

UNIVERSITE DE LILLE – SECTEUR DROIT ET SANTE
FACULTE DE MEDECINE HENRI WAREMBOURG
Année 2021

THESE POUR LE DIPLOME D'ÉTAT
DE DOCTEUR EN MEDECINE

Modules d'intelligence artificielle au cabinet de ville : qu'en attendent les médecins généralistes français ?

Présentée et soutenue publiquement
le Mardi 21 Septembre 2021 à 16h00
au Pôle Formation

Par Smaïn TABLA

JURY

Président :

Monsieur le Professeur Emmanuel CHAZARD

Assesseurs :

Madame le Docteur Sabine BAYEN

Madame le Docteur Gabrielle LISEMBARD

Monsieur le Docteur Pierre BALAYÉ

Directeur de thèse :

Monsieur le Professeur Emmanuel CHAZARD

Avertissement

La Faculté n'entend donner aucune approbation aux opinions émises dans les thèses : celles-ci sont propres à leurs auteurs.

Sigles

ANN	<i>Artificial neural network</i>
CDSS	<i>Clinical Decision Support System</i>
DL	<i>Deep Learning</i>
DMLA	Dégénérescence maculaire liée à l'âge
EHR	<i>Electronic Health Record</i>
EIM	Effet Indésirable Médicamenteux
EKG / ECG	<i>Electrocardiogram / Electrocardiogramme</i>
GP	<i>General Practitioner</i>
HTA	Hypertension Artérielle
IA	Intelligence Artificielle
ML	<i>Machine Learning</i>
OCT	<i>Optical Coherence Tomographie</i> (tomographie par cohérence optique)
OMS	Organisation Mondiale de la Santé
RPD	Rétinopathie Diabétique
SADM	Systèmes d'Aide à la Décision Médicale
SVM	<i>Support vector machine</i>

Sommaire

Avertissement	2
Remerciements	Erreur ! Signet non défini.
Sigles	3
Sommaire.....	4
Préambule.....	6
Introduction	7
1 L'intelligence artificielle	7
1.1 Définition.....	7
1.2 Historique.....	7
1.3 Trois niveaux d'IA	8
2 Domaines d'application de l'IA.....	9
2.1 Panorama des applications en santé.....	9
2.2 Application de l'IA à la médecine générale	10
3 Les enjeux	10
3.1 Enjeux liés à l'évaluation des technologies de santé	10
3.2 Enjeux liés à la responsabilité médicale : éthiques et morale.....	10
3.3 Enjeux liés à l'acceptabilité en pratique des modules théoriquement valides 11	
3.4 Enjeux au niveau mondial.....	11
4 Objectif.....	12
Abstract.....	13
Article en Anglais	14
1 Introduction	14
2 Material and methods	15
2.1 Design of the survey	15
2.2 Recruitment of participants	15
2.3 Data Analysis	15
2.4 Statistical analysis.....	16
3 Results.....	16
3.1 Analyzed questionnaires.....	16
3.2 Participants' characteristics	16
3.3 Practitioner's attitude toward AI modules characteristics.....	17
4 Discussion	20

5 Conclusion	21
Conclusion	22
Liste des tables	23
Liste des figures	24
Références.....	25
Annexe 1 : courrier envoyé aux participants	26
Annexe 2 : questionnaire	27

Préambule

Le travail scientifique présenté dans cette thèse de médecine fait l'objet d'une publication d'article international en anglais. Il suit le plan suivant :

- Une introduction longue en français, qui poursuit deux objectifs : présenter le contexte médical avec une orientation principalement pédagogique, et présenter le contexte scientifique et l'objectif, comme le fait également l'introduction de l'article en anglais
- L'abstract en anglais, tel qu'il a été soumis en complément de l'article reproduit juste après.
- L'article en anglais, tel qu'il a déjà été soumis à une revue scientifique internationale. Cet article suit le plan classique, dans le format imposé par le journal (introduction, matériel et méthodes, résultats, discussion)

Le document est structuré ainsi en application de la circulaire Toubon¹.

Les références présentées en fin de document, ainsi que les listes de figures et tables, résultent de la fusion des parties en anglais et en français. La numérotation est donc incrémentée dans l'ensemble du document, que les parties soient anglophones ou francophones.

¹ Circulaire du 19 mars 1996 concernant l'application de la loi no 94-665 du 4 août 1994 relative à l'emploi de la langue française. JORF n°68 du 20 mars 1996 page 4258. NOR: PRMX9601403C

Introduction

1 L'intelligence artificielle

1.1 Définition

L'intelligence (naturelle) est définie comme un ensemble de capacités mentales de raisonnement, de résolution de problème et d'apprentissage ; une aptitude à intégrer des fonctions cognitives telles que la perception, l'attention, la mémoire, le langage ou la planification [1]. L'intelligence artificielle (IA) désigne la capacité d'un processus automatisé à répondre à des stimuli environnementaux (ou à des nouvelles données), d'une manière qui paraît comparable à celle des êtres humains [2].

1.2 Historique

L'histoire des débuts de l'informatique biomédicale, y compris les premières approches de l'IA, peuvent être considérées comme une série de tentatives pour étudier, comprendre et construire des modèles informatiques des connaissances scientifiques. Une première publication de Warren Weaver en 1949 émet l'idée qu'une machine pourrait effectuer une tâche qui relève de l'intelligence humaine [3]. En 1950, Alan Turing pose la question : « Une machine peut-elle penser ? », pour cela il élabore le test de Turing permettant de différencier les humains des machines [4]. Le terme IA est reconnu comme véritable domaine scientifique en 1956, lors d'une conférence aux États-Unis au Dartmouth Collège par John McCarthy [5,6]. En 1968, apparaît la première génération des réseaux neuronaux artificiels (*Artificial neural network*, ANN) exposée par Minsky [7], constituant un réseau de neurones artificiels basés sur le concept du perceptron imaginé dix ans plus tôt. C'est au milieu des années 1970, que les travaux menés par l'université de Stanford [8] et l'université Rutgers [9] ont permis la conception de système d'expert [10,11]. La généralisation de ces systèmes a la fin des années 1970, se révèle comme véritable piste d'aide à la décision clinique [12]. Durant les années 1980, les modèles statistiques qui seront utilisés par la suite ont vu leur développement poursuivi [12]. Une véritable rupture survient dans les années 1990, lorsqu'est émise l'idée de générer de nouvelles connaissances en fouillant automatiquement des données historiques [13,14]. C'est ainsi qu'est né le concept de data mining, puis de *machine learning* : les résultats de ces méthodes pouvaient dès lors être injectés directement dans les modules d'intelligence artificielle. Au cours des années 2000, un « boom » de l'IA apparaît, avec des méthodes d'apprentissages automatiques innovantes tel que les machines à vecteurs de supports (SVM), le *deep learning* (DL), ou encore les *multi-layer neural networks* [15].

