

**THESE
POUR LE DIPLOME D'ETAT
DE DOCTEUR EN PHARMACIE**

**Soutenue publiquement le 25/08/2025
Par M. DEMAILLY ALEXANDRE PIERRE ELIE**

**Prise de décision d'investissement dans les biotechs : vers une disruption par
l'Intelligence Artificielle ?**

L'« evidence-based investment » à la portée des fonds de Venture Capital

Membres du jury :

Président & Directeur de thèse : M. MORGENROTH THOMAS, Maître de
Conférence des Universités en Droit et d'Economie Pharmaceutique – UFR3S
Pharmacie

Assesseur(s) : Mme. PINÇON CLAIRE, Maître de Conférence des Universités en
Biomathématiques – UFR3S Pharmacie

Membre extérieur : Mme. FOUTEL VERONIQUE, CEO – InBrain Pharma

Membre extérieur : M. MARCHAND LAURENT, Responsable du programme
scientifique – Paris Santé Campus

Université de Lille

Président
Premier Vice-président
Vice-présidente Formation
Vice-président Recherche
Vice-président Ressources Humaine
Directrice Générale des Services

Régis BORDET
Bertrand DÉCAUDIN
Corinne ROBACZEWSKI
Olivier COLOT
Jean-Philippe TRICOIT
Anne-Valérie CHIRIS-FABRE

UFR3S

Doyen
Premier Vice-Doyen, Vice-Doyen RH, SI et Qualité
Vice-Doyenne Recherche
Vice-Doyen Finances et Patrimoine
Vice-Doyen International
Vice-Doyen Coordination pluriprofessionnelle et Formations sanitaires
Vice-Doyenne Formation tout au long de la vie
Vice-Doyen Territoire-Partenariats
Vice-Doyen Santé numérique et Communication
Vice-Doyenne Vie de Campus
Vice-Doyen étudiant

Dominique LACROIX
Hervé HUBERT
Karine FAURE
Emmanuelle LIPKA
Vincent DERAMECOURT
Sébastien D'HARANCY
Caroline LANIER
Thomas MORGENROTH
Vincent SOBANSKI
Anne-Laure BARBOTIN
Victor HELENA

Faculté de Pharmacie

Vice - Doyen
Premier Assesseur et
Assesseur à la Santé et à l'Accompagnement
Assesseur à la Vie de la Faculté et
Assesseur aux Ressources et Personnels
Responsable de l'Administration et du Pilotage
Représentant étudiant
Chargé de mission 1er cycle
Chargée de mission 2eme cycle
Chargé de mission Accompagnement et Formation à la Recherche
Chargé de mission Relations Internationales
Chargée de Mission Qualité
Chargé de mission dossier HCERES

Pascal ODOU

Anne GARAT

Emmanuelle LIPKA
Cyrille PORTA
Honoré GUISE
Philippe GERVOIS
Héloïse HENRY
Nicolas WILLAND
Christophe FURMAN
Marie-Françoise ODOU
Réjane LESTRELIN

Professeurs des Universités - Praticiens Hospitaliers (PU-PH)

Civ.	Nom	Prénom	Service d'enseignement	Section CNU
Mme	ALLORGE	Delphine	Toxicologie et Santé publique	81
M.	BROUSSEAU	Thierry	Biochimie	82
M.	DÉCAUDIN	Bertrand	Biopharmacie, Pharmacie galénique et hospitalière	81
M.	DINE	Thierry	Pharmacologie, Pharmacocinétique et Pharmacie clinique	81
Mme	DUPONT-PRADO	Annabelle	Hématologie	82
Mme	GOFFARD	Anne	Bactériologie - Virologie	82
M.	GRESSIER	Bernard	Pharmacologie, Pharmacocinétique et Pharmacie clinique	81
M.	ODOU	Pascal	Biopharmacie, Pharmacie galénique et hospitalière	80
Mme	POULAIN	Stéphanie	Hématologie	82
M.	SIMON	Nicolas	Pharmacologie, Pharmacocinétique et Pharmacie clinique	81
M.	STAELS	Bart	Biologie cellulaire	82

Professeurs des Universités (PU)

Civ.	Nom	Prénom	Service d'enseignement	Section CNU
M.	ALIOUAT	El Moukhtar	Parasitologie - Biologie animale	87
Mme	ALIOUAT	Cécile-Marie	Parasitologie - Biologie animale	87
Mme	AZAROUAL	Nathalie	Biophysique - RMN	85
M.	BERLARBI	Karim	Physiologie	86
M.	BERTIN	Benjamin	Immunologie	87
M.	BLANCHEMAIN	Nicolas	Pharmacotechnie industrielle	85
M.	CARNOY	Christophe	Immunologie	87
M.	CAZIN	Jean-Louis	Pharmacologie, Pharmacocinétique et Pharmacie clinique	86
M.	CUNY	Damien	Sciences végétales et fongiques	87
Mme	DELBAERE	Stéphanie	Biophysique - RMN	85

Mme	DEPREZ	Rebecca	Chimie thérapeutique	86
M.	DEPREZ	Benoît	Chimie bio inorganique	85
Mme	DUMONT	Julie	Biologie cellulaire	87
M.	ELATI	Mohamed	Biomathématiques	27
M.	FOLIGNÉ	Benoît	Bactériologie - Virologie	87
Mme	FOULON	Catherine	Chimie analytique	85
M.	GARÇON	Guillaume	Toxicologie et Santé publique	86
M.	GOOSSENS	Jean-François	Chimie analytique	85
M.	HENNEBELLE	Thierry	Pharmacognosie	86
M.	LEBEGUE	Nicolas	Chimie thérapeutique	86
M.	LEMDANI	Mohamed	Biomathématiques	26
Mme	LESTAVEL	Sophie	Biologie cellulaire	87
Mme	LESTRELIN	Réjane	Biologie cellulaire	87
Mme	LIPKA	Emmanuelle	Chimie analytique	85
Mme	MELNYK	Patricia	Chimie physique	85
M.	MILLET	Régis	Institut de Chimie Pharmaceutique Albert Lespagnol	86
M.	MOREAU	Pierre-Arthur	Sciences végétales et fongiques	87
Mme	MUHR-TAILLEUX	Anne	Biochimie	87
Mme	PERROY	Anne-Catherine	Droit et Economie pharmaceutique	86
Mme	RIVIÈRE	Céline	Pharmacognosie	86
Mme	ROMOND	Marie-Bénédicte	Bactériologie - Virologie	87
Mme	SAHPAZ	Sevser	Pharmacognosie	86
M.	SERGHERAERT	Éric	Droit et Economie pharmaceutique	86
M.	SIEPMANN	Juergen	Pharmacotechnie industrielle	85
Mme	SIEPMANN	Florence	Pharmacotechnie industrielle	85
M.	WILLAND	Nicolas	Chimie organique	86

Maîtres de Conférences - Praticiens Hospitaliers (MCU-PH)

Civ.	Nom	Prénom	Service d'enseignement	Section CNU
Mme	CUVELIER	Élodie	Pharmacologie, Pharmacocinétique et Pharmacie clinique	81
Mme	DANEL	Cécile	Chimie analytique	85
Mme	DEMARET	Julie	Immunologie	82
Mme	GARAT	Anne	Toxicologie et Santé publique	81
Mme	GENAY	Stéphanie	Biopharmacie, Pharmacie galénique et hospitalière	81
Mme	GILLIOT	Sixtine	Biopharmacie, Pharmacie galénique et hospitalière	80
M.	GRZYCH	Guillaume	Biochimie	82
Mme	HENRY	Héloïse	Biopharmacie, Pharmacie galénique et hospitalière	80
M.	LANNOY	Damien	Biopharmacie, Pharmacie galénique et hospitalière	80
Mme	MASSE	Morgane	Biopharmacie, Pharmacie galénique et hospitalière	81
Mme	ODOU	Marie-Françoise	Bactériologie - Virologie	82

Maîtres de Conférences des Universités (MCU)

Civ.	Nom	Prénom	Service d'enseignement	Section CNU
M.	ANTHÉRIEU	Sébastien	Toxicologie et Santé publique	86
M.	BANTUBUNGI-BLUM	Kadiombo	Biologie cellulaire	87
M.	BERTHET	Jérôme	Biophysique - RMN	85
M	BEDART	Corentin	ICPAL	86
M.	BOCHU	Christophe	Biophysique - RMN	85
M.	BORDAGE	Simon	Pharmacognosie	86
M.	BOSC	Damien	Chimie thérapeutique	86
Mme	BOU KARROUM	Nour	Chimie bioinorganique	
M.	BRIAND	Olivier	Biochimie	87
Mme	CARON-HOUDE	Sandrine	Biologie cellulaire	87
Mme	CARRIÉ	Hélène	Pharmacologie, Pharmacocinétique et Pharmacie clinique	86

Mme	CHABÉ	Magali	Parasitologie - Biologie animale	87
Mme	CHARTON	Julie	Chimie organique	86
M.	CHEVALIER	Dany	Toxicologie et Santé publique	86
Mme	DEMANCHE	Christine	Parasitologie - Biologie animale	87
Mme	DEMARQUILLY	Catherine	Biomathématiques	85
M.	DHIFLI	Wajdi	Biomathématiques	27
M.	EL BAKALI	Jamal	Chimie thérapeutique	86
M.	FARCE	Amaury	Institut de Chimie Pharmaceutique Albert Lespagnol	86
M.	FLIPO	Marion	Chimie organique	86
M.	FRULEUX	Alexandre	Sciences végétales et fongiques	
M.	FURMAN	Christophe	Institut de Chimie Pharmaceutique Albert Lespagnol	86
M.	GERVOIS	Philippe	Biochimie	87
Mme	GOOSSENS	Laurence	Institut de Chimie Pharmaceutique Albert Lespagnol	86
Mme	GRAVE	Béatrice	Toxicologie et Santé publique	86
M.	HAMONIER	Julien	Biomathématiques	26
Mme	HAMOUDI-BEN YELLES	Chérifa-Mounira	Pharmacotechnie industrielle	85
Mme	HANNOTHIAUX	Marie-Hélène	Toxicologie et Santé publique	86
Mme	HELLEBOID	Audrey	Physiologie	86
M.	HERMANN	Emmanuel	Immunologie	87
M.	KAMBIA KPAKPAGA	Nicolas	Pharmacologie, Pharmacocinétique et Pharmacie clinique	86
M.	KARROUT	Younes	Pharmacotechnie industrielle	85
Mme	LALLOYER	Fanny	Biochimie	87
Mme	LECOEUR	Marie	Chimie analytique	85
Mme	LEHMANN	Hélène	Droit et Economie pharmaceutique	86
Mme	LELEU	Natascha	Institut de Chimie Pharmaceutique Albert Lespagnol	86
M.	LIBERELLE	Maxime	Biophysique - RMN	
Mme	LOINGEVILLE	Florence	Biomathématiques	26

Mme	MARTIN	Françoise	Physiologie	86
M.	MARTIN MENA	Anthony	Biopharmacie, Pharmacie galénique et hospitalière	
M.	MENETREY	Quentin	Bactériologie - Virologie	87
M.	MORGENROTH	Thomas	Droit et Economie pharmaceutique	86
Mme	MUSCHERT	Susanne	Pharmacotechnie industrielle	85
Mme	NIKASINOVIC	Lydia	Toxicologie et Santé publique	86
Mme	PINÇON	Claire	Biomathématiques	85
M.	PIVA	Frank	Biochimie	85
Mme	PLATEL	Anne	Toxicologie et Santé publique	86
M.	POURCET	Benoît	Biochimie	87
M.	RAVAUX	Pierre	Biomathématiques / Innovations pédagogiques	85
Mme	RAVEZ	Séverine	Chimie thérapeutique	86
Mme	ROGEL	Anne	Immunologie	
M.	ROSA	Mickaël	Hématologie	87
M.	ROUMY	Vincent	Pharmacognosie	86
Mme	SEBTI	Yasmine	Biochimie	87
Mme	SINGER	Elisabeth	Bactériologie - Virologie	87
Mme	STANDAERT	Annie	Parasitologie - Biologie animale	87
M.	TAGZIRT	Madjid	Hématologie	87
M.	VILLEMAGNE	Baptiste	Chimie organique	86
M.	WELTI	Stéphane	Sciences végétales et fongiques	87
M.	YOUS	Saïd	Chimie thérapeutique	86
M.	ZITOUNI	Djamel	Biomathématiques	85

Professeurs certifiés

Civ.	Nom	Prénom	Service d'enseignement
Mme	FAUQUANT	Soline	Anglais
M.	HUGES	Dominique	Anglais
Mme	KUBIK	Laurence	Anglais
M.	OSTYN	Gaël	Anglais

Professeurs Associés

Civ.	Nom	Prénom	Service d'enseignement	Section CNU
M.	BAILLY	Christian	ICPAL	86
M.	DAO PHAN	Haï Pascal	Chimie thérapeutique	86
M.	DHANANI	Alban	Droit et Economie pharmaceutique	86

Maîtres de Conférences Associés

Civ.	Nom	Prénom	Service d'enseignement	Section CNU
M	AYED	Elya	Pharmacie officinale	
M.	COUSEIN	Etienne	Biopharmacie, Pharmacie galénique et hospitalière	
Mme	CUCCHI	Malgorzata	Biomathématiques	85
Mme	DANICOURT	Frédérique	Pharmacie officinale	
Mme	DUPIRE	Fanny	Pharmacie officinale	
M.	DUFOSSEZ	François	Biomathématiques	85
M.	FRIMAT	Bruno	Pharmacologie, Pharmacocinétique et Pharmacie clinique	85
Mme	GEILER	Isabelle	Pharmacie officinale	
M.	GILLOT	François	Droit et Economie pharmaceutique	86
M.	MITOUMBA	Fabrice	Biopharmacie, Pharmacie galénique et hospitalière	86
M.	PELLETIER	Franck	Droit et Economie pharmaceutique	86

M	POTHIER	Jean-Claude	Pharmacie officinale	
Mme	ROGNON	Carole	Pharmacie officinale	

