration.modified",
	"edx.cohort.created",
	"edx.cohort.creation_requested",
	"edx.cohort.email_address_preassigned",
	"edx.cohort.user_add_requested",
	"edx.cohort.user_added",
	"edx.cohort.user_removed",
	"edx.grades.grading_policy_changed",
	"edx.grades.problem.state_deleted",
	"edx.special_exam.timed.allowance.created",
	"edx.special_exam.timed.allowance.deleted",
	"edx.special_exam.timed.attempt.deleted",
	"edx.special_exam.proctored.attempt.deleted",
	"edx.special_exam.proctored.updated",
	"edx.special_exam.timed.created",
	"edx.special_exam.timed.updated"

Category name	edX events
cat_admin	"edx.certificate.configuration.activated",
	"edx.certificate.configuration.created",
	"edx.certificate.configuration.modified",
	"edx.certificate.generation.enabled",
	"edx.cohort.created",
	"edx.cohort.creation_requested",
	"edx.cohort.email_address_preassigned",
	"edx.cohort.user_add_requested",
	"edx.cohort.user_added",
	"edx.cohort.user_removed",
	"edx.course.index.reindexed",
	"edx.grades.grading_policy_changed",
	"edx.grades.problem.state_deleted",
	"edx.special_exam.timed.attempt.deleted"
cat_assessments	"edx.special_exam.proctored.attempt.created",
	"edx.special_exam.proctored.attempt.download_software_clicked",
	"edx.special_exam.proctored.attempt.ready_to_submit",
	"edx.special_exam.proctored.attempt.started",
	"edx.special_exam.proctored.attempt.submitted",
	"edx.special_exam.proctored.option-presented",
	"edx.special_exam.proctored.updated",
	"edx.special_exam.timed.attempt.created",
	"edx.special_exam.timed.attempt.ready_to_submit",
	"edx.special_exam.timed.attempt.started",
	"edx.special_exam.timed.attempt.submitted"
cat_problem_sessions	"problem_check",
	"problem_graded",
	"problem_save",
	"problem_show",
	"edx.grades.problem.submitted"
cat_system	"edx.course.index.reindexed",
	"edx.user.settings.change_initiated",
	"edx.grades.subsection.grade_calculated",
	"edx.librarycontentblock.content.assigned",
	"edx.certificate.created",
	"edx.certificate.evidence_visited",
	"edx.certificate.generation.enabled",
	"edx.course.enrollment.mode_changed",
	"edx.course.enrollment.upgrade.clicked"
cat_navigation	"edx.bookmark.accessed",
	"edx.bookmark.added",
	"edx.bookmark.listed",
	"edx.bookmark.removed",
	"edx.ui.lms.link_clicked",

Category name	edX events
	"edx.ui.lms.sequence.next_selected", "edx.ui.lms.sequence.previous_selected", "edx.ui.lms.sequence.tab_selected", "page_close"
cat_general	"edx.course.enrollment.activated", "edx.course.enrollment.deactivated", "edx.course.goal.added", "edx.course.goal.updated", "edx.course.home.resume_course.clicked", "edx.course.search.initiated", "edx.course.search.result_selected", "edx.course.search.results_displayed", "edx.course.share_clicked", "edx.course.tool.accessed", "edx.course_discovery.search.initiated", "edx.course_discovery.search.results_displayed", "edx.user.settings.viewed", "edx.user.settings.changed"
cat_unknown	any other.

Note: edX events description is detailed in the EdX Research Guide (*cf.* footnote 132 above).

Appendix 14. – Constructed fields for trace analysis

Table 0-2

Constructed Fields xACTION

Fieldname (xACTION)	Description
_id	Document identifier
original_id	Repeated _id
_id_origin	GdP's anonymized identifier
username	MOOC's anonymized identifier
time	Document Timestamp
name	Event name
session	MOOC's session identifier
SECS_epoch_time	Document timestamp transformed to Unix time
INT_day_week	Document timestamp transformed to a weekday value
STRING_time_day	Document timestamp transformed to a typical French day's textual separations (morning, afternoon, night, etc.)
STRING_edx_cat	Event name re-classified
STRING_edx_path	Absolute MOOC path position
event	Original branch event
context	Original branch context

Table 0-3

Constructed Fields xSESSION

Fieldname (xSESSION)	Description
Total_valid_events	Count of all valid events
Types_of_valid_events	Number of distinct types of valid_events
List_of_valid_events	Verification for both previous fields
List_of_valid_STRING_edx_cats	Concatenation of distinct STRING_edx_cat
First_Conn	Timestamp
Last_Conn	Timestamp
SECS_First_Conn	First_Conn converted to Unix time
SECS_Last_Conn	Last_Conn converted to Unix time
STRING_First_Conn	Text string multi conversion verification for First_Conn
BOOL_multiday_session	Comparison of partial string conversion of First_Conn and Last_Conn
BOOL_sunday_q	Comparison of conversions weekday of Last_Conn against First_Conn
DAYWEEK_First_Conn	Conversion of First_Conn to weekday
MINUTES_session_length	Difference between SECS_Last_Conn and SECS_First_Conn in minutes
AverageActionsPerMinute	Ratio between Total_valid_events and MINUTES_session_length