Malgré l'absence réelle de nouveau concept, la dernière décennie semble assister à une explosion des applications de l'IA visibles du grand public et des professionnels pour plusieurs raisons [16,17] :

- l'émergence relativement récente des méthodes de DL au sein des méthodes de Machine Learning (ML), et l'amélioration continue des autres méthodes de ML
- une augmentation des capacités de calcul et de mémoire des ordinateurs ou serveurs

- une explosion en volume et en diversité des données d'apprentissage disponibles (*big data* [18])
- la diffusion ubiquitaire de terminaux de calculs et de capture que sont les smartphones
- la multiplicité des applications possibles, y compris avec des débouchés commerciaux immédiats (ex : e-commerce, publicité ciblée, etc.)

Pour le grand public (incluant les décideurs et journalistes), l'IA résonne comme une discipline émergente alors qu'elle a été définie dans les années 1950, et que les concepts actuellement exploités datent des années 1990.

1.3 Trois niveaux d'IA

Trois niveaux d'IA peuvent classiquement être décrits.

Le premier niveau consiste à l'exécution de règles simples par un moteur d'inférence. Ces règles doivent au préalable avoir été décrites dans une base de règles par un expert humain. Un exemple en est la génération automatique d'alertes sur les interactions médicamenteuses dans les systèmes d'aide à la décision médicale (SADM) [19].

Le second niveau consiste à ce qu'un humain écrive un programme dont l'exécution sur des données réelles d'apprentissage permet à la machine de générer automatiquement la base de règles. Ces règles s'appuient donc sur des observations passées. On parle de ML, car l'écriture des règles découle d'un « apprentissage » fait par la machine sur des données réelles annotées ou contenant « le futur ». Une intervention humaine demeure nécessaire pour dicter au processus la manière d'analyser les données. Les règles sont ensuite exécutées automatiquement par un moteur d'inférence, de la même manière que précédemment. Parfois, la mécanique d'apprentissage s'appuie sur des techniques d'une complexité telle que la base de règles est inintelligible pour un humain, et constitue une « boîte noire ». C'est le cas par exemple des techniques de DL. Nous pouvons citer le développement d'outils destinés à rendre plus performant les SADM pour prévenir les effets indésirables médicamenteux (EIM), en réutilisant des données cliniques réelles [20].

Enfin, le troisième niveau, purement théorique, permettrait à une machine de mélanger l'exploitation de règles prédéfinies, de l'expérience et d'une faculté de digression et d'imagination, à l'aide de logique floue, avec intégration de conclusions contradictoires. On parlerait d'IA forte, par opposition aux deux premiers niveaux qui relèvent de l'IA faible. Ce troisième niveau, s'il existait, permettrait une adaptation intelligente du comportement du processus, et lui permettrait d'explicitement ses propres raisonnements. Ce dernier niveau inexistant à ce jour, suscite l'excitation au sein du grand public, alimente le fantasme d'un possible transhumanisme, avec des représentations futuristes, et d'un possible remplacement du médecin par la machine. Il relève à ce jour de la pure science-fiction.

2 Domaines d'application de l'IA

2.1 Panorama des applications en santé

L'IA montre actuellement des champs d'application de plus en plus nombreux en santé, principalement dans quatre catégories de tâches cliniques [21]: le dépistage, le diagnostic, l'analyse des risques, et le traitement. Nous en citerons ici quelques exemples.

En radiologie, la croissance des publications de IA [22] met en évidence le potentiel et les perspectives très prometteuses. Le flux de travail peut être segmenté en 4 étapes [23] : planification, acquisition des images, interprétation, résultats. Ainsi à chaque étape une série de tâches typiques est répertoriée et pourrait être effectuée par une IA afin de rendre plus efficace la prise en charge de patient entrant dans ce circuit [24]. Ainsi par exemple, un module d'IA est développé pour évaluer la pertinence des ordonnances afin d'améliorer l'efficacité et l'orientation [25] des prises en charges des trouble musculo-squelettiques. Selon *Vivanti et al*, des modules d'IA ont pu détecter de nouvelles tumeurs hépatiques automatiquement [26]. *Ben-Cohen et al*, ont développé un modèle d'IA, pouvant retrouver l'origine primaire des métastases hépatiques à l'aide d'image tomodensitométrie parmi quatre sites (mélanome, cancer colorectal, cancer du pancréas et cancer du sein) [27]. Des algorithmes sont proposés [28], pour la détection de cancers et la prédiction du risque en mammographie. Cependant, la validation clinique fait largement défaut, et la manière dont l'IA devrait être utilisée pour optimiser la pratique n'est pas claire.

En ophtalmologie, l'application de l'IA se concentre principalement sur les maladies à forte incidence, telles que la rétinopathie diabétique (RPD), la dégénérescence maculaire liée à l'âge (DMLA), le glaucome, la rétinopathie du prématuré, ou la cataracte [29]. Ces applications s'intéressent tantôt à la mesure de la gravité d'une maladie, à la progression de la maladie ou à l'aide à la décision thérapeutique [30].