Assistants Hospitalo-Universitaire (AHU)

Civ.	Nom	Prénom	Service d'enseignement	Section CNU
M.	BOUDRY	Augustin	Biomathématiques	
Mme	DERAMOUDT	Laure	Pharmacologie, Pharmacocinétique et Pharmacie clinique	
M.	GISH	Alexandr	Toxicologie et Santé publique	
Mme	NEGRIER	Laura	Chimie analytique	

Hospitalo-Universitaire (PHU)

	Nom	Prénom	Service d'enseignement	Section CNU
M.	DESVAGES	Maximilien	Hématologie	
Mme	LENSKI	Marie	Toxicologie et Santé publique	

Attachés Temporaires d'Enseignement et de Recherche (ATER)

Civ.	Nom	Prénom	Service d'enseignement	Section CNU
Mme	BERNARD	Lucie	Physiologie	
Mme	BARBIER	Emeline	Toxicologie	
Mme	COMPAGNE	Nina	Chimie Organique	
Mme	COULON	Audrey	Pharmacologie, Pharmacocinétique et Pharmacie clinique	
M.	DUFOSSEZ	Robin	Chimie physique	
Mme	FERRY	Lise	Biochimie	
M	HASYEOUI	Mohamed	Chimie Organique	
Mme	HENRY	Doriane	Biochimie	
Mme	KOUAGOU	Yolène	Sciences végétales et fongiques	

M	LAURENT	Arthur	Chimie-Physique	
M.	MACKIN MOHAMOUR	Synthia	Biopharmacie, Pharmacie galénique et hospitalière	
Mme	RAAB	Sadia	Physiologie	

Enseignant contractuel

Civ.	Nom	Prénom	Service d'enseignement
Mme	DELOBEAU	Iris	Pharmacie officinale
M	RIVART	Simon	Pharmacie officinale
Mme	SERGEANT	Sophie	Pharmacie officinale
M.	ZANETTI	Sébastien	Biomathématiques

LRU / MAST

Civ.	Nom	Prénom	Service d'enseignement
Mme	FRAPPE	Jade	Pharmacie officinale
M	LATRON-FREMEAU	Pierre-Manuel	Pharmacie officinale
M.	MASCAUT	Daniel	Pharmacologie, Pharmacocinétique et Pharmacie clinique

CYCLE DE VIE DU DOCUMENT

Version	Modifié par	Date	Principales modifications
1.0		20/02/2020	Création
2.0		02/01/2022	Mise à jour
2.1		21/06/2022	Mise à jour
2.2		01/02/2024	Mise à jour
2.3		15/11/2024	Mise à jour
2.4		18/02/2025	Mise à jour

Table des matières

Table des illustrations	15
A] Introduction :	20
I] La technologie comme solution à la crise d'efficacité de l'industrie pharmaceutique ?	20
II] Pourquoi s'intéresser spécifiquement à l'investissement en Biotech ?	23
III] Structure et objectifs de la thèse	25
B] Qu'est-ce que l'Intelligence Artificielle ?	26
I] Intelligence Artificielle : Définitions et concepts	26
A) L'intelligence artificielle faible	26
B) L'intelligence artificielle forte	27
C) L'intelligence artificielle symbolique et l'intelligence artificielle connexionniste	28
D) L'apprentissage automatique :	30
E) Quelques exemples d'algorithmes utilisés en Intelligence Artificielle :	32
F) Focus sur le Deep Learning :	32
G) La révolution de l'intelligence artificielle générative :	32
II] Une histoire de l'Intelligence Artificielle : des hommes et des concepts	33
A) Entre mythes et légendes, une idée vieille comme l'humanité ?	33
B) La mécanisation de la pensée humaine :	34
C) La naissance de l'intelligence artificielle :	34
D) Le premier grand hiver de l'IA :	35
E) Le retour de l'IA, portée par l'industrie :	35
F) Le second grand hiver de l'IA :	35
G) Le retour progressif de l'IA depuis les années 1990 :	35
H) L'émergence de l'intelligence artificielle générative	36
III] L'intelligence artificielle en pratique : les étapes de développement d'un modèle	36
A) La collecte et la structuration des données brutes :	37
B) Le prétraitement des données :	37
C) La constitution d'un jeu de données préparé :	37
D) L'entraînement des algorithmes de Machine Learning :	37
E) La sélection et la validation du modèle :	38
F) Le déploiement et l'intégration dans les applications métiers :	38
G) Le post-déploiement du modèle :	38
H) Les contraintes spécifiques à l'implémentation de l'intelligence artificielle :	38
C] Qu'est-ce qu'un fonds de Venture Capital ?	40
I] Définition du Venture Capital et comparaison avec les autres formes de financements	40
A) Historique du capital-risque et évolution jusqu'à nos jours	41

II] Fonctionnement des fonds de Venture Capital	42
III] Les particularités des fonds de Venture Capital dédiés à la biotech	47
D] Utilisation de l'Intelligence Artificielle dans le domaine de l'investissement : un état des lieux	52
E] Etude de l'utilisation de l'IA dans les fonds d'investissement biotech	59
I] Méthodologie et élaboration d'un questionnaire :	59
II] Le questionnaire	60
III] Les résultats du questionnaire	64
F] IA et Propriété intellectuelle dans le contexte des VC Biotechs :	76
I] Introduction et enjeux de l'analyse de brevets dans les décisions d'investissement en biotech.....	76
A) Importance de l'analyse des brevets	76
B) Défis spécifiques liés au domaine de la biotech	77
II] Applications de l'IA dans les analyses de brevets	77
A) Analyses sémantiques et extractions d'informations	77
B) Cartographie des technologies et analyse des tendances	78
C) Evaluation de la qualité des brevets	79
D) Détection des conflits et risques juridiques.....	79
E) Simulation de stratégies de propriété intellectuelle	79
III] Intégration de l'IA & bénéfices et limites de son utilisation par les VC dans l'analyse des brevets biotech.....	80
G] IA & Données pré-cliniques	84
I] Introduction	84
II] Les données disponibles	84
III] Les critères d'évaluation de la pertinence des études menées et des données générées ..	84
A) Evaluation de la pertinence de la cible thérapeutique	84
B) Evaluation de la pertinence de la molécule	85
C) Evaluation de la pertinence translationnelle.....	85
D) Robustesse statistique.....	86
E) Adhérence aux guidelines.....	86
IV] Les apports possibles de l'IA	86
V] L'émergence des agents IA scientifiques	87
H] Utilisation de l'IA pour prédire le succès des études cliniques	90
I] Les données utilisables : quelles données pour nourrir les algorithmes ?	90
II] Evolution des modèles d'IA et performances publiées	91
III] Quelle (future) place des agents IA dans la prédiction des résultats cliniques ?	92
IV] Bénéfices et limites.....	94

I] Conclusion et Perspectives	96
I] Le no-code comme moteur d'accélération de l'intégration de l'IA au sein des fonds d'investissement	96
II] Mise en place de la transformation data-driven d'un fonds de capital risque : par où commencer ? Quelle feuille de route ?.....	100
A) Alignement initial et préparation	100
B) Prototypage no-code des premiers cas d'usage	101
C) Extension, itérations et décision « make or buy ».....	101
D) Structuration à moyen terme et changement d'échelle	102
III] Un VC Data-Driven, et après ?.....	102
A) L'émergence des nouveaux VC Data-Driven : les Quant-VC	103
B) Quelle stratégie d'évaluation des approches data driven ?	103
C) Vers un co-pilote des VC Biotech ?	105
J] Annexe : L'intégration de l'IA aux Venture Builders Biotech	108
K] Annexe 2 : L'utilisation de l'IA pour sourcer et scorer des assets inemployés ou disponibles à la vente	111

Table des illustrations

Figure 1 - Evolutions des retours financiers de la R&D pharmaceutique	21
Figure 2 - Diminution de la productivité pharmaceutique à travers le temps	22
Figure 3 - Différence entre l'IA faible et l'IA forte	27
Figure 4 - Différences entre IA Symbolique et Connexionniste	28
Figure 5 - Fonctionnement de l'IA symbolique.....	29
Figure 6 - Apprentissage automatique & IA.....	30
Figure 7 - Les différents types d'IA connexionnistes.....	30
Figure 8 - Développement d'un modèle d'IA	37
Figure 9 - Différents types de financements de l'innovation	41
Figure 10 - Organisation d'une société de gestion avec différents fonds successeurs	43
Figure 11 - Illustration d'un processus de sélection et d'analyse	44
Figure 12 - Illustration du fonctionnement d'un fonds de Venture Capital.....	47
Figure 13 - Part de la biotech dans la totalité des financements VC.....	49
Figure 14 - Cas d'usage de l'IA chez les VC data-driven (non-spécifique de la biotech) – Panorama Data-Driven VC 2025 – Andre Retterath	53
Figure 15 - Niveau d'adoption interne de la solution IA développée – Panorama Data-Driven VC 2025 – Andre Retterath	54
Figure 16 - Principales problématiques liées à l'usage des LLMs - Panorama Data-Driven VC 2025 – Andre Retterath	54
Figure 17 - Estimation du cycle de hype par les VC - Panorama Data-Driven VC 2025 – Andre Retterath.....	56
Figure 18 - Panorama des fonds VC data-driven - Panorama Data-Driven VC 2025 – Andre Retterath.....	56
Figure 19 - Illustration de la base de données constituée (fonds)	59
Figure 20 - Illustration de la base de données constituée (investisseurs).....	60
Figure 21 - Rôle du répondant	65
Figure 22 - Nature du fonds du répondant.....	66
Figure 23 - Typologie d'intervention des fonds des répondants	67
Figure 24 - Verticales des fonds des répondants	68
Figure 25 – Atteinte des objectifs en termes de deal sourcing	68
Figure 26 - Atteinte des objectifs en termes de deal-closing	69
Figure 27 - Atteinte des objectifs financiers.....	69
Figure 28 - Evaluation de l'amélioration possible des métriques financières.....	70
Figure 29 - Evaluation des connaissances en matières d'utilisation de l'IA en VC.....	70
Figure 30 - Evaluation des bases de données Biotech	71
Figure 31 - L'IA fera-t-elle partie de l'avenir du VC ?	71
Figure 32 - Pensez-vous que l'IA permettra d'améliorer les perf. de votre fonds ?	72
Figure 33 - Evaluation de la compréhension de l'IA par les VC.....	72
Figure 34 - Evaluation de la présence d'une équipe data au sein du fonds VC	73
Figure 35 - Les raisons de l'absence d'une équipe data (par % de répondants)	73
Figure 36 - Les domaines d'intervention des équipes data	74
Figure 37 - Avez-vous déjà collaboré avec des entreprises proposant des services d'analyse data pour les VC ?	74

Figure 38 - Envisagez-vous, en fonction du développement de l'offre, d'utiliser des outils d'aide à la décision ?	75
Figure 39- Quelles fonctionnalités d'un tel outil seraient les plus importantes ?	75
Figure 40 - Caractéristiques de l'outil Octimine.....	78
Figure 41 - Schéma de fonctionnement de l'outil de cartographie et d'identification de l'état de l'art de LexisNexis (issu du site web de la société).....	79
Figure 42 - L'utilisation de l'IA en IP devient de plus en plus populaire	80
Figure 43 - Schéma de fonctionnement d'un agent IA	87
Figure 44 - Schéma de fonctionnement du système multi-agents de Google (issu de la publication pré-citée)	88
Figure 45 - Les 4 agents autonomes de FutureHouse	89
Figure 46 - Stades d'utilisation d'un outil de prédiction des études cliniques.....	90
Figure 47 - Performance du modèle de prédiction du succès des études cliniques.....	92
Figure 48 - Schéma de fonctionnement du système multi-agents de prédiction du succès de l'étude clinique(50).....	94
Figure 49 - Proposition d'un système multi-agents pouvant être déployé en no-code	97
Figure 50 - Exemple d'un tableau de bord VC data-driven (généré par IA)	106
Figure 51 - Les raisons de l'abandon du développement d'un candidat médicament	113
Figure 52 - Proposition d'organisation d'un Venture Builder hybride	115

Remerciements

Je remercie chaleureusement les membres de mon jury pour leur accompagnement et leur confiance. Je suis très heureux de vous avoir rencontré et vous remercie pour nos nombreuses discussions. Puissent nos chemins se croiser encore à de nombreuses reprises !

Je remercie la Faculté de Pharmacie de l'Université de Lille pour la qualité de ses enseignements, pour les différentes opportunités qu'elle a pu m'apporter, pour les belles rencontres qu'elle a permises et pour son soutien dans les différentes initiatives que j'ai pu porter durant mon cursus.

Je remercie mes collègues et employeurs, présents et passés, pour leur confiance et pour tout ce qu'ils ont apporté et apporteront encore à la construction de mon avenir professionnel.

Je remercie avec beaucoup d'affection ma famille – et tout particulièrement ma mère – pour son soutien et son amour constants depuis le début de ma vie et de mes études. J'espère que ce travail, pour le long chemin scolaire parcouru qu'il symbolise et pour l'avenir professionnel qu'il augure, saura les rendre fiers.

Je remercie mes amis et mes proches qui contribuent à faire de ma vie une belle et joyeuse aventure. Puissent les liens de l'amitié ne jamais se distendre.

سپاس از آن دوشیزه با نگاه آهوی دلربا

Le meilleur est à venir !

Cette thèse est dédiée à tous les innovateurs qui contribuent, par leur esprit de conquête et par leur prise de risque, à faire émerger un monde dans lequel la souffrance et la maladie tiendront moins de place.

Je pense notamment à Alexandra Elbakyan, fondatrice de Sci-Hub, site de mise à disposition d'articles scientifiques visant à rendre la science plus ouverte et l'innovation plus facile sans lequel ce travail – comme beaucoup d'autres – n'aurait pas été possible.

Si cette thèse vous a été utile ou vous a apporté quelques idées neuves, aidez Sci-Hub à faire face à la pression judiciaire du cartel des éditeurs scientifiques en contribuant à leur campagne de dons et en leur envoyant des bitcoins.