Table 0-4

Constructed Field xUSER

Fieldname (xUSER)	**Description**
Total_valid_events	Count of all valid events
Types_of_valid_events	Number of distinct types of valid_events
List_of_valid_events	[a] Verification for both previous fields
List_of_valid_STRING_edx_cats	[a] Concatenation of distinct STRING_edx_cat
SECS_First_Conn	First_Conn converted to Unix time
SECS_Last_Conn	Last_Conn converted to Unix time
DAYWEEK_First_Conn	Conversion of First_Conn to weekday
SECS_total_timespan	[a] Difference between SECS_Last_Conn and SECS_First_Conn
AverageActionsPerMinute	[a] Ratio between Total_valid_events and MINUTES_session_length
AverageActionsPerMinute100	[a] 100x ratio between Total_valid_events and MINUTES_session_length
cat_general-monday, …	¬ Count of STRING_edx_cat per INT_day_week
Total_valid_sessions	Count of [sessions]
valid_sessions-monday, …	¬ Count of [sessions] per INT_day_week
valid_sessions-matin, …	¬ Count of [sessions] per STRING_time_day
valid_sessions-monday-matin, …	¬ Count of [sessions] per INT_day_week per STRING_time_day
SECS_average_session_length	60 * Mean of [MINUTES_session_length]
SECS_average_session_length-monday	¬ Average session length per INT_day_week
Total_multiday_session	[a] Count of [BOOL_multiday_session]
multiday_session-monday, …	¬ Count of [BOOL_multiday_session] per INT_day_week

Note:

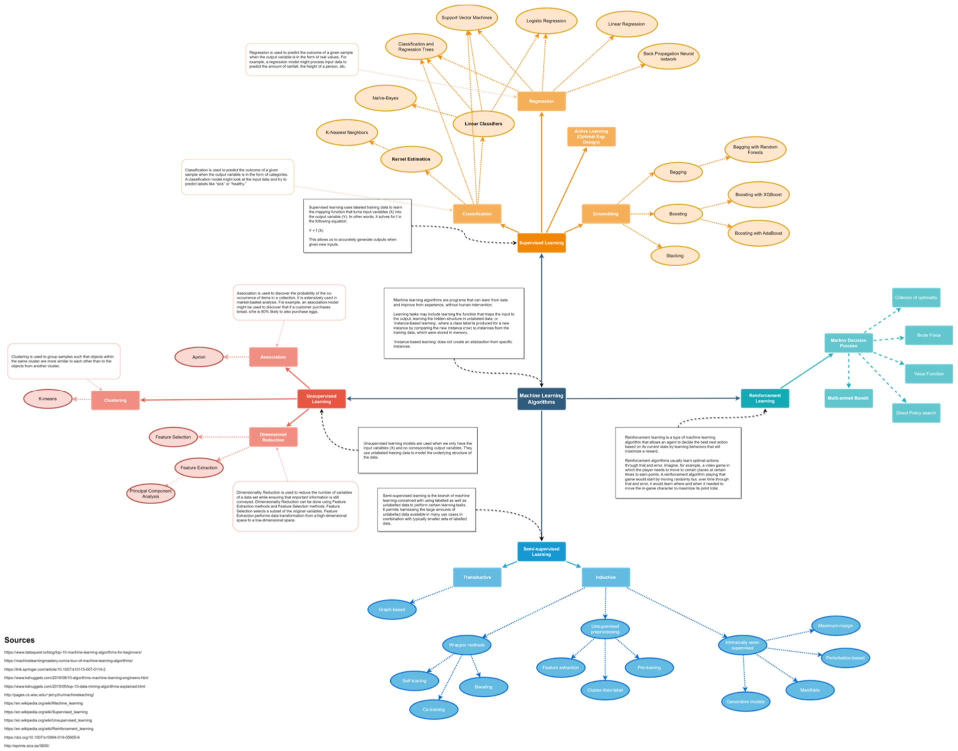
¬ Fields were discarded specifically to avoid the curse of dimensionality.

ª Fields were discarded because they carried no relevant information.

Appendix 15. – Machine Learning algorithms

Figure 0-6

Overview of a Mental Map of Machine Learning Algorithms (Ramírez Luelmo, El Mawas, & Heutte, 2022)



Note: For readability purposes, this Mental Map is available for download at https://drive.google.com/file/d/1xPPpSsFW3eiI67VyiHSLInMSrzZf3ssU/view?usp=drive_link