En oncologie, les applications de l'IA sont nombreuses, et dépassent le champ de l'imagerie [31] (anatomopathologie [32], aide à la décision en temps réel pendant l'endoscopie [33], microbiologie, séquençage du génome, etc.). De nombreuses applications nécessitent au préalable la constitution de banques de données d'apprentissage [31].

En cardiologie, des modules ont été développés pour améliorer le diagnostic et affiner la prise en charge thérapeutique [34]. Ainsi par exemple, alors que des modules d'IA visent à l'interprétation immédiate des ECG depuis plusieurs décennies [35], certains s'intéressent désormais à la prédiction d'un arrêt cardiaque dans les heures suivant l'ECG [36].

En psychiatrie, l'évaluation du risque suicidaire est subjective, par conséquent les mécanismes de prédiction sont limités. Des approches d'IA sont développées pour prédire les suicides en cascade et identifier les populations à risque [37].

En infectiologie, dans le contexte d'urgence sanitaire lié à la COVID-19, des applications d'aide à la décision ont été développées pour les patients symptomatiques [38]. De nouvelles méthodes basées sur l'IA sont développées pour le diagnostic du paludisme sur goutte épaisse [39]. Un système basé sur des capteurs électroniques [40] est étudié pour le dépistage de masse et la détection précoce des patients tuberculeux dans les zones endémiques.

À travers ces quelques exemples, nous pouvons voir que les applications de l'IA sont très diversifiées, tant en termes de disciplines et pathologies, que d'examens cliniques ou paracliniques concernés.

2.2 Application de l'IA à la médecine générale

Les modules d'IA se développent progressivement au cabinet de ville. Nous en citerons quelques-uns.

L'auscultation cardio-pulmonaire à l'aide de stéthoscope intelligent permet la détection de quatre classes de sons tels les sibilants, les ronchis, les crépitants fins et grossiers [41] ou encore de différencier les bruits du cœur normaux et pathologiques [34] avec une possibilité d'échanger à distance avec d'autres confrères.

Un SADM est développé pour la prise en charge personnalisée telle que la détermination optimale de la pression artérielle, ou encore l'identification de schéma médicamenteux d'antihypertenseur le plus efficace [42].

Les interprètes automatisés d'ECG sont largement utilisés au cabinet de médecine ville [43]. En France, en 2018 déjà, un quart des médecins généralistes français étaient équipés d'ECG avec interpréteurs automatisés [43]. Le défibrillateur automatique analyse le rythme, pose un diagnostic, et aide à la décision thérapeutique dans un contexte d'urgence.

En dermatologie, les praticiens pourront utiliser des modules d'IA pour la distinction des images dermoscopiques de mélanome et de nævus [44]. Cette classification pourrait améliorer la précision du diagnostic [45].

Ainsi les nouvelles technologies peuvent apparaître comme des outils potentiellement performants. Ces SADM pourraient contribuer à la pratique d'une médecine prédictive, préventive, personnalisée et participative. Cependant ces modules d'IA posent également des questions sociétales, technologiques, éthiques, pratiques, et économiques.

3 Les enjeux

3.1 Enjeux liés à l'évaluation des technologies de santé

L'évaluation de ces nouvelles technologies semble indispensable pour pouvoir intégrer un module d'IA dans la démarche clinique. Les performances des applications d'IA doivent être évaluées dans des contextes cliniques réels [46]. Certaines études mènent leur évaluation dans un contexte clinique défini et réaliste, dans un flux de travail (workflow) également défini. C'est le cas par exemple d'une étude de gastroentérologie où le module d'IA assiste la réalisation d'une fibroscopie [33], de l'analyse secondaire d'images échographiques de lésions mammaires [47], ou de la détection de fractures osseuses sur des clichés radiologiques standards en soin primaire [48].

3.2 Enjeux liés à la responsabilité médicale : éthiques et morale

De nombreux questionnements moraux et éthiques sont soulevés lorsque les systèmes d'IA sont impliqués dans le processus de prise de décisions.

Deux conditions définissent qu'une personne soit moralement responsable d'une action : la première est la condition de contrôle (sur la décision ou l'action) et la seconde est épistémique (connaissance propre à un groupe social, à une époque donnée) c'est-à-dire une compréhension suffisante de la décision et des conséquences probables [49].

Les systèmes intégrant un processus décisionnel seraient remis en question sur le plan des responsabilités morales.

Néanmoins, les initiatives européennes visent à élaborer des limites directrices éthiques et morales pour l'IA. Ces initiatives contribuent à établir des normes mondiales afin d'assurer la responsabilité, la confidentialité, la transparence, l'équité et la sécurité nécessaires à la société pour garantir la confiance envers ces nouvelles technologies [50].

3.3 Enjeux liés à l'acceptabilité en pratique des modules théoriquement valides

Bien que les recherches sur l'IA en santé soient en constante augmentation, leur mise en œuvre dans la pratique clinique réelle n'est pas ubiquitaire dans tous les champs [51]. Plusieurs explications [21] sont soutenues pour expliquer l'écart entre la recherche qui établit des modules théoriques d'IA et leur implication pratique.

D'une part, un ensemble de facteurs [51] politiques, économiques, de normes de pratiques médicales déterminent la manière dont les soins de santé sont dispensés. D'autre part concernant le substrat de l'IA de niveau 2, qui est le recueil de données, plusieurs problématiques se dégagent. L'infrastructure[51] n'est pas répandue dans chaque centre de soin, indispensable pour collecter, harmoniser, partager, protéger la confidentialité des données recueillies. Nous pouvons également citer les difficultés techniques, le manque de transparence des algorithmes, la nature changeante dans le domaine de la santé. Par ailleurs, nous devons également prendre en compte le consentement du patient.

Pour cela une étude [52] a permis de mettre en évidence la satisfaction de la part des patients pour le dépistage de la rétinopathie diabétique (RD) par une IA. A contrario, la grande majorité des patients ont souhaité avoir une consultation physique sans intervention de IA lors des entretiens de santé sexuelle reproductive [53]. Cela pourrait nous laisser penser que dans certaines disciplines l'IA aura moins sa place.