UFR3S-Pharmacie

L'Université n'entend donner aucune approbation aux opinions émises dans les thèses ; celles-ci sont propres à leurs auteurs.

« [...] while it is not a competitive disadvantage for a fund to not have such an AI platform now, it will be one in two years » - Mohammad Rasouli, Stanford University.(1)

A] Introduction :

I] La technologie comme solution à la crise d'efficacité de l'industrie pharmaceutique ?

Ces dernières décennies, les succès remportés par l'industrie pharmaceutique ont été très nombreux et ont eu, pour certains d'entre eux, un impact majeur sur la vie de millions de patients et de familles à travers le monde. Songeons, à titre d'exemple, aux enfants atteints de cancers, dont 80% survivent alors qu'ils n'étaient que 50% il y a quelques décennies(2), songeons aux progrès dans la lutte contre des maladies infectieuses comme l'hépatite C, dont la mortalité a chuté de manière significative dans les pays développés depuis l'arrivée des traitements antiviraux(3), songeons, enfin, aux vaccins qui permettent de prévenir des dizaines de maladies infectieuses et qui ont conduit à la quasi-éradication de la poliomyélite au niveau mondial, pathologie qui condamnait par le passé des milliers d'enfants.(4)

Pour autant, le modèle qui a permis ces victoires majeures des êtres humains sur la souffrance, n'est-il pas perfectible ? Est-il économiquement soutenable dans le temps ? Les dizaines d'années et les dépenses pharaoniques nécessaires à l'élaboration d'une nouvelle thérapeutique ne sont-elles pas compressibles ? A l'heure de l'augmentation exponentielle des connaissances scientifiques, à l'heure à laquelle des centaines de pathologies dont les mécanismes physiopathologiques et les déterminants génétiques sont connus restent sans traitement, à l'heure à laquelle, avec le vieillissement de la population et l'explosion des maladies métaboliques, l'industrie pharmaceutique – qui porte finalement bien mal son nom, puisqu'il s'agit à chaque étape du développement d'un médicament d'un patient et très spécialisé artisanat – sera-t-elle en mesure de faire face avec son modèle actuel à ces nouveaux défis ?

Ainsi, l'humanité pourra-t-elle répondre à l'ambition portée par la Chan Zuckerberg Initiative(5), la fondation du fondateur de Facebook et de son épouse, celle de guérir toutes les maladies à moyen terme ? Comme l'illustre leur manifeste en faveur de ce projet, ce rêve ne pourra devenir un objectif atteignable qu'en imaginant de nouveaux outils et en bouleversant la méthode de création actuelle d'un nouveau médicament et en s'appuyant toujours plus sur les opportunités qui nous sont offertes par les technologies numériques et en particulier par l'expansion de nos capacités de calcul. Ce thème de la technologie au service de l'éradication de toutes les pathologies est – depuis l'émergence de l'IA générative – au cœur des préoccupations des dirigeants de grandes entreprises : Demis Hassabis, le CEO de Google DeepMind prédit l'éradication de toutes les maladies sous 10 ans(6), Dario Amodei, le CEO d'Anthropic, suggère que l'espérance de vie pourrait doubler pour atteindre 150 ans, grâce à l'IA(7), Sam Altman lui-même, le CEO d'OpenAI a déclaré que l'IA permettrait de guérir les cancers et les maladies cardiaques à un rythme sans précédent. (8)

Parmi les différentes initiatives visant à utiliser les technologies numériques afin d'augmenter la productivité de l'industrie pharmaceutique – et en particulier l'intelligence artificielle générative, nous y reviendrons – il apparaît également indispensable de citer le récent partenariat entre Moderna et OpenAI(9), ayant pour objectif de diviser par 100 les effectifs nécessaires au développement d'un nouveau médicament ou encore l'historique levée de fonds d'amorçage de

1 milliards de dollars de la « mega-start-up » Xaira Therapeutics(10) dont l'ambition est de diviser par trois les capitaux nécessaires au développement d'un nouveau médicament.

Une loi empirique, nommée loi d'Eroom(11), en référence à la loi de Moore, décrit le catastrophique déclin de l'efficacité de l'industrie pharmaceutique : tous les 9 ans, le nombre de médicaments approuvés par la FDA par milliard de dollars dépensé en R&D diminue de moitié. Différents facteurs sont combinés pour tenter d'expliquer ce phénomène : la nature structurellement incrémentale de la qualité de chaque nouvelle molécule développée (elle doit présenter une efficacité et une sécurité au moins non-inférieures à celles des molécules qui sont déjà commercialisées) ; le durcissement progressif des réglementations ; la tendance au surinvestissement et aux dépenses inutiles des laboratoires pharmaceutiques et enfin, la concentration des projets de recherche dans certaines aires thérapeutiques ou se focalisant sur certains types de thérapeutiques « en vogue ». En résumé, les pommes les plus faciles à atteindre ont été cueillies, il faut désormais grimper plus haut dans l'arbre.

Sous l'effet combiné de ces différents facteurs, chaque nouvelle molécule mise sur le marché, coûte en moyenne plus cher à développer que la précédente. Le cabinet Deloitte(12) a ainsi mesuré que la rentabilité de la R&D pharmaceutique a ainsi atteint son plus bas historique à 1,6% en 2019 alors qu'elle était encore de 10,1% en 2010.

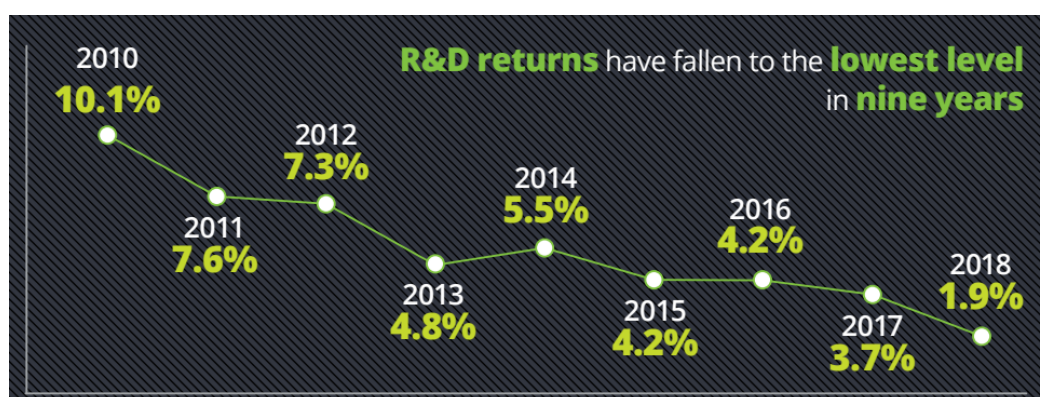


Figure 1 - Evolutions des retours financiers de la R&D pharmaceutique

Atteindre un niveau de rentabilité aussi faible est particulièrement préoccupant en matière de perspective économique : en effet, ces niveaux de rentabilité deviennent inférieurs au coût du capital (le taux de rendement requis par les investisseurs eu égard de la rémunération qu'ils pourraient obtenir d'un placement présentant un profil de risque similaire sur les marchés financiers). Par conséquent, mécaniquement, les budgets dédiés à la recherche et au développement de nouveaux produits dans l'industrie pharmaceutique ne pourront que diminuer.

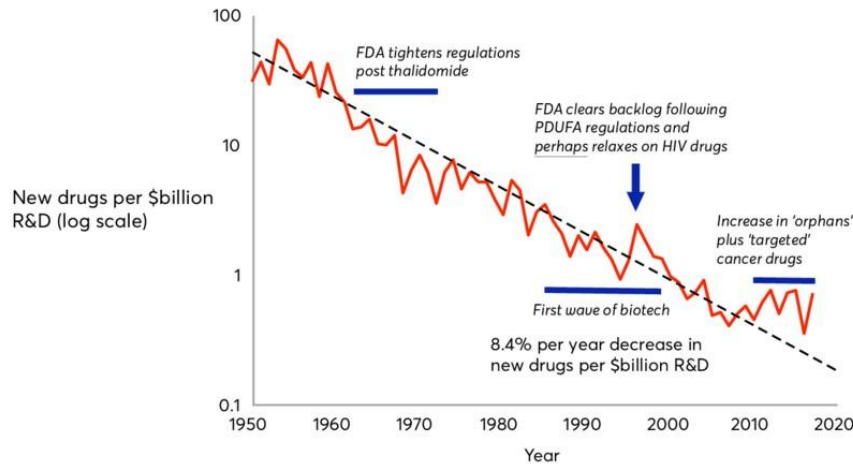


Figure 2 - Diminution de la productivité pharmaceutique à travers le temps

Les différentes recommandations du rapport du cabinet Deloitte mentionnées plus haut dans cette introduction vont dans le sens du plan de création de nouveaux outils numériques envisagé par Sam Altman ou les autres techno-optimistes de la Silicon Valley. Ainsi, le cabinet de conseil recommande notamment aux entreprises pharmaceutiques d'augmenter leur implication dans l'utilisation d'approches transformatives, d'investir dans l'intelligence artificielle et dans les technologies digitales mais également de mettre en place des stratégies d'acquisition et de rétention de talents en data science.

Les grandes structures présentent beaucoup d'inertie et modifier des processus très ancrés prend beaucoup de temps. Dans une interview datant de 2018 reprise par Forbes, Vas Narasimhan, le CEO de Novartis⁽¹³⁾ expliquait à quel point il était difficile d'implémenter des outils d'intelligence artificielle dans les structures des entreprises pharmaceutiques. Pour lui, le potentiel exceptionnel de ce qu'il appelle les DST (pour « Data Science and Technologies » n'avait pas encore trouvé d'expression significative dans la plupart des organisations de développement de médicaments. Pour autant, il avait la conviction que l'utilisation de l'IA et de la Data Science pourrait permettre d'économiser 20% des coûts de développement d'un médicament. Plus loin dans cet article, il avouait qu'en dépit des efforts de Novartis pour se digitaliser, l'impact de l'IA était encore extrêmement faible. Dans d'autres publications, V. Narasimhan expliquait le défi culturel qui compliquait la bonne intégration de l'intelligence artificielle au sein des laboratoires pharmaceutiques.

C'est dans ce contexte compliqué, que nous voyons apparaître un modèle collaboratif avec la création des PharmaTechs, également appelées TechBios – entreprises de technologie au service de la R&D pharmaceutique – la plupart de ces start-up ne développant pas leurs propres médicaments, du moins dans un premier temps, et ayant un business model basé sur les services offerts aux biotechs ou groupes pharmaceutiques. On ne compte plus les annonces de collaborations entre ces jeunes pousses et l'industrie pharmaceutique, à l'image de celle réunissant Insilico Medicine, start-up hong-kongaise dirigée par Alex Zhavoronkov spécialisée dans l'utilisation de l'intelligence artificielle dans le drug discovery et Boehringer Ingelheim⁵.

Bien qu'une partie importante des PharmaTechs soient spécialisées dans le drug discovery, ce n'est pas le seul domaine de la R&D pharmaceutique dans lequel ces start-ups explorent l'utilisation possible de la puissance de l'Intelligence Artificielle (pour ne citer que cette technologie – des PharmaTechs travaillent également sur l'utilisation de la technologie

Blockchain par exemple). Ainsi, il existe de nombreuses entreprises spécialisées dans le développement clinique ou pré-clinique développant pour chacune d'elle des méthodes différentes. Se contenteront-elles de subvenir aux besoins technologiques des grands groupes pharmaceutiques ou réinventeront-elles une nouvelle manière plus rapide et moins cher de développer des médicaments ?

Un domaine très spécifique, à la lisière des TechBios et des Fintech – start-ups mettant la technologie au service de la réinvention de la finance – reste pourtant inexploré, en dépit de son importance. Il s'agit de la rationalisation et de l'optimisation financière de l'investissement dans le développement de nouveaux médicaments. En effet, de nombreux acteurs de la chaîne de financement – à l'instar des fonds d'investissement – ne disposent pas encore d'outils d'aide à la décision s'appuyant sur la puissance de l'Intelligence Artificielle.

II] Pourquoi s'intéresser spécifiquement à l'investissement en Biotech ?

Depuis plusieurs années désormais, l'utilisation de l'Intelligence Artificielle dans les fonds d'investissement tend à se généraliser, avec pour principale moteur la compétition avec les autres fonds d'investissement pour capter les fonds des Limited Partners (nous y reviendrons, les Limited Partners, ou LPs sont les personnes morales ou physiques confiant une partie de leur patrimoine à l'équipe de gestion du fonds d'investissement). Ainsi, 75% des LPs pensent que l'IA peut être utile pour trouver de nouvelles opportunités d'investissement et 64% d'entre eux pensent que la technologie peut être utile pour réaliser de meilleures évaluations des start-ups.(14)

Les tendances macroéconomiques ont également eu pour conséquence d'exacerber la compétition entre les fonds pour les opportunités d'investissement les plus prometteuses (ralentissement dans le financement et la création de nouvelles start-up, lié plus ou moins directement à la hausse des taux d'emprunt), ce qui renforce la nécessité de trouver des manières innovantes de sourcer des opportunités.

Ainsi, alors que l'intelligence artificielle est déjà couramment utilisée par les hedge funds et les investisseurs s'intéressant aux entreprises cotées, elle ne commence que maintenant à se faire une place dans le monde du capital risque, avec pour premiers cas d'usage l'optimisation des temps humains dédiés au deal sourcing (avec pour objectif de trouver les bonnes start-up avant les autres) et à la due diligence (réaliser l'analyse de l'entreprise pour s'assurer qu'elle constitue une opportunité d'investissement intéressante). Nous reviendrons en détail sur ces différents processus des fonds d'investissement.

Les raisons de s'intéresser à l'optimisation des méthodes actuelles d'investissement plus spécifiquement en Sciences de la Vie sont nombreuses :

- Le financement par des fonds de capital-risque de l'innovation en biotech rencontre des problématiques structurelles :
 - L'inadéquation entre les cycles financiers et les cycles scientifiques, ce qui peut parfois pousser des biotechs à faire leur introduction en bourse (IPO) avant d'être suffisamment matures, conduisant à une déception générale des investisseurs et à une chute de leurs cours.