Les opinions réfractaires à l'IA s'appuient essentiellement sur l'immaturation de la technologie et la méfiance à l'égard des entreprises liées [54]. Enfin, nous devons également mentionner que de nombreux patients croient que les modules d'IA sont amenés à remplacer complètement ou partiellement les médecins humains [54]. Inversement, les médecins généralistes semblent considérer que le potentiel de l'IA est limité sur le plan des technologies futures, les préoccupations sociales et éthiques[55]. Selon eux, l'IA ne les remplacera donc pas.

3.4 Enjeux au niveau mondial

L'Organisation Mondiale de la santé (OMS), a reconnu l'urgence d'aborder ces questions de santé numériques et d'IA, en vue de proposer des soins issus de ces nouvelles technologies de manière plus universelle.

Plusieurs concertations ont eu lieu pour aborder ce sujet [56–58]. Par exemple, en mai 2018, l'Assemblée mondiale de la santé a adopté une résolution sur les technologies numériques pour la couverture sanitaire universelle [59]. En octobre 2019, The Lancet [60] et le Financial Times ont inauguré une commission conjointe axée sur la convergence de la santé numérique et de l'IA.

En 2021, le marché mondial de l'IA dans le domaine de la santé devrait atteindre 6,6 milliards de dollars américains et pourrait être multiplié par 10 au cours des cinq prochaines années [61]. Le sujet a attiré l'attention de nombreux auteurs, et fait couler beaucoup d'encre [62].

4 Objectif

L'IA s'immisce de plus en plus dans la pratique médicale. Le domaine de l'ingénierie met à disposition de nouveaux instruments pouvant aider les professionnels de santé. Le développement des technologies de santé suit un processus induit par la technologie : des recherches rendent une technique possible, l'évaluent plus ou moins, puis un industriel parvient à lever des fonds pour financer un développement de produit. L'approche initiale est donc d'abord technique puis financière et par conséquent n'émane pas des besoins de l'utilisateur. L'impulsion des start-up motivées par un intérêt financier n'est pas toujours en relation avec le bénéfice médical de ces nouvelles technologies. De nombreux procédés techniquement matures ne trouvent pas de marché car ils ne répondent pas à un besoin, ou y répondent de manière non-utilisable ou non-acceptable.

Notre hypothèse est qu'il faudrait clairement identifier les critères d'acceptabilité d'un module d'IA avant de chercher à le commercialiser. Cela n'a pas été clairement fait ou pas de manière générique.

Notre objectif est d'identifier et de prioriser les critères d'acceptabilité ou de désirabilité d'un module d'IA au cabinet de médecine générale.

Abstract

Introduction

Development of artificial intelligence (AI) modules should rely on technical progress, but also on users' needs. Our objective is to identify criteria that make a hypothetical AI module desirable for general practitioners (GPs).

Method:

Random selection of 200 French GPs, and paper-based questionnaire.

Results:

The population was representative. GPs expect AI modules to diagnose or eliminate an urgent pathology for which they are not competent and for which specialists are not available. They also demand interoperability, automated electronic health record integration and facilitated information sharing. GPs would like AI modules to make them save time, simplify some procedures and delegate tasks to the secretary. They expect AI modules to allow them to associate the patient with the care, to reassure him or her, and to personalize the care. Interestingly, GPs would also rely on a machine to cut off abusive requests, such as work stoppages or certificates of convenience.

Article en Anglais

1 Introduction

Intelligence denotes as a set of mental capacities for reasoning, problem solving and learning; an ability to integrate cognitive functions such as perception, attention, memory, language or planning [1]. Artificial intelligence (AI) refers to the ability of a process to respond to environmental stimuli (or new data) in a manner whose results are comparable to that of human beings [2].

The first level of AI consists in setting up an inference engine connected to a knowledge base explicitly described by experts. This inference engine analyzes the data it receives and can, for example, issue a detailed message. This type of AI module is currently widely deployed in healthcare. The second level of AI takes advantage of machine learning techniques. A first step consists in automatically generating the rule base, generally in the form of predictive models, by using a learning dataset. The second step is then similar to the one described in level 1 AI. This second level takes advantage of deep learning methods.

In healthcare, AI is mainly used to support medical decision. More precisely, it can be used for disease screening or triage, diagnostic assistance, risk analysis, or assistance in setting up, adapting and monitoring treatment [21].

The applications of AI in healthcare are ubiquitous. We will cite a few examples here. In any case, it is important to keep in mind that the outputs proposed by AI modules are subject to error, and most of the time do not aim at replacing a health professional, but only at helping him or her, or saving time [21]. AI applications are numerous in imaging. They are also developing in ophthalmology for the screening and diagnosis of glaucoma [63] and diabetic retinopathy on fundus photographs. Applications also exist in oncology [31], in cardiology, in suicide prevention [37] or in the COVID-19 pandemic [38].

Some medical devices with artificial intelligence are also intended for use in general medicine. We can mention "intelligent" stethoscopes, which automatically classify the sounds heard during pulmonary auscultation [41] or cardiac auscultation [34], with the possibility of remote transmission. Automated electrocardiogram (EKG) interpreters have also invaded the daily life of general practitioners (GPs) [43]. In France in 2018 already, a quarter of French GPs were equipped with electrocardiographs with automated interpreters [43].

The development of health technologies follows a technology driven process: researchers make a technique possible, evaluate it, and then an industrialist manages to raise funds to finance a product development. Usability evaluations of an AI module do exist [52], but they remain relatively rare.

It seems to us that the process is first technical, then financial, but does not necessarily stem from a user need. We have been asked to develop products that are technically feasible, but totally unrealistic in terms of workflow, such as the automated interpretation of an ultrasonographic image, when it is obvious that only an experienced sonographer is capable of capturing the image to be analyzed, and that this same person no longer needs an AI module to interpret it. Many technically mature

processes may not find a market if they do not meet a need, or meet it in a non-useful or non-acceptable way.

Our hypothesis is that the acceptability criteria of an AI module should be clearly identified before seeking to develop it. Our objective is therefore to identify the domain independent criteria that would make a hypothetical AI module desirable for GPs.

2 Material and methods

2.1 Design of the survey

The questionnaire was defined by the authors. It was tested with 8 other GPs and iteratively improved. It includes a part allowing to describe the interviewed practitioner. The main part of the questionnaire consists of a set of 7-modality Likert scales. The main question is: "Imagine in the future a hypothetical device with a diagnostic or therapeutic decision support module (e.g. EKG or otoscope with computerized interpretation). This excludes web sites and medical biology sample analysis. For each of the following features, assuming it applies to this hypothetical device, would it encourage you to purchase such a device or not?" The items then proposed can be seen in table 2 of the results section. The 7 possible responses are for each item (for some statistical analyses, we will use the number in brackets):

- I would never buy it (-3)
- I am very discouraged (-2)
- It discourages me a little (-1)
- It has no impact (0)
- It encourages me a little (+1)
- It strongly encourages me (+2)
- I would definitely buy it (+3)

Intentionally, the term "artificial intelligence" is never used in the questionnaire, because from our previous experience its definition is not clearly known by practitioners, and this term is subject to fantasies.

The complete form is available in Annex 2.

2.2 Recruitment of participants

Participants were randomly drawn from a national directory of GPs. A paper letter with a prestamped return envelope was systematically sent to them. As the survey was strictly anonymous, all included GPs (even those who had responded) were called back by telephone after 2 weeks to improve the response rate.

The cover letter is available in Annex 1.

2.3 Data Analysis

Descriptive analyses were performed for each variable. The items evaluated using Likert scales were classified as "neutral items" when the proportion of "0 (it has no impact)" was equal or greater than 1/3, otherwise they were classified as "incentive" or

“deterrent” items according to the average of answers. We tested the independence between each Likert scale (an integer ranging from -3 to +3) and the participant’s sex, the urban location of the practice, and the participant’s age. Only statistically significant results were then reported.

2.4 Statistical analysis

Categorical variables were expressed in numbers and percentages. Quantitative variables were expressed as mean and standard deviation (SD) if the histogram revealed a symmetrical pattern distribution, and median first and third quartile (Q1, Q3) otherwise.

The independence between a quantitative variable and a categorical variable was tested using Welch’s t-test. The independence between two quantitative variables was tested using Pearson’s correlation coefficient nullity test.

Statistical tests were bilateral. The p values were considered significant at the 5% threshold.

3 Results

3.1 Analyzed questionnaires

Two hundred GPs were randomly selected. All of them were sent the questionnaire. No mail was returned in error. Among them, 139 (69.5%) answered the survey. Thirteen questionnaires (9.3%) were excluded because of poor quality (e.g., all the answers to Likert scales were identical). In total 126 questionnaires could be analyzed.

The flowchart is presented in Figure 1.

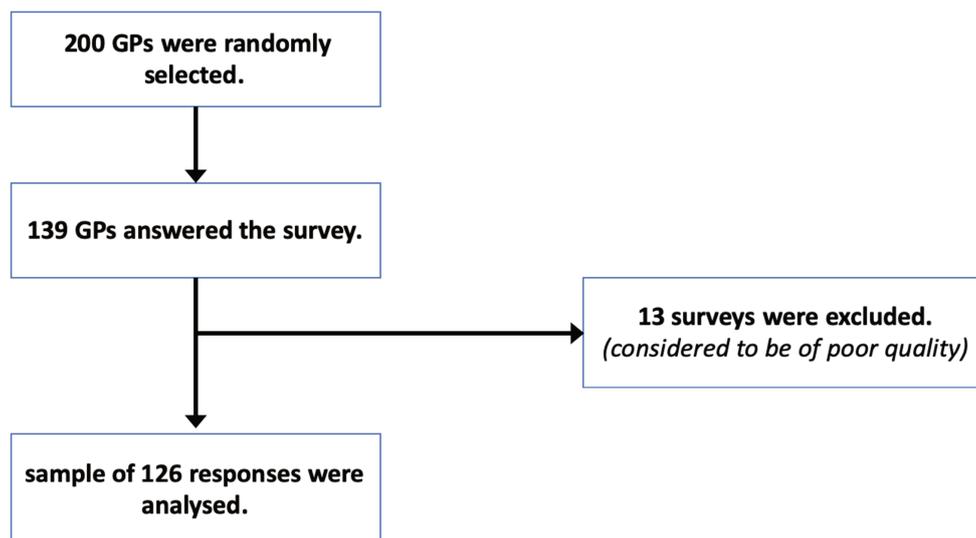


Figure 1. Flowchart

3.2 Participants’ characteristics

There were 66 (52.4%) women, the average age was 47.8 years (SD=11.2). Men were significantly older than women (50.5 vs 45.4, $p= 0.011$). Among participants, 88

(71.0%) worked in a group practice, 32 (25.8%) worked alone, and 4 (3.2%) worked in another organization. The practice location was urban in 90 cases (71.4%), semi-rural in 30 cases (23.8%) and rural in 6 cases (4.8%). Participants were significantly younger in urban practices than in rural or semi-rural practices (respectively 45.5, 52.0 and 54.0 years, $p=0.0006$). Table 1 reports the devices the GPs report having already used or had at disposal during consultations.

Table 1. Devices with AI that GPs already used

Device	n (%)
EKG with computerized interpretation	46 (36.5%)
Connected blood pressure monitor	8 (6.3%)
(Semi-)automatic defibrillator	8 (6.3%)
Connected saturometer	3 (2.4%)
Connected bathroom scale	2 (1.6%)
Connected glucose meter	1 (0.8%)
Connected thermometer	1 (0.8%)
Connected smart stethoscope	0 (0.0%)
Dermatological pathology diagnosis	0 (0.0%)

3.3 Practitioner's attitude toward AI modules characteristics

Table 2 reports the 33 items that were evaluated using Likert scales. Items are sorted according to the average of the answers analyzed as numbers ranging from “-2 (I would never buy it)” to “+2 (I would definitely buy it)”. The proportion of neutral answers is the proportion of “0 (it has no impact)”.