- Le financement en séquentiel, d'une levée de fonds à l'autre, peut parfois mener à la prise de mauvaises décisions stratégiques (nécessité de survivre jusqu'à la prochaine levée, d'atteindre des jalons créateurs de valeurs, etc)
 - L'industrie du capital risque fait parfois preuve d'un certain grégairisme en surinvestissant dans des modes thérapeutiques ou technologiques (CRISPR, ADC, mRNA, etc). Ce qui a pour conséquence de laisser moins de capital aux autres projets mais aussi de créer des bulles en matière de valorisation, et par conséquent des corrections financières brutales.
 - Même si cela est en train d'évoluer, l'industrie a souvent été composée de financiers sans réelle expertise scientifique, conduisant à mal évaluer les risques translationnels ou cliniques et à financer des projets mal conçus ou sous-estimer des approches originales mais peu marketables.
- Le financement des projets de recherche et de développement est une étape fondamentale du développement de nouveaux médicaments. Ne pourrions-nous pas accélérer le cycle de développement des médicaments en optimisant leurs financements ? En concentrant les investissements sur les projets dont la probabilité de réussite est la plus grande et en évitant les mauvaises décisions et en détectant plus précocement les projets à haut potentiel, ne pourrions-nous pas porter plus de nouveaux médicaments jusqu'au marché tout en augmentant les performances financières des investisseurs ?
 - Le mécanisme présidant à la prise de décision d'investissement, au sein des fonds d'investissement est encore en grande partie artisanal, il s'agit d'une méthodologie très chronophage mêlant expertise humaine et ressenti sur le devenir de l'entreprise, l'ensemble étant finalement assez peu performant, du fait notamment du manque d'expertise translationnelle d'un certain nombre d'investisseurs.
 - Les technologies développées par les Biotechs ont la particularité de pouvoir être jugées de manière objective sur les données qu'elles ont déjà générées, qu'il s'agisse de données pré-cliniques, de données cliniques ou encore de données liées à la propriété intellectuelles.
 - Les données permettant de mettre en place un modèle d'aide à la décision d'investissement sont relativement faciles d'accès, en grande quantité, pensons notamment aux bases de données internationales de brevets, aux bases de données cliniques structurées mais aussi au très grand nombre de publications scientifiques disponibles et dont les données importantes sont facilement extractibles en utilisant la puissance de l'Intelligence Artificielle.
 - Un certain nombre de fonds d'investissement spécialisés en santé sont en train de développer leurs propres outils propriétaires. Nous pouvons notamment citer à ce titre Sofinnova Partners, l'un des leaders européens du domaine, qui a développé un outil d'accompagnement⁽¹⁵⁾ de ses équipes reposant sur une gigantesque base de données regroupant plus de 90 millions d'auteurs de publications scientifiques, plus de 40 millions de publications scientifiques, plus de 100 000 instituts de recherche, plus de 4,5 millions de brevets publiés, plus de 450 000 essais cliniques et plus de 2 millions d'entreprises.

En l'absence de la généralisation de l'utilisation de l'IA, l'émergence de tels outils pourrait in fine ne produire qu'un effet négatif sur le marché de l'investissement : en augmentant potentiellement les retours sur investissement des plus gros fonds, au détriment des plus petits ne pouvant pas s'offrir la création d'un tel outil, elle accentuerait le goulot d'étranglement du financement conduisant à concentrer les décisions d'investissement dans un très petits nombre de mains (c'est pourquoi cette thèse sera également l'opportunité de développer des méthodologies

applicables aux plus petits fonds – notamment sur la base des Agents IA et du no-code, nous y reviendrons)

Enfin, parmi les différentes raisons de s'intéresser à l'IA dans le capital risque, il est également important de noter que l'évolution récente de l'IA et la démocratisation de l'utilisation des LLMs (Large Language Models, popularisés par Open AI et ChatGPT) joueront un rôle dans le développement de l'evidence-based investment (réduction des coûts et du temps de développement des outils). Il n'est ainsi plus nécessaire d'embaucher des armées de développeurs pour pouvoir bénéficier d'un outil fonctionnel.

En clair, les outils d'aide à la décision spécifiques aux investissements dans les biotechnologies peuvent constituer un nouveau pan des TechBios et Fintechs restant encore à développer dont il est en tout cas intéressant d'imaginer le futur. C'est ce à quoi s'emploiera cette thèse.

III] Structure et objectifs de la thèse

Afin de répondre à ces objectifs, cette thèse s'attachera dans un premier temps à dresser un panorama général et très descriptif de l'intelligence artificielle, de son histoire, des différents types d'algorithmes et de modèles existants. Puis, de la même manière, à détailler les pratiques et le fonctionnement des fonds de capital-risque.

Dans un second temps, nous nous attarderons sur les initiatives actuelles dans le monde du Venture Capital généraliste en ce qui concerne l'utilisation de l'Intelligence Artificielle. Nous présenterons également la méthodologie et les résultats d'une étude menée dans le cadre de cette thèse auprès d'investisseurs spécialisés en biotech afin de les questionner sur leurs besoins spécifiques et sur leur rapport à l'utilisation de l'intelligence artificielle.

Enfin, au sein d'une troisième partie représentant le cœur de cette thèse, beaucoup plus prospective, nous chercherons à imaginer différentes briques fondamentales d'un outil d'aide à la décision d'investissement dans les biotechs :

- Analyse de la pertinence des données pré-cliniques générées
- Analyse de la propriété intellectuelle générée, qui permettra de (pré)qualifier sa liberté d'exploitation et de mesurer l'intensité de l'environnement concurrentiel
- Analyse de la qualité des études cliniques et prédiction du succès des études à venir.

Enfin, en conclusion, nous détaillerons les modalités de la mise en place de ces technologies dans la prise de décision d'investissement. En annexe, nous développerons l'opportunité que l'IA peut représenter pour une catégorie très spécifique des fonds de capital-risque, les venture builders, spécialisés dans le fait de sourcer des innovations technologiques et de créer des entreprises en interne.

B] Qu'est-ce que l'Intelligence Artificielle ?

L'intelligence artificielle est un continent, une jungle foisonnante de concepts, de méthodes et de techniques, dont l'histoire tumultueuse est vieille de plusieurs décennies, mais dont un certain nombre de concepts remontent à l'antiquité et à la logique aristotélicienne, nous le verrons. Du fait de la grande médiatisation récente de l'intelligence artificielle, de sa complexité et de ses contours mouvants, il est facile de se perdre sur ce continent.

Ainsi, cette thèse étant destinée en priorité aux professionnels de santé, il nous apparaissait comme nécessaire de prendre le temps de poser les grands concepts, de définir les mots importants, de faire un rapide résumé historique et enfin de faire un point opérationnel à travers l'explicitation de plusieurs cas d'usage afin de montrer de quelle manière l'intelligence artificielle peut être utilisée en entreprise. Cette partie sera généraliste et ne sera qu'un prélude aux descriptions plus détaillées qui viendront au cours de cette thèse, au fil de l'exposition de leurs utilisations actuelles ou potentielles dans les fonds d'investissement.

I] Intelligence Artificielle : Définitions et concepts

L'intelligence artificielle n'est pas une discipline, mais plutôt un ensemble de méthodes reliées à un grand nombre de disciplines différentes : les sciences neurocognitives, la neurobiologie computationnelle, l'informatique, la linguistique, et les mathématiques. Ainsi, l'intelligence artificielle se définit de manière transversale à travers ces différents champs de la connaissance comme étant « l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine », notamment la perception, la compréhension, l'apprentissage et le raisonnement. Cette ambition part du principe, établi par les pères fondateurs de l'intelligence artificielle lors du camp d'été de Dartmouth en 1956 – qui occupe une place importante dans la naissance de l'intelligence artificielle, nous y reviendrons dans la partie historique – que chaque aspect de l'intelligence et de l'apprentissage, ainsi que de toutes les autres caractéristiques de l'esprit humain, est décrit de manière suffisante à permettre la conception d'une machine en mesure de les simuler.

La recherche sur l'intelligence artificielle et la science-fiction se sont régulièrement abreuvées mutuellement, de telle sorte qu'il demeure difficile pour le grand public – en dépit du fait que les principes concourant à l'intelligence artificielle ne soient pas extrêmement récents – de fixer la limite, certes mouvante, de ce qui relève de la science et de ce qui relève de la fiction. Cédric Villani, mathématicien et lauréat de la médaille Fields, à l'occasion de la publication en Mars 2018 de son rapport intitulé « Donner un sens à l'Intelligence Artificielle » rapportait que la non-connaissance du concept d'algorithme concernait deux tiers de Français et que la croyance que l'intelligence artificielle allait dominer le monde, entre autres craintes, plus ou moins justifiées que nous n'aborderont pas ici, concernait, elle aussi, deux français sur trois. Ainsi, ces craintes et interrogations nous poussent à différencier, pour dessiner en pointillé la frontière du possible et du fantasme, l'intelligence artificielle forte de l'intelligence artificielle faible. Contrairement à ce que pourrait laisser penser cette dénomination, la différence entre les deux est bien de nature et non pas de degré.

A) L'intelligence artificielle faible

Commençons par l'intelligence artificielle faible : elle représente l'ensemble des techniques actuellement mises en place et accessibles. Il s'agit d'une intelligence artificielle non-sensible, non dotée de conscience et qui se concentre sur une ou des tâches précises. Elle n'est pas

adaptative à des problèmes pour lesquels elle n'a pas été conçue et c'est le concepteur humain qui définit à quelles sources de données elle aura accès (un logiciel conçu pour analyser des images d'IRM ne peut pas comprendre le langage humain). L'intelligence artificielle faible nous entoure, il s'agit par exemple du logiciel de reconnaissance vocal et de communication en langage naturel de l'iPhone développée par Apple, ou encore comme illustré ci-dessous des logiciels permettant de jouer aux échecs contre « l'ordinateur », dont le plus célèbre restera l'ordinateur Deep Blue qui défia Garry Kasparov, alors champion du monde en titre, à la fin des années 1990. La grande caractéristique de l'intelligence artificielle faible est d'être extrêmement spécialisée (bien que cette caractéristique puisse être remise en cause étant donné l'immensité des tâches dans lesquelles peuvent être utilisés les LLMs).

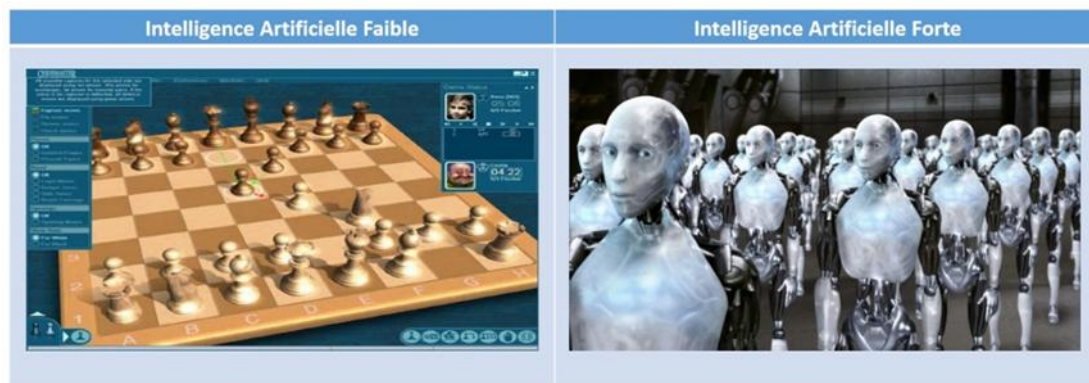


Figure 3 - Différence entre l'IA faible et l'IA forte

B) L'intelligence artificielle forte

L'intelligence artificielle forte est à l'heure actuelle de l'ordre de la projection, du futur, de la fiction. Elle correspond à une machine dotée d'une conscience d'elle-même, de sensibilité et d'esprit, capable de comprendre le sens de ses actions et de les mettre en perspective, de faire preuve de créativité et, pourquoi pas, d'émotion. Le thème de l'intelligence artificielle forte est omniprésent dans la culture de science-fiction, nous pouvons mentionner HAL 9000, l'ordinateur de bord machiavélique de 2001 : L'odyssée de l'espace de Stanley Kubrick, mais aussi, sous un jour moins négatif, les intelligences décrites par Isaac Asimov dans son cycle des Robots. Il occupe également les philosophes et intellectuels, en effet, l'avènement d'une autre forme de « vie » dotée d'intelligence conceptuelle sur Terre aurait un impact qu'il est encore difficile à évaluer pour Homo Sapiens. Ainsi, Nick Bostrom, philosophe et professeur à Oxford a travaillé sur les risques inhérents à l'émergence d'une intelligence artificielle forte et sur concept de « superintelligence » dans son livre éponyme. D'autres personnalités du monde scientifique ou technologique sont également concernées par ces problématiques, comme l'illustre par exemple la lettre ouverte signée 2015 par plus de 150 personnalités dont Stephen Hawking, Elon Musk ou Stuart J. Russell, intitulée « Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter » mentionnant un risque existentiel pour l'humanité en cas d'« explosion d'intelligence ».

Ces problématiques philosophiques sont très intéressantes et concernent l'ensemble de l'humanité, mais nous ne les traiterons pas dans cette thèse, qui se concentrera sur des intelligences artificielles faibles appliquées au domaine de la recherche scientifique et du développement de nouvelles solutions thérapeutiques. Cependant, aux frontières de l'intelligence forte, existe le concept d'intelligence artificielle générale (AGI), qui verra probablement le jour à plus court terme. L'intelligence artificielle générale désigne des systèmes

en mesure de performer dans toutes les tâches cognitives (nous perdons donc l'extrême spécificité des systèmes d'intelligence artificielle faible) mais sans conscience d'elle-même et sans comprendre le sens profond des tâches mathématiques auxquelles elle travaille.

Avec l'émergence des LLMs, les prédictions concernant l'émergence prochaine de l'AGI ont été nombreuses : Sam Altman, le CEO d'Open AI, estime que l'AGI pourrait être opérationnelle avant la fin 2025(16), Elon Musk prédit quant à lui que l'IA pourrait être plus intelligente que le plus intelligent des êtres humains avant la fin de l'année prochaine.(17)

En décalage avec ces déclarations, des acteurs reconnus comme Yann LeCun (directeur des laboratoires de recherche IA de Meta) estiment que l'AGI pourrait au contraire nécessiter des décennies, soulignant que les systèmes actuels basés sur les LLMs manquent de capacités cognitives. Pour lui et pour d'autres détracteurs des modèles LLMs, les déclarations sur l'AGI seraient finalement relativement marketing, avec pour objectif de soutenir la valorisation et les volontés de levée de fonds des entreprises travaillant sur ces modèles.

Une piste privilégiée par de nombreux laboratoires de recherche dans le monde pour développer ce type d'intelligence, qui constitue un pas en avant vers l'intelligence artificielle forte, est de faire se rencontrer deux types d'intelligence artificielle, l'intelligence artificielle symboliste et l'intelligence artificielle connexionniste.

C) L'intelligence artificielle symbolique et l'intelligence artificielle connexionniste

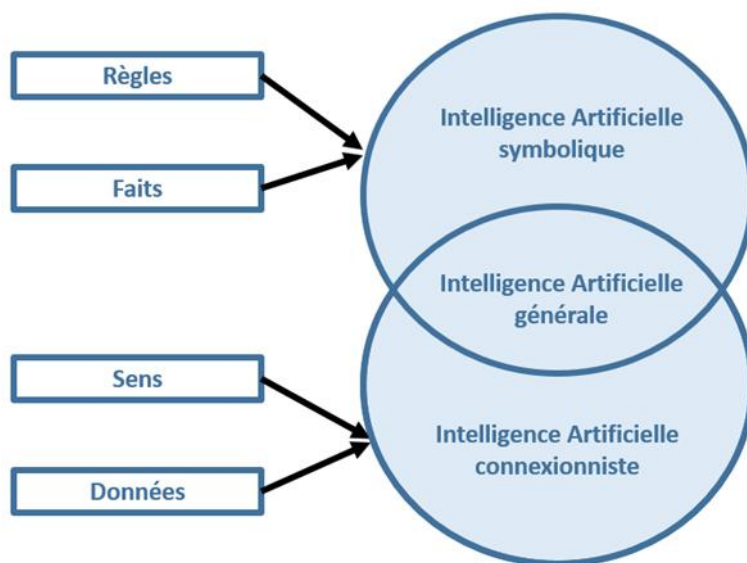


Figure 4 - Différences entre IA Symbolique et Connexionniste

Commençons par l'intelligence artificielle symbolique : les recherches récentes à ce sujet sont relativement rares tant l'intelligence artificielle connexionniste écrase les débats ces dernières années. L'intelligence artificielle symbolique regroupe l'ensemble des méthodes prenant en charge le raisonnement automatique, avec pour objectif d'imiter puis de systématiser les capacités de raisonnement généralistes de l'Homme. Ces techniques s'appuient sur la logique formelle, aristotélicienne et l'exploitation des connaissances existantes (des règles acquises par l'expérimentation humaine du réel) et l'exploitation de faits.

Cette branche de l'intelligence artificielle permet de mettre en place des systèmes experts (*Knowledge Based Expert Systems*) qui sont, par définition, des outils de mécanisations du raisonnement pour obtenir des déductions et des conclusions à partir de règles et de faits.

L'avantage des systèmes experts fabriqués à partir de la méthode symbolique est que les résultats qu'ils produisent sont vérifiables, à partir des règles scientifiques établies et prouvables, ce qui n'est pas le cas dans l'intelligence artificielle connexionniste, nous le verrons.

L'intelligence artificielle symbolique a perdu du terrain au fil du temps face aux méthodes connexionnistes pour des raisons variées :

- A un moment donné du développement de l'intelligence artificielle, où l'on souhaitait principalement traiter le langage, les méthodes symboliques n'étaient pas très adéquates pour cette tâche
- La collecte et la structuration des règles du savoir humain sont longues et complexes
- Les évolutions constantes des capacités de stockage de données, et des capacités de calcul, la baisse de leur coût et l'émergence du réseau Internet mondial, des technologies de réseau sans fil et avec elles, de sources de recueil de données comme les objets connectés (IoT pour Internet of Things) ont favorisé la domination du connexionnisme

Ci-dessous, le schéma explicatif du fonctionnement conceptuel de l'intelligence artificielle symbolique. Nous apporterons des explications plus techniques sur le fonctionnement du système expert dans la partie de cette thèse explorant la possibilité d'automatiser le rapport aux autorités de santé, qui l'un des points primordiaux du développement des entreprises de HealthTech.

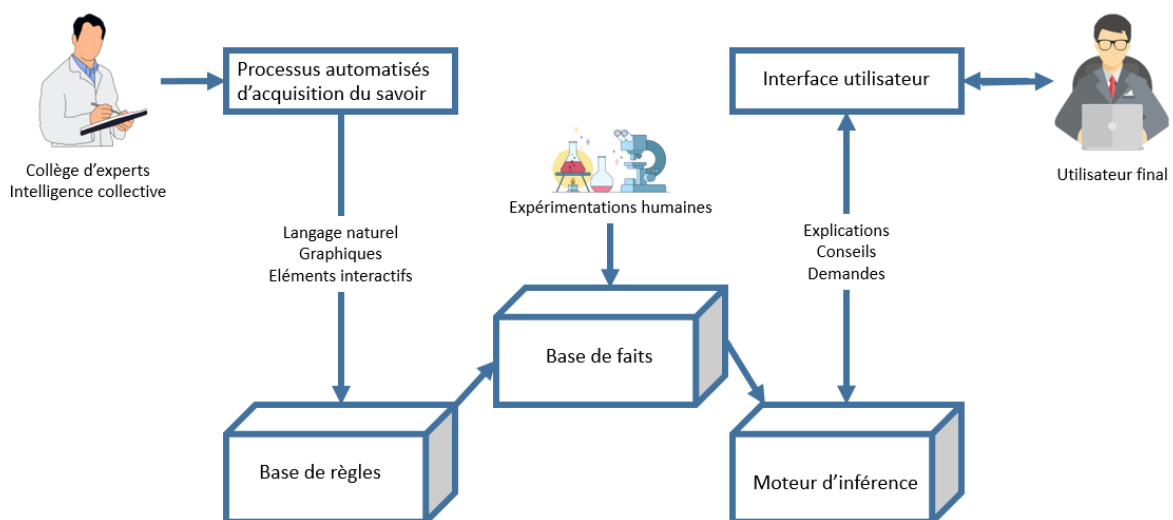


Figure 5 - Fonctionnement de l'IA symbolique

Le moteur d'inférence présent sur ce schéma représente le cœur de l'intelligence artificielle symbolique, il correspond à un logiciel mettant en œuvre des algorithmes de simulation des raisonnements déductifs.

Pour creuser le sujet de l'IA symbolique, dont il sera assez peu question dans cette thèse, nous recommandons la lecture de l'excellent article « Que devient l'IA symbolique ? » d'Olivier Ezratty.(18)

D) L'apprentissage automatique :

Passons désormais à ce qui constitue, pour des raisons évoquées précédemment, la majeure partie de l'intelligence artificielle, la majeure partie des techniques utilisées en entreprises, la majeure partie des recherches actuellement menées sur l'intelligence artificielle, la majeure partie de ce que l'on enseigne aux étudiants et la majeure partie de la couverture médiatique de l'intelligence artificielle : le connexionnisme.

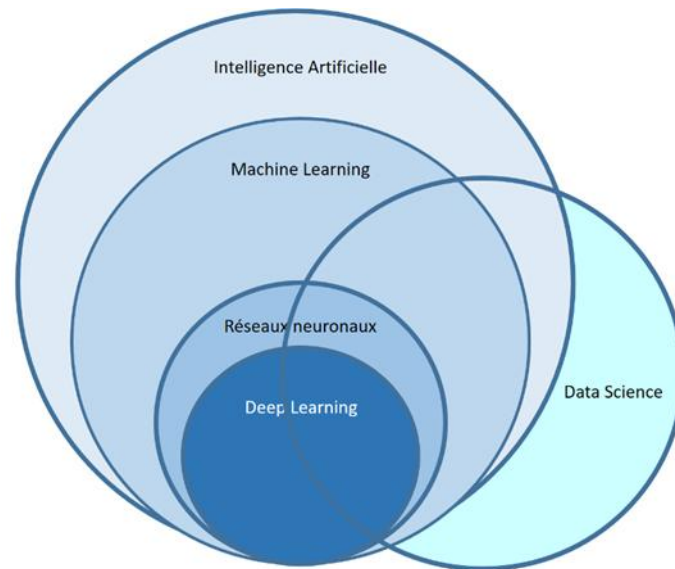


Figure 6 - Apprentissage automatique & IA

En effet, le Machine Learning et le Deep Learning, sont actuellement les formes d'intelligence artificielle les plus utilisées et médiatisées. C'est par conséquent sur ces formes d'intelligence que nous nous attarderons le plus.

Comme l'illustre l'arborescence ci-dessous, le Machine Learning et le Deep Learning sont des technologies que nous pouvons elles-mêmes diviser en plusieurs autres.

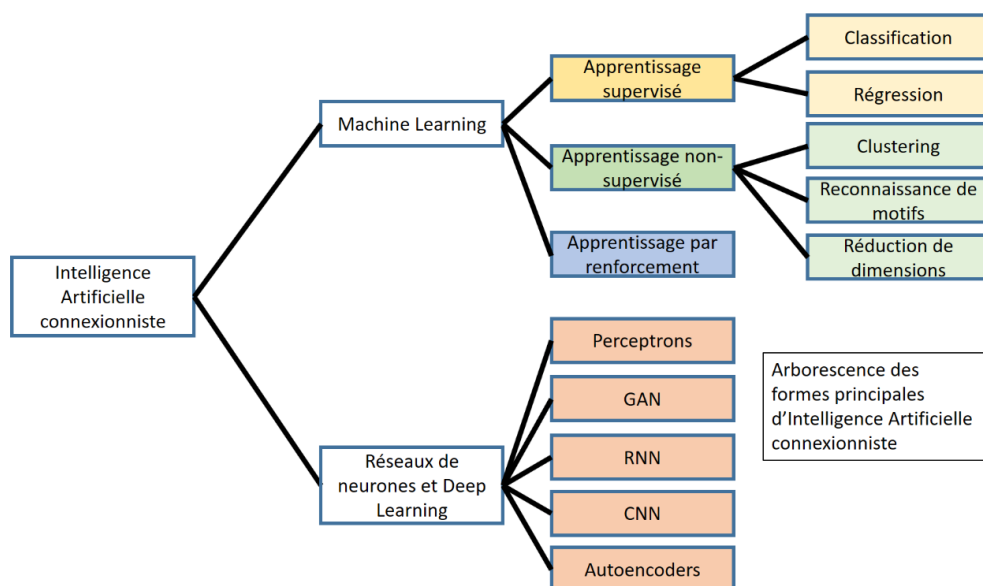


Figure 7 - Les différents types d'IA connexionnistes

Le Machine Learning, ou apprentissage automatique, désigne un ensemble de méthodes permettant à un programme informatique d'identifier des régularités (ou patterns) dans de grandes quantités de données, et d'en tirer des prédictions ou des décisions sans avoir été explicitement programmé pour cela. Ces techniques s'appuient principalement sur des fondements mathématiques et statistiques.

Contrairement à l'intelligence artificielle symbolique, fondée sur des règles logiques et des connaissances humaines formalisées, le Machine Learning repose sur l'exploitation directe de données brutes. L'objectif est d'extraire automatiquement des corrélations ou des modèles implicites à partir de ces données, sans recours à des règles explicites définies à priori.

Si l'on devait établir une analogie avec le fonctionnement du cerveau humain, le Machine Learning correspondrait à la partie traitant les signaux sensoriels, apprenant à reconnaître des régularités à partir de l'expérience accumulée. Ce n'est pas un hasard si les technologies de Machine Learning et en particulier le Deep Learning (ou apprentissage profond), qui reposent sur des réseaux de neurones artificiels multicouches se sont imposées dans des tâches complexes telles que la reconnaissance d'images, ou la compréhension du langage naturel.

Le processus d'apprentissage automatique va comporter deux grandes phases :

- **La phase d'apprentissage** : à partir d'un ensemble de données (le dataset), l'algorithme est entraîné à reconnaître des relations entre les variables d'entrée et des variables de sortie. Ce jeu de données est généralement divisé en deux sous-ensembles : un pour l'entraînement du modèle et l'autre pour l'évaluation de sa performance. Cette évaluation consiste à comparer les prédictions du modèle avec les valeurs réelles de la partie du dataset dédiée au test.
- **La phase de production** : une fois le modèle entraîné et validé, il peut être utilisé pour faire des prédictions sur de nouvelles données issues du monde réel. Par exemple, un algorithme peut être entraîné à reconnaître des objets sur des images pour permettre à une voiture autonome de les éviter sur la route.

Nous reviendrons plus en détail sur les différentes phases du développement d'un modèle plus tard.

Certains algorithmes sont capables d'apprendre en continu, à mesure qu'ils reçoivent de nouvelles données et qu'ils peuvent évaluer l'exactitude de leurs prédictions. C'est ce que l'on appelle l'apprentissage par renforcement, dans lequel l'algorithme améliore ses performances par essai-erreur, en recevant un signal de récompense ou de pénalité.