“Incentive items” (Table 2) relate to AI modules that:

- Enable to make or eliminate an urgent diagnosis
- Enable to make a diagnosis for which the GP is not competent, or for which no specialist is easily available
- Enable to register information in the electronic health record (EHR) or to send information to healthcare professionals
- Enable to simplify a procedure, save time or spend more time with the patient
- Enable to predict a risk category or to perform personalized medicine
- Enable to involve the patient and reassures him/her
- Enable to cut short patients' abusive requests (e.g.: certificates, work stoppages, etc.)
- Train the GP him/herself

“Deterrent items” (Table 2) relate to AI modules that: automatically trigger an emergency call, automatically decide without the GP intervention, or make a diagnosis for which the GP is already competent.

“Neutral items” (Table 2) relate to:

- Task delegation
- Reputation with patients, “high tech” image of the practice, selfuse by patients, patient entertainment
- The way the information has to be captured, the need for training, the use of personal smartphones
- The economic model of the device
- Digression ability

Table 2 also shows that many items are preferred by young GPs, and that some items are preferred according to the gender or the practice location.

Table 2. Results analysis of Likert scales: "Would this feature encourage or discourage you from purchasing this module?"

Form order	Item	Mean	Conclusion	Preferred by
9	It quickly eliminates an urgent diagnosis	2.25	Incentive	-
10	It quickly makes an urgent diagnosis	2.19	Incentive	young ($p=0.003$)
21	It automatically adds information to the patient's electronic records	1.86	Incentive	young ($p=0.01$)
20	It allows me to quickly share an examination result with colleagues	1.80	Incentive	-
1	It makes a diagnosis for which I am not competent	1.71	Incentive	young ($p=0.008$)
2	It makes a diagnosis for which no specialist is accessible within an acceptable time or distance	1.59	Incentive	-
4	It saves me time	1.47	Incentive	young ($p=0.004$) men ($p=0.02$)
11	It allows me to manage the patient entirely in the office without handing over	1.40	Incentive	-
8	It allows me to predict a risk category of the patient	1.26	Incentive	-
6	It detects a specific pathology in an asymptomatic subject	1.17	Incentive	young ($p=0.04$)
14	It allows me to involve the patient in the therapeutic management	1.10	Incentive	-
32	It reassures the patient	1.10	Incentive	young ($p=0.01$)
22	It sometimes provides me with training	1.09	Incentive	young ($p=0.0002$)
15	It removes me from a situation of doubt	0.99	Incentive	young ($p=0.003$)
12	It allows me to adapt the treatment to a specific sub-category of patients (personalized medicine)	0.98	Incentive	young ($p=0.01$) urban ($p=0.05$)
33	It allows me to cut short abusive requests from the patient (e.g.: certificates, work stoppages, etc.)	0.87	Incentive	young ($p=0.01$) men ($p=0.02$)
5	It simplifies a procedure that I already know how to perform	0.84	Incentive	young ($p=0.03$) urban ($p=0.02$)
29	It increases the time spent with the patient	0.72	Incentive	-
7	It allows me to delegate a task to the secretary	0.71	Incentive	-
26	It is possible to capture the information (e.g. image) and then analyze it asynchronously	0.51	Neutral	young ($p=0.04$)
25	It shows me how to capture information (e.g. positioning a probe)	0.37	Neutral	-
13	It has a good reputation with patient associations	0.24	Neutral	-
31	It is very expensive but allows me to be profitable after a few months	0.21	Neutral	young ($p=0.003$)
19	It reflects a "high tech" image of my practice	0.09	Neutral	-
30	It reduces time spent with the patient	0.04	Neutral	-
28	It requires specific training from the manufacturer	-0.23	Neutral	-
18	It entertains the patient	-0.26	Neutral	-
24	It uses exclusively my personal smartphone	-0.36	Neutral	non-urban ($p=0.03$)
17	It gives information on elements for which it has not been requested (digression faculty)	-0.42	Neutral	-
27	It can be used freely in the waiting room by patients and their companions	-0.98	Deterrent	-
23	It automatically triggers an emergency call in some cases	-1.05	Deterrent	men ($p=0.03$)
16	It decides automatically without my intervention	-1.29	Deterrent	women ($p=0.01$)
3	It makes a diagnosis for which I am already competent	-1.39	Deterrent	-

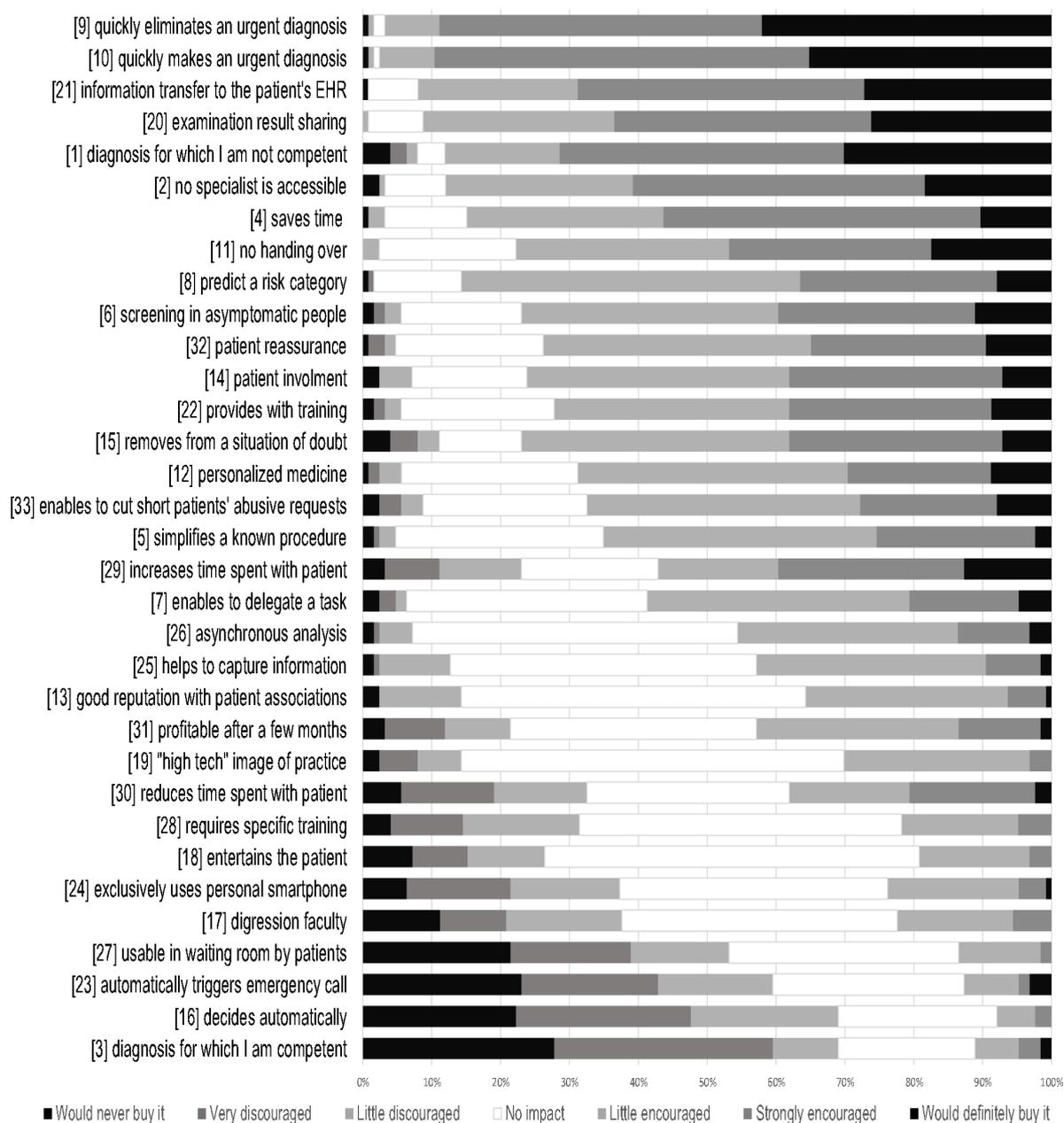


Figure 2. Results of Likert scales: "Would this feature encourage or discourage you from purchasing this module?". For original labeling, refer to Table 2.

4 Discussion

We conducted a survey among GPs to find out what criteria would make them buy or refuse to buy a device with an AI module that could be used to support care.

As a consequence of the random draw and the high response rate, the characteristics of the surveyed population show a real representativeness of the French GP population.

Regarding the results, first criteria were in relation with the disease the module could enable to diagnostic: GPs expect such a module to diagnose or eliminate an urgent pathology for which they are not competent and for which specialists are difficult to access. They do not want confirmation of a diagnosis that they are already able to make. They are not specifically interested on digression ability.

There is also a strong demand from GPs for interoperability, with results integrated into the patient's EHR and easy transmission to colleagues.

GPs would like such modules to make them save time and simplify some procedures, but they do not want such modules to automatically make a call or any action without confirmation.

As far as relations with the patient are concerned, GPs expect such a module to allow them to associate the patient with the care, to reassure him or her, but do not pay particular attention to the image that the module sends back to the patient, nor to what the patient can do with it without the doctor. They are also looking for modules that allow them to personalize the care. Interestingly, GPs also want to be able to rely on the neutrality of a machine to cut off abusive requests, such as work stoppages or certificates of convenience.

5 Conclusion

We believe that, in addition to taking advantage of technical opportunities enabled by research progress, manufacturers should also prioritize AI module developments according to what endusers, especially general practitioners, expect from future AI modules.

Conclusion

La présente étude nous permet de montrer que, en plus de s'appuyer sur les résultats obtenus par les équipes de recherche en intelligence artificielle, les fabricants devraient tenir compte des attentes des praticiens auxquels ils destinent les modules d'IA. Leurs attentes sont en particulier indissociables du flux de prise en charge des patients (*workflow*), de la notion d'urgence et des besoins de transmission et d'intégration des résultats.

Liste des tables

Table 1. Devices with AI that GPs already used	17
Table 2. Results analysis of Likert scales: "Would this feature encourage or discourage you from purchasing this module?"	19

Liste des figures

Figure 1. Flowchart	16
Figure 2. Results of Likert scales: “Would this feature encourage or discourage you from purchasing this module?”. For original labeling, refer to Table 2.	20

Références

Annexe 1 : courrier envoyé aux participants



Docteur,

Vous avez été tiré au sort pour participer à une étude. Nous souhaiterions connaître votre avis sur les **modules automatisés d'aide au diagnostic ou à la décision thérapeutique**. Ces modules sont de plus en plus répandus dans la pratique quotidienne (ex : interpréteurs automatisés d'ECG), mais les industriels qui les développent semblent parfois peu à l'écoute des besoins des Médecins Généralistes.

Nous souhaiterions comprendre vos attentes concernant ce sujet. Accepteriez-vous de remplir le questionnaire ci-joint (temps moyen : 4 minutes et 50 secondes) ? Il est entièrement anonyme.

Une fois l'ensemble du questionnaire rempli, merci de le glisser dans l'**enveloppe retour préaffranchie** et de nous le renvoyer. Si vous souhaitez recevoir un exemplaire numérique de nos résultats dans quelques mois, merci de m'envoyer un message à smain.tabla@gmail.com (mais renvoyez quand même le questionnaire par la poste indépendamment, afin de préserver son caractère anonyme).

Chacune des réponses retournées apportera une pierre à l'édifice de notre recherche. Je vous suis d'avance très reconnaissant pour le temps que vous me consacrerez.

Smaïn TABLA

Annexe 2 : questionnaire

Début du questionnaire :

Quel âge avez-vous ? ans

Quel est votre mode d'exercice principal ? Urbain Semi-rural Rural

Quel est votre sexe ? Femme Homme

Vous exercez principalement : Seul au cabinet En cabinet de groupe Autre

Parmi les appareils suivants, lesquels avez-vous déjà eu à disposition en consultation ? (plusieurs réponses possibles. Laissez vide si aucun de ces appareils.)