Enfin, il est important de différencier deux grandes familles de Machine Learning :

- **L'apprentissage supervisé** : repose sur des données préalablement annotées, dans lesquelles chaque exemple est associé à une étiquette ou à une catégorie. L'objectif est alors de prédire l'étiquette d'un nouvel exemple. Cette méthode est très utilisée en imagerie médicale par exemple, lorsque les données peuvent être annotées par des experts.
- **L'apprentissage non supervisé** : à l'inverse, s'applique à des données non étiquetées. L'algorithme tente alors de découvrir par lui-même des structures ou des regroupements naturels dans les données. Dans un contexte d'application médicale, il pourrait ainsi regrouper automatiquement des images selon leur similarité. L'interprétation sera ensuite laissée à l'expert humain.

E) Quelques exemples d'algorithmes utilisés en Intelligence Artificielle :

Afin de rentrer plus en détail dans la présentation de l'IA, voici quelques exemples d'algorithmes simples et fréquents utilisés en Machine Learning :

- **La régression linéaire** : modélise la relation linéaire entre une variable dépendant y et une ou plusieurs variables indépendantes x . Les avantages de cet algorithme sont qu'il est simple, interprétable et rapide à entraîner. Il aura en revanche de nombreuses limitations en ne permettant pas de capturer les relations non-linéaires et sensible aux valeurs aberrantes.
- **La régression logarithmique** : va permettre de calculer la probabilité qu'une instance appartienne à une classe via une fonction sigmoïde. Les avantages sont qu'il est interprétable et bien adapté aux problèmes médicaux binaires. Cet algorithme sera en revanche limité aux séparations linéaires et sera peu performant en cas de forte colinéarité.
- **Les arbres de décision** : cet algorithme simple va diviser les données en fonction de règles de décision simples (si... alors...). L'avantage de cet algorithme tient notamment à son interprétabilité et à sa rapidité d'entraînement. Les limites de cet algorithme concernent la tendance au surapprentissage.
- **Les forêts aléatoires (Random Forest)** : il s'agit d'un ensemble d'arbres de décision construits sur des sous-échantillons aléatoires des données. Cet algorithme sera utilisé pour réaliser des classifications. Il est très robuste et corrige le surapprentissage des arbres individuels et fonctionne bien sur de nombreux types de données.

F) Focus sur le Deep Learning :

Le Deep Learning, ou apprentissage profond, est une sous-catégorie du Machine Learning fondée sur l'utilisation de réseaux de neurones artificiels à plusieurs couches. Il s'inspire de manière abstraite du fonctionnement des réseaux neuronaux biologiques, en empilant des couches de neurones artificiels organisés en architecture hiérarchique.

L'idée fondamentale du Deep Learning est que chaque couche du réseau apprend à extraire un niveau croissant d'abstraction à partir de données brutes : dans le cas d'une image par exemple, les premières couches détectent des motifs simples (bords, textures, etc), les couches intermédiaires détectent des formes (yeux, visages, etc) et enfin les couches profondes seront en mesure d'interpréter l'image.

Cette approche se distingue du Machine Learning « classique » par sa capacité à apprendre automatiquement les représentations pertinentes à partir des données, sans qu'il soit nécessaire de spécifier manuellement les variables explicatives. En contrepartie, le Deep Learning nécessite des volumes de données massifs, des ressources de calcul importantes et présente une moindre explicabilité des résultats (ce qui peut poser problème dans un certain nombre de cas d'usage).

Le Deep Learning est également au cœur des modèles génératifs modernes, comme les LLMs ou les architectures de diffusion utilisées en génération d'images.

G) La révolution de l'intelligence artificielle générative :

Depuis 2022 et le lancement de ChatGPT par OpenAI, l'intelligence artificielle générative connaît une croissance fulgurante et suscite un intérêt inédit dans les milieux scientifiques, industriels et grand public. Cette branche de l'IA, encore relativement confidentielle il y a quelques années, est

désormais au cœur d'une révolution technologique comparable à celle d'Internet ou des smartphones.

Les premiers travaux sur des modèles génératifs remontent aux années 2010, avec notamment l'émergence des GANs (Generative Adversarial Networks), introduits par Ian Goodfellow en 2014, capables de générer des images réalistes à partir du bruit aléatoire. Par la suite les chercheurs se sont orientés vers des architectures plus puissantes et stables, en particulier dans le domaine du traitement du langage naturel.

L'explosion médiatique et industrielle de l'IA générative s'est produite entre 2022 et 2023, avec le déploiement de modèles LLM, comme GPT-3.5 et GPT-4 développés par OpenAI, LLaMA par Meta, Claude par Anthropic, ou encore Gemini (ex-Bard) par Google. Ces modèles ont rapidement montré une capacité impressionnante à générer du texte, du code, des images, des musiques ou encore des molécules, à un niveau qui rivalise parfois avec des humains experts.

Cette évolution s'explique par la conjonction de plusieurs facteurs : une augmentation exponentielle de la puissance de calcul disponible (notamment grâce aux GPU et au cloud), l'accès à des volumes massifs de données d'entraînement, des avancées dans les architectures de réseaux de neurones (en particulier les modèles dits « transformers » introduits en 2017 dans l'article fondateur « Attention is All you need »).(19)

L'IA générative repose principalement sur des modèles de Deep Learning capables de modéliser une distribution de probabilité à partir d'un corpus de données (quelle est la probabilité que le mot suivant soit celui-ci ?) puis de générer de nouvelles instances qui en respectent les caractéristiques (génération de texte, par exemple).

Parmi les grandes familles de modèles génératifs, nous pouvons mentionner :

- Les LLMs, qui comme mentionné plus haut sont des modèles pré-entraînés (entraînés en amont sur de grandes quantités de données avant d'être réutilisés sur des tâches spécifiques) capables de prédire la suite la plus probable d'un texte donné.
- Les modèles de diffusion, initialement conçus pour générer des images à partir du texte
- Les modèles multimodaux, capables de traiter simultanément plusieurs types de données, ouvrant la voie à des applications transversales et multiples
- Les modèles génératifs propres à la chimie et à la biologie qui seront capables de prédire ou de générer des structures biologiques comme des protéines ou des molécules avec des applications immédiates en drug design par exemple.

II] Une histoire de l'Intelligence Artificielle : des hommes et des concepts

A) Entre mythes et légendes, une idée vieille comme l'humanité ?

L'histoire de l'intelligence artificielle débute avec le vieux rêve des hommes d'insuffler l'intelligence à des entités mécaniques. Ce rêve, bien qu'il ne soit toujours pas aujourd'hui réalisé s'est incarné à travers de nombreux mythes, légendes ou œuvres de fiction. Nous pouvons notamment citer les robots dorés d'Héphaïstos dans la mythologie grecque ou encore le Golem de la culture hébraïque. Cette idée d'insuffler une intelligence autonome à un corps non vivant a par la suite été reprise dans un nombre incalculable de fois dans la littérature de science-fiction, de la célèbre créature du Docteur Frankenstein, créée par Mary Shelley aux robots d'Isaac Asimov. De nos jours, des entreprises comme Tesla ou Boston Dynamics travaillent activement à

faire de ce rêve une réalité à travers le développement d'androïdes pouvant être utilisés dans un cadre industriel, domestique, ou militaire.

B) La mécanisation de la pensée humaine :

L'idée de pouvoir recréer ou simuler tout ou partie de l'intelligence repose sur l'hypothèse que le processus de pensée humaine peut être mécanisé. Le formalisme du raisonnement a été étudié tout au long de l'histoire de la philosophie. De nombreux systèmes de pensée et d'hypothèses décrits par ces philosophes ont été utilisés dans le développement de l'intelligence artificielle symbolique. Nous pouvons citer bien sûr les syllogismes d'Aristote, mais aussi Euclide ou Al-Khawarizmi, à l'origine de modèles de raisonnements formels.

Au XVII^{ème} siècle, des philosophes comme Leibniz, Hobbes ou Descartes creusèrent le sillon de la systématisation et mécanisation de la réflexion humaine. Cela aboutit à des systèmes de symboles. Enfin, au début du XX^{ème} siècle, les efforts conjugués de différents philosophes et mathématiciens dans l'étude de la logique mathématique ont permis de commencer à imaginer les contours d'une intelligence artificielle encore dans les limbes. Ces différents efforts amèneront aux travaux d'Alan Turing, mathématicien et cryptologue britannique sur la calculabilité.

Turing joua ensuite un rôle déterminant dans l'histoire de la Seconde Guerre mondiale en imaginant une série de machines dont la plus connue est Enigma qui permirent aux alliés de décrypter les messages codés nazis et qui contribua à poser les bases de l'informatique. Turing a également laissé son nom au test qu'il a conçu et qui permet de définir si un programme est intelligent ou pas. En effet, si une machine peut mener une discussion avec un humain (par téléscripateur interposé) sans que celui-ci ne soit en mesure de déterminer s'il s'agit d'un humain ou d'une machine, alors elle pourra être déterminée comme intelligente.

C) La naissance de l'intelligence artificielle :

L'intelligence artificielle a été fondée officiellement en tant que discipline en 1956, en réunissant sur le même projet de construire une machine intelligente des scientifiques de différents domaines comme les mathématiques, la psychologie, l'ingénierie, l'économie ou encore les sciences politiques. En effet, c'est la Conférence de Dartmouth, à l'initiative de Marvin Minsky et de John McCarthy qui est à l'origine de la fondation de ce champ de recherches. Cette conférence réunissait des chercheurs académiques ainsi que des scientifiques issus de l'entreprise IBM. Les retombées et percées scientifiques permises par cette première conférence et par les possibilités de recherches qui en découlèrent ont été massives et ont donné naissance à une période de presque 20 ans qui fut qualifiée de « L'âge d'or de l'IA »

En effet, cette période voit apparaître des révolutions majeures dans le domaine, notamment du traitement automatisé du langage naturel. Il s'agissait dans le contexte de la guerre froide d'un enjeu majeur afin de comprendre facilement le russe sans forcément disposer de grandes quantités d'agents le lisant couramment. Ces avancées scientifiques concrètes et le grand optimisme des pères fondateurs (Marvin Minsky pensait par exemple en 1970 qu'une machine aurait l'intelligence équivalente à celle d'un homme en moins de 10 ans ; Herbert Simon, lui aussi présent à la conférence de 1956, pensait quant à lui en 1965 que les machines seraient capables de remplacer intégralement les humains dans le monde du travail en 20 ans), ont fait affluer les investissements dans ce champ de recherche, aux USA et également au Japon.

A cette époque, l'Intelligence Artificielle est majoritairement symbolique. Cependant, des chercheurs explorent d'autres voies, à l'image du psychologue Frank Rosenblatt, inventeur du perceptron, machine mécanique ancêtre du Deep Learning moderne. Alors que Rosenblatt mourrait dans un accident de la route, le perceptron fut attaqué avec tant de virulence par les tenants de l'IA symbolique, que toute recherche sur le connexionnisme fût stoppé pendant une décennie. Depuis, l'affrontement ne s'est jamais réellement interrompu, même si les connexionnistes semblent avoir gagné la guerre, pour l'instant.

D) Le premier grand hiver de l'IA :

Malheureusement, les résultats industriellement transposables se faisant trop attendre, les flux d'investissement se tarirent, la plupart des gouvernements et institutions étant déçues après l'enthousiasme pour des objectifs hors-de-portée qu'avait suscité l'optimisme premier et prenant conscience des grandes difficultés techniques se présentant face à l'intelligence artificielle. Parmi ces difficultés techniques, nous pouvons mentionner : les limites des puissances de calcul, le manque de données (sans Internet, comment récupérer des millions de photos de chiens et de chats pour créer un programme en mesure de les différencier ?), le paradoxe de Moravec (énonçant qu'un raisonnement de haut niveau est bien plus facile à modéliser qu'un autre plus simple, relatif par exemple aux aptitudes sensorielles humaines) ainsi que les limites structurelles de la logique classique qui impliqua de devoir en constituer d'autres types.

L'intelligence artificielle finit par acquérir auprès du grand public une image de vaste fumisterie et nombre de chercheurs, faute de financements, se détournèrent du domaine.

E) Le retour de l'IA, portée par l'industrie :

Dans les années 80, un nouveau type d'intelligence artificielle symbolique fait son apparition : le système expert. Il s'agit d'un programme qui répond à des questions et résout des problèmes dans un domaine de connaissances donné, à l'aide de règles logiques dérivées de la connaissance des experts humains de ce domaine. Les centres industriels de recherche et développement sont cette fois le fer de lance du développement de l'IA.

Du côté du connexionnisme, de nouvelles techniques et de nouveaux types de réseaux de neurones sont mis au point, sous l'impulsion notamment de John Hopfield, en plus d'applications commerciales se profilant, comme la reconnaissance optique des caractères (OCR) et la reconnaissance, permettent de redynamiser l'intérêt pour ce domaine de recherche.

F) Le second grand hiver de l'IA :

Le second hiver de l'IA s'explique encore par des raisons économiques, mais cette fois, non en raison de coupes budgétaires. Les centres industriels des firmes informatiques de l'époque furent en effet victime de la disruption apportée par les ordinateurs de bureau d'Apple et d'IBM ce qui fit, du jour au lendemain disparaître leur industrie et tous les budgets qu'ils allouaient à la recherche en IA.

G) Le retour progressif de l'IA depuis les années 1990 :

Depuis les années 1990, l'intelligence artificielle a connu une succession de vagues d'enthousiasme et de périodes de stagnation, mais les avancées technologiques constantes ont progressivement levé les principaux verrous techniques. Conformément à la loi de Moore, la puissance de calcul n'a cessé d'augmenter, tandis que la généralisation d'Internet a permis l'accès à des quantités de données massives, conditions indispensables à l'entraînement de

modèles performants. En 1997, la victoire de Deep Blue, l'ordinateur d'IBM, contre le champion du monde d'échecs Garry Kasparov marque un tournant symbolique dans la reconnaissance publique de l'IA. Deux décennies plus tard, en 2016, AlphaGo de DeepMind bat le champion du monde de Go, un jeu réputé inabordable pour les approches algorithmiques traditionnelles, illustrant la montée en puissance des méthodes d'apprentissage profond (deep learning). Parallèlement, les géants du numérique (GAFA) ont massivement investi dans l'intelligence artificielle, bénéficiant d'un positionnement stratégique à l'intersection des données, des infrastructures cloud et des capacités de financement.