- ECG avec interprétation automatisée
- Stéthoscope intelligent connecté
- Tensiomètre connecté
- Saturomètre connecté
- Glucomètre connecté
- Thermomètre connecté
- Diagnostic de pathologie dermatologique (skindiag)
- Défibrillateur semi-automatique ou automatique
- Pèse-personne connecté

Imaginons dans le futur un **appareil hypothétique disposant d'un module d'aide au diagnostic** ou à la décision thérapeutique (ex : ECG ou otoscope avec diagnostic automatique). Ceci exclut les sites web et les analyses d'échantillon de biologie médicale. Pour chacune des caractéristiques suivantes, en supposant qu'elle s'applique à cet appareil hypothétique, **vous inciterait-elle ou non à acquérir un tel appareil ?**

Concernant cet appareil hypothétique...	Je n'en ferai JAMAIS l'acquisition	Cela me dissuade beaucoup	Cela me dissuade un peu	Cela n'a pas d'impact	Cela m'incite un peu	Cela m'incite fortement	J'en ferai A COUP SÛR l'acquisition
Il pose un diagnostic pour lequel je ne suis pas compétent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il pose un diagnostic pour lequel aucun spécialiste n'est accessible dans un temps ou une distance acceptable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il pose un diagnostic pour lequel je suis déjà compétent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il me fait gagner du temps	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il simplifie un geste que je sais déjà réaliser	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il dépiste une pathologie spécifique chez un sujet asymptomatique	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il me permet de déléguer une tâche au secrétariat	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il permet de prédire une catégorie de risque du patient	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il élimine rapidement un diagnostic urgent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il pose rapidement un diagnostic urgent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il permet de prendre en charge le patient intégralement au cabinet sans passer la main	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Concernant cet appareil hypothétique...	Je n'en ferai JAMAIS l'acquisition	Cela me dissuade beaucoup	Cela me dissuade un peu	Cela n'a pas d'impact	Cela m'incite un peu	Cela m'incite fortement	J'en ferai A COUP SUR l'acquisition
Il permet d'adapter le traitement à une sous-catégorie spécifique de patients (médecine personnalisée)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il a bonne réputation auprès des associations de patients	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il permet d'associer le patient à la prise en charge thérapeutique	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il m'extrait d'une situation de doute	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il décide automatiquement sans mon intervention	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il donne des indications sur des éléments pour lesquels il n'a pas été sollicité (faculté de digression)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il divertit le patient	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il renvoie une image « high Tech » de mon cabinet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il me permet rapidement de partager un résultat d'examen avec des confrères	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il ajoute automatiquement une information au dossier électronique du patient	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il me délivre parfois une formation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il déclenche automatiquement un appel au 15 dans certains cas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il utilise exclusivement mon smartphone personnel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il m'indique comment réaliser la capture d'information (ex : positionner une sonde)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il est possible dans un premier temps de capturer l'information (ex : image), et dans un deuxième temps (asynchrone) d'analyser l'information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il peut être utilisé librement en salle d'attente par les patients et leurs accompagnants	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il nécessite une formation spécifique délivrée par le constructeur	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il augmente le temps passé avec le patient	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il diminue le temps passé avec le patient	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il coûte très cher mais me permet d'être bénéficiaire au bout de quelques mois	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il rassure le patient	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Il permet de couper court à des demandes abusives du patient (ex : certificats, arrêts de travail, etc.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Merci d'avoir pris le temps de répondre au questionnaire.

AUTEUR : Nom : TABLA

Prénom : Smaïn

Date de Soutenance : 21/09/2021

Titre de la Thèse : Modules d'intelligence artificielle au cabinet de ville : qu'en attendent les médecins généralistes français ?

Thèse - Médecine - Lille 2021

Cadre de classement : Médecine Générale

DES + spécialité : Médecine Générale

Mots-clés : Intelligence artificielle, interpréteurs automatisés, utilisabilité.

Résumé :

Contexte : Le développement de modules d'intelligence artificielle (IA) doit s'appuyer sur les progrès techniques, mais aussi sur les besoins des utilisateurs. Notre objectif est d'identifier les critères qui rendent un hypothétique module d'IA désirable pour les médecins généralistes (MG).

Matériel et Méthodes : Nous avons tiré au sort 200 médecins généralistes français, et leur avons envoyé un questionnaire papier avec enveloppe réponse préimprimée.

Résultats : La population des répondants était représentative de la population des médecins généralistes français. Les médecins généralistes attendent des modules d'IA qu'ils diagnostiquent ou éliminent une pathologie urgente pour laquelle ils ne sont pas compétents et pour laquelle des spécialistes ne sont pas disponibles. Ils demandent également une interopérabilité, une intégration automatisée des résultats dans les dossiers médicaux électroniques et un partage facilité des informations avec leurs confrères. Les médecins généralistes souhaitent que les modules d'IA leur fassent gagner du temps, simplifient certaines procédures et leur permettent de déléguer des tâches au secrétariat. Ils attendent des modules d'IA qu'ils leur permettent d'associer le patient aux soins, de le rassurer et de personnaliser les soins. Il est intéressant de noter que les médecins généralistes compteraient également sur de tels modules pour couper court aux demandes abusives, comme certains arrêts de travail ou les certificats de complaisance. En revanche, les MG rejetteraient des modules qui agiraient automatiquement à leur place, ou confirmeraient un diagnostic qu'ils savent déjà porter.

Conclusion : les fabricants de dispositifs médicaux équipés d'un module d'IA devraient tenir compte des attentes des praticiens de terrain.

Composition du Jury :

Président : Monsieur le Professeur Emmanuel Chazard

Asseseurs :

Madame le Docteur Sabine BAYEN

Madame le Docteur Gabrielle LISEMBARD

Monsieur le Docteur Pierre BALAYÉ