H) L'émergence de l'intelligence artificielle générative

Depuis 2018, une nouvelle révolution s'est amorcée avec l'émergence de l'intelligence artificielle générative, portée par des modèles de très grande taille (Large Language Models) tels que GPT, BERT ou encore DALL-E. En 2022, la sortie de ChatGPT démocratise l'usage de ces technologies, marquant un changement de paradigme : l'IA ne se contente plus d'optimiser ou de prédire, elle est désormais capable de créer, rédiger, programmer, voire raisonner dans certains cas. Ces avancées spectaculaires reposent sur trois piliers : l'augmentation de la puissance de calcul (notamment via les GPU), l'accumulation de vastes corpus de données textuelles et visuelles, et l'optimisation des architectures de réseaux neuronaux. L'IA s'invite désormais dans tous les secteurs, de la santé à la finance, de l'éducation à l'art, en passant par la conduite autonome. L'arrivée prochaine des réseaux 5G et l'essor des objets connectés promettent d'étendre encore davantage les cas d'usage en permettant un traitement distribué et en temps réel. L'intelligence artificielle s'affirme ainsi comme une technologie systémique, structurante et omniprésente, avec un impact profond sur les organisations humaines, les modèles économiques et les équilibres sociétaux.

III] L'intelligence artificielle en pratique : les étapes de développement d'un modèle

Le développement d'un modèle d'IA repose sur une chaîne d'actions rigoureuses, qui va de la collecte des données à l'implémentation dans des applications métiers. Le schéma ci-dessous illustre cette démarche itérative, constituée de plusieurs étapes successives, chacune jouant un rôle clé dans la performance finale du modèle.

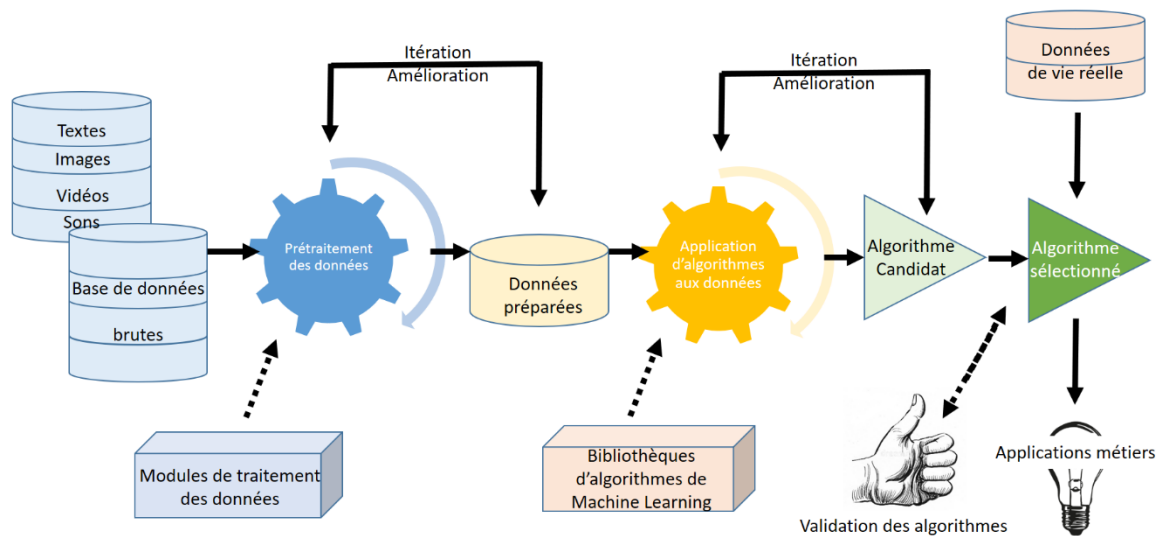


Illustration du processus de développement d'un modèle d'intelligence artificielle

Figure 8 - Développement d'un modèle d'IA

A) La collecte et la structuration des données brutes :

Le processus débute par la collecte de données dites brutes, qui peuvent provenir de multiples sources : textes, images, vidéos, sons, ou encore bases de données classiques. Ces données sont rarement exploitables en l'état. Elles nécessitent une phase de traitement préalable pour les nettoyer, les normaliser et les structurer.

B) Le prétraitement des données :

Le prétraitement des données est une étape déterminante. Elle mobilise des modules de traitement (souvent automatisés) pour transformer les données brutes en un format cohérent, sans valeurs manquantes, sans anomalies, et souvent réduit à des variables pertinentes. Cette étape inclut des techniques telles que la vectorisation de texte, la normalisation numérique, l'échantillonnage, ou encore l'encodage de variables catégorielles.

C) La constitution d'un jeu de données préparé :

Les données ainsi traitées forment un jeu de données préparé, qui peut être utilisé pour entraîner un modèle d'IA. La qualité de cette préparation influence fortement les performances des algorithmes à venir. Cette phase est souvent sujette à des itérations, dans lesquelles des améliorations du jeu de données sont effectuées après évaluation des premiers modèles.

D) L'entraînement des algorithmes de Machine Learning :

Les algorithmes de machine learning sont ensuite appliqués aux données préparées. Ils proviennent généralement de bibliothèques open-source telles que scikit-learn, TensorFlow ou PyTorch. Plusieurs modèles candidats sont testés afin d'identifier celui qui offre les meilleures performances sur un ensemble de validation. Cette étape est également itérative : les paramètres des modèles peuvent être ajustés (optimisation hyperparamétrique), ou les données révisées pour améliorer les résultats.

E) La sélection et la validation du modèle :

Une fois les modèles candidats évalués, un algorithme sélectionné est retenu sur la base de critères de performance (précision, rappel, AUC, etc.), mais aussi de robustesse et d'explicabilité. Ce modèle est ensuite confronté à des données de vie réelle pour vérifier sa généralisation en dehors du cadre strict du développement. La validation des algorithmes peut nécessiter des tests supplémentaires en environnement contrôlé (sandbox) avant le déploiement final.

F) Le déploiement et l'intégration dans les applications métiers :

Enfin, le modèle validé est intégré dans les applications métiers : outils de diagnostic, systèmes de recommandation, moteurs de recherche d'informations, plateformes d'analyse prédictive, etc. Ce déploiement nécessite souvent une phase de mise à l'échelle, ainsi qu'une surveillance post-déploiement pour détecter toute dérive ou perte de performance.

G) Le post-déploiement du modèle :

Une fois le modèle déployé en production, le travail ne sera pas terminé. En effet, l'IA, en particulier lorsqu'elle est fondée sur l'apprentissage automatique, est dynamique par nature : ses performances peuvent se dégrader avec le temps, notamment lorsque l'environnement ou les données évoluent. Il est donc essentiel de mettre en place une gestion rigoureuse du cycle de vie des modèles, c'est ce que l'on peut appeler le MLOps (Machine Learning Operations).

- **La surveillance continue :** une fois le modèle intégré dans une application métier, il conviendra de suivre ses performances prédictives (l'accuracy, l'AUC, etc), les erreurs éventuelles ou les cas de non-concordance avec l'avis d'experts et les métriques d'usage.
- **La détection de dérives :** le data-drift, désignant l'évolution des caractéristiques statistiques des données en entrée, ou encore le concept drift, désignant quant à lui l'évolution de la relation entre les variables explicatives et la cible. Ces deux dérives justifient souvent un réentraînement périodique du modèle.
- **La collecte systématique des retours utilisateurs :** Les retours des utilisateurs (erreurs signalées, décisions corrigées, niveaux de confiance) peuvent alimenter une boucle d'amélioration continue du modèle. Ces feedbacks peuvent être intégrés automatiquement (en apprentissage supervisé continu) ou via un processus semi-manuel avec vérification humaine.

H) Les contraintes spécifiques à l'implémentation de l'intelligence artificielle :

Enfin, pour conclure cette partie sur le déploiement pratique d'un modèle d'IA, il est nécessaire de discuter des contraintes et problématiques spécifiques à l'IA, dont l'implémentation se heurte parfois à plusieurs obstacles d'ordre technique, réglementaire, organisationnel et éthique.

- **La qualité et l'accessibilité des données :**

La performance d'un modèle d'IA est en effet directement corrélée à la qualité et à la pertinence des données utilisées pour son entraînement. Les données sont en effet souvent : hétérogènes, incomplètes, stockées dans des formats incompatibles ou non-standardisées. L'accès aux données peut également être particulièrement compliqué.

- **La réglementation, la confidentialité et l'éthique :**

Ces sujet sont d'autant plus prégnants dans le domaine de la santé. La structure souhaitant mettre en place un modèle d'IA devra se conformer à un certain nombre de réglementations : le RGPD (Règlement Général sur la production des données), qui impose un traitement loyal, transparent, sécurisé et proportionné des données personnelles, mais également au règlement européen sur l'IA (AI Act) qui décrit des obligations spécifiques de transparence, de traçabilité et d'explicabilité, notamment.

- **L'interopérabilité et l'intégration dans les systèmes métiers existants :**

La mise en œuvre d'un modèle d'IA ne peut se faire en silo. Il doit s'intégrer de manière fluide dans les systèmes d'information existants (SIH, LIMS, ERP, etc.). Cela nécessite une compatibilité technique (API, formats de données, protocoles) et organisationnelle (formation des équipes, évolution des processus métiers).

- **L'acceptabilité et la résistance au changement :**

Même lorsque les outils sont performants, leur adoption par les professionnels et par les équipes peut être freinée par une méconnaissance des principes de fonctionnement de l'IA, une crainte d'automatisation excessive ou de substitution ou encore par une absence de formation spécifique, limitant la compréhension et la confiance dans les résultats fournis.

C] Qu'est-ce qu'un fonds de Venture Capital ?

I] Définition du Venture Capital et comparaison avec les autres formes de financements

Le capital-risque (ou en anglais Venture Capital) désigne l'investissement en fonds propres par des investisseurs professionnels dans de jeunes entreprises innovantes à fort potentiel de croissance, non cotées en bourse. Ce financement intervient au démarrage ou aux premiers stades de développement de l'entreprise. Contrairement à un emprunt, le capital-risque n'endette pas l'entreprise : l'investisseur apporte des capitaux contre une participation au capital, généralement minoritaire et espère réaliser une plus-value en revendant ses parts après quelques années (généralement de 5 à 7 ans). C'est ce que l'on appelle le financement dilutif (car il dilue le pourcentage de détention des parts de la sociétés des fondateurs). Ce financement dilutif sera essentiel afin de pouvoir accéder à du financement non-dilutif (prêts bancaires ou subventions).

Le venture capital est une sous-partie du capital-investissement (private equity). Les fonds de Venture Capital investissent tôt, comme mentionné précédemment, dans des entreprises en création ou relativement jeunes, alors que les fonds de capital-développement finance des sociétés déjà matures en phase d'expansion. Enfin, les fonds de LBO – le leverage buy-out est une opération financière consistant à acquérir une entreprise en utilisant majoritairement de la dette remboursée ensuite grâce aux flux de trésorerie futurs générés par l'entreprise rachetée. Ainsi, le fonds bénéficiera d'un effet de levier important sur la petite partie de capital injectée. En santé, les cibles de LBO sont souvent des laboratoires d'analyses médicales, des cliniques privées, des entreprises de dispositifs médicaux matures, etc.

Le venture capital se distingue d'autres formes de financement des entreprises innovantes :

- La love money désigne les fonds apportés par la famille, les amis et les proches d'un entrepreneur au moment de la création ou du démarrage d'une start-up, en général sur la base de liens affectifs et sans réelle analyse ou exigence de rentabilité.
- Les business angels sont des personnes physiques qui investissent à titre personnel une partie de leur patrimoine financier dans des start-ups innovantes. Ils interviennent seuls ou en groupe via des clubs d'investissement (club deals) ou des plateformes de crowdequity. Les business angels sont souvent d'anciens entrepreneurs ayant réalisé des exits avantageux. En plus de leurs fonds, ils seront, dans la plupart des cas, en mesure d'apporter leurs réseaux et leur expertise entrepreneuriale.
- Les prêts bancaires pourront également être très utiles aux entreprises. Par rapport à ces formes de financement, le Venture Capital accepte le risque d'échec sans garantie ni remboursement, en échange d'un potentiel rendement élevé en cas de succès. Un banquier exigera un certain nombre de garanties. Une levée de fonds est souvent constituée de financements dilutifs et de financements non-dilutifs, ces derniers s'appuieront sur la sécurisation apportée par les premiers.
- A un certain stade de son développement, l'entreprise peut aussi réaliser une introduction en bourse (IPO). Cela peut représenter une opportunité d'exit pour l'investisseur, mais ce n'est généralement pas la voie choisie si d'autres sont possibles. En effet, les investisseurs subiront une période de vesting durant laquelle ils ne pourront pas céder leurs parts de l'entreprise sur les marchés durant une période définie. Par ailleurs, l'introduction en bourse peut être à double tranchant pour l'entreprise si cette dernière

n'a pas atteint un niveau de maturité suffisant : coût élevé de la cotation, newsflow défaillant, perte d'intérêt du marché, etc...

- Enfin, dans le contexte de la biotech, il existe un dernier type de financement privé, non dilutif cette fois : les royalty funds. Ces fonds vont financer des biotechs en échange de droit futurs sur leurs revenus, généralement sous forme de royalties. Plutôt que de prendre des parts au capital (comme un VC), ces fonds investissent pour recevoir une portion des revenus commerciaux futurs générés par un actif selon un schéma contractuel défini à l'avance (durée fixe, plafond cumulé, etc). Ces fonds, généralement cotés en bourse, maîtrisent leur risque en investissant généralement des tickets importants (plusieurs dizaines à centaines de millions d'euros) exclusivement dans des sociétés disposant d'actifs approuvés et commercialisés ou dans des sociétés en fin de développement clinique (phase III validée ou avec des données intermédiaires encourageantes).

Finalement, à chaque stade de développement de l'entreprise (et de risque associé !) correspond une certaine modalité de financement :

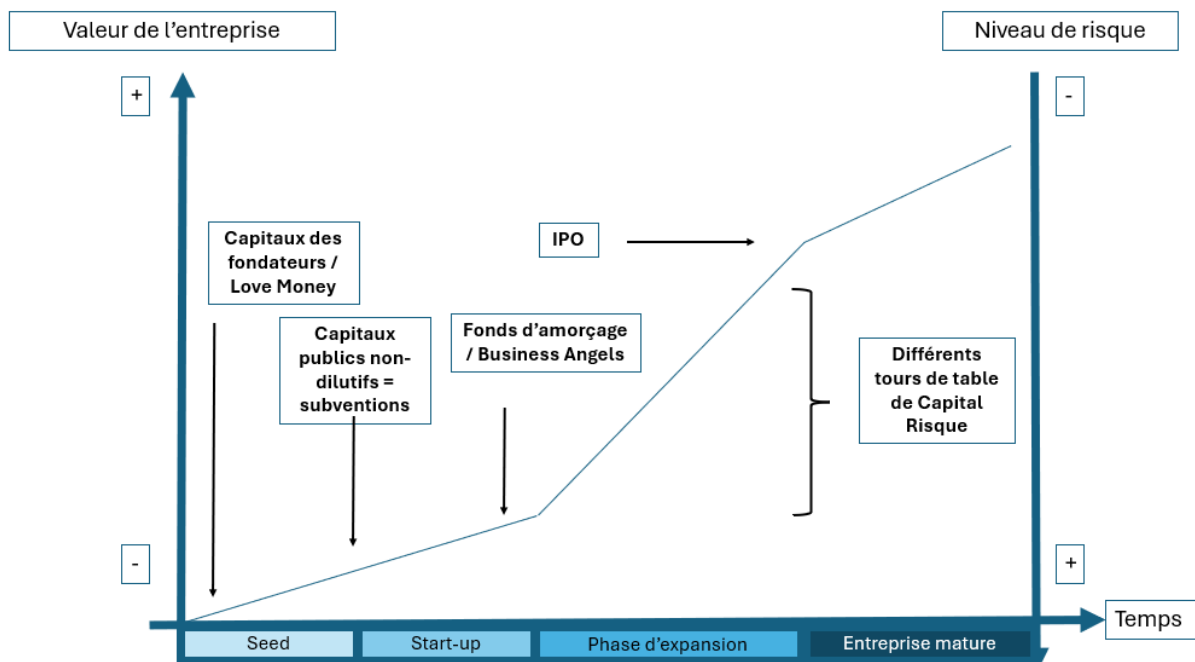


Figure 9 - Différents types de financements de l'innovation

En plus de son investissement, un fonds de Venture Capital apporte souvent son accompagnement (conseils stratégiques, réseau, expérience, etc) pour aider l'entreprise à croître le plus rapidement possible. Cet accompagnement se matérialise dans la plupart des cas par une implication dans la gouvernance de l'entreprise via une place au comité stratégique pour l'investisseur ayant réalisé le deal.

A) Historique du capital-risque et évolution jusqu'à nos jours

Les origines du capital-risque moderne remontent au milieu du XXème siècle. En 1946, l'investisseur franco-américain Georges Doriot(20) créa à Boston l'American Research & Development, considéré comme le premier fonds de capital-risque de l'histoire. Le concept était en effet particulièrement novateur pour l'époque : financer la création de jeunes entreprises innovantes en prenant des risques sur des projets technologiques issus du monde universitaires.

Le grand succès d'ARD a fait école et a inspiré la création d'autres fonds pionniers aux USA, notamment au sein de la Silicon Valley.

Jusqu'aux années 1970, le secteur se structure aux USA avec notamment l'apparition de fonds emblématiques comme Kleiner Perkins ou Sequoia Capital qui profitent de changements réglementaires importants et notamment, l'autorisation pour les caisses de retraite américaines d'investir dans le Venture Capital, ce qui augmente drastiquement les capitaux disponibles (et est encore aujourd'hui du moins en partie à l'origine de l'avance technologique des Etats-Unis). Le modèle de Limited Partnership, dans lequel une société de gestion gère un fonds pour le compte d'investisseurs extérieurs s'impose à cette période et devient le standard de l'industrie.

Des années 1980 à 2000, le secteur connaît une grande structuration avec l'émergence de nombreux fonds, portés notamment par le dynamisme des secteurs technologiques – à ce titre, nous pouvons notamment mentionner le PC et le début des biotechnologies, conduisant à des gains importants pour le secteur, liés à l'investissement dans des entreprises comme Apple, Microsoft ou encore Genentech. La frénésie d'investissement dans des start-up qui fait suite à ces succès conduit à la bulle Internet du début des années 2000, provoquant un ralentissement temporaire du secteur.

Durant les années 2000-2010, les fonds de VC diversifient leurs géographies. L'Europe et l'Asie développent leurs propres écosystèmes VC (souvent avec le soutien d'argent public). Durant les années 2000, le rythme des introductions en bourse ralentit mais les VC continuent de financer l'innovation.

Les années 2010 et 2020 correspondent quant à elles à une période d'expansion massive du capital-risque. Les montants investis explosent dans les années 2010, portés par l'émergence de licornes (start-ups valorisées à plus d'un milliard de dollars), et l'afflux de capitaux institutionnels. Après la crise financière de 2008, une longue phase haussière culmine en 2021, année record de l'investissement dans la biotech au niveau mondial.

Depuis 2022, le marché VC marque un coup d'arrêt, essentiellement provoqué par des facteurs macro-économiques (inflation, hausse des taux d'intérêt, incertitudes géopolitiques) et la correction des survalorisations post-Covid. De nombreuses start-ups qui s'étaient financées aisément en 2020 ou 2021 doivent revoir leur valorisation à la baisse. Le nombre d'IPO baissent et de nombreux fonds de VC n'arrivent plus à lever leurs fonds successeurs. Dans cette industrie en crise, ayant besoin de se réinventer, certains secteurs comme la biotech présentent toutefois des signes de résilience et dynamisme.

II] Fonctionnement des fonds de Venture Capital

L'objectif de cette sous-partie est de bien comprendre les modalités de fonctionnement d'un fonds de venture capital.

○ La levée des fonds :

Lorsqu'un gérant de fonds (General Partner – GP) va souhaiter lever un fonds, il va s'adresser aux Limited Partners, qui sont des personnes physiques fortunées ou des personnes morales (entreprises, family offices, institutionnels, banques, mutuelles, etc) qui, sur la base de la présentation de la stratégie d'investissement portée par la société de gestion, vont s'engager à investir leur argent lors des différents appels de fonds (la société de gestion appellera les fonds auprès de ses LPs à chaque investissement). Il est commun que les GPs contribuent en investissant personnellement environ 1 ou 2% de la totalité du fonds. Le Venture Capital

représentant une classe d'actifs particulièrement illiquide et risquée, les LPs vont généralement attendre des rendements supérieurs à ce qui est proposé sur les marchés financiers : entre 15 et 30% par an. Ainsi, les meilleurs fonds pourront rendre à leurs LPs 3 fois les sommes investies.

○ Les cycles de vie d'un fonds d'investissement

En dehors des fonds dits « Evergreen » ayant une durée de vie de 99 ans – et souvent alimentés par des institutionnels – la plupart des fonds ont une durée de vie beaucoup plus courte de 10 ans (auxquels peuvent éventuellement s'ajouter 2 années d'extension). Cette durée de vie est généralement divisée en deux périodes : la période d'investissement, durant laquelle l'équipe de gestion va sélectionner et investir dans des start-ups prometteuses. Durant la période de suivi et de désinvestissement, l'équipe de gestion cesse de réaliser de nouveaux investissements et se concentre sur l'accompagnement des participations existantes, la consolidation et finalement la cession progressive des parts. A partir de la 8^{ème} année de vie du fonds, la phase de liquidation intervient : il est désormais temps de vendre les dernières participations et de rendre l'argent des LPs afin de pouvoir clore le fonds.

Afin de pouvoir déployer du capital en permanence, l'équipe de gestion se mettra en quête de la levée du fonds successeur environ 2 ans avant la fin de la période d'investissement. Ainsi, plusieurs millésimes de fonds co-existeront au sein d'une même société de gestion. Pour un entrepreneur, il est toujours préférable de recevoir un investissement d'un fonds en début de période d'investissement, car il est à prévoir que les pressions pour l'organisation de la cession seront beaucoup plus tardives – avec davantage de temps, l'entreprise aura eu le temps de générer beaucoup plus de valeur.

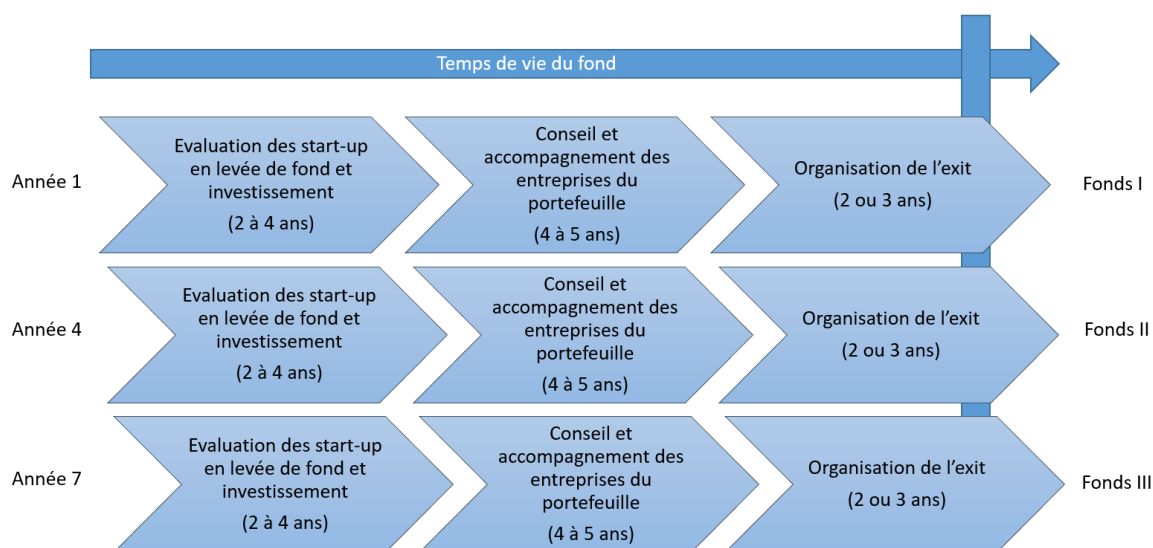


Figure 10 - Organisation d'une société de gestion avec différents fonds successeurs

○ La due-diligence et la sélection des opportunités d'investissement :

Ces tâches font partie du cœur du métier de VC. C'est sur celles-ci que se concentrera la majeure partie de cette thèse.

Avant d'investir, les VC procèdent à une due-diligence approfondie. Les équipes du fonds vont ainsi analyser le business model, la technologie, le marché adressé, l'équipe fondatrice et les perspectives de l'entreprise. Il s'agit d'un véritable audit d'investissement afin d'obtenir une vision à 360° des risques et des potentiels de rendement de l'investissement. Il s'agit d'un processus de sélection très compétitif, en moyenne entre 1 à 2% des start-ups rencontrées feront l'objet d'un

investissement. Ainsi, la qualité du deal-flow (les start-ups entrant en contact avec le fonds) et la capacité de l'équipe à discerner les pépites seront cruciales pour la performance finale du fonds. C'est ce qui explique les grandes variations de performance entre les fonds.

La prise de décision suit généralement un processus bien établi mais toutefois variable d'un fonds à l'autre : analyse du deck, rencontre avec les dirigeants, présentation en comité interne (généralement composé du management de la société de gestion), signature de NDA (non-disclosure agreement / accord de confidentialité) et ouverture de la data room (un espace sécurisé virtuel dans lequel sont centralisés et mis à disposition les informations confidentielles nécessaires à la due-diligence). Puis, la phase qui prendra le plus de temps : la due diligence. Durant cette phase d'analyse du dossier, des sessions de questions et réponses pourront être mises en œuvre. Dans certains fonds, cette phase s'accompagne également d'une rencontre avec certains souscripteurs. Si l'intérêt de l'équipe pour le dossier se maintient, ce dernier sera présenté en comité d'investissement qui validera le mandat d'investissement.

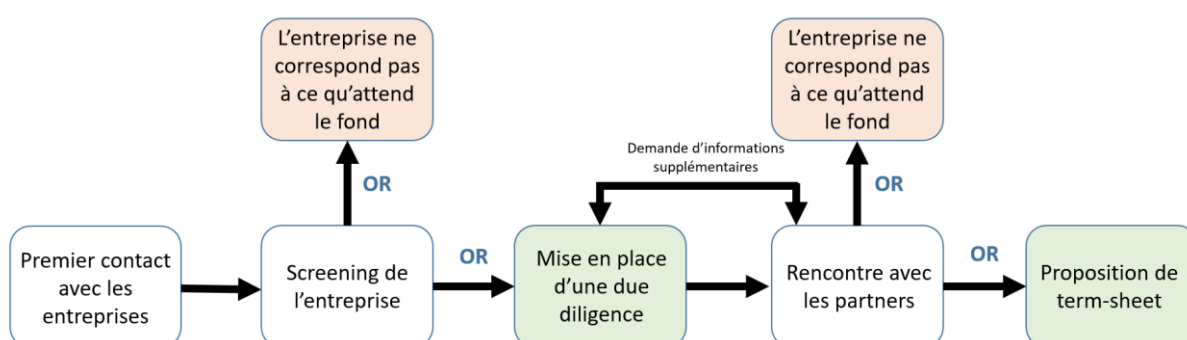


Figure 11 - Illustration d'un processus de sélection et d'analyse

Une fois la volonté d'investir établie, l'équipe d'investissement se lancera – en collaboration avec les co-investisseurs – dans la rédaction de la term-sheet : un document précontractuel qui résume les principaux termes et conditions d'un futur investissement entre une entreprise et un investisseur. L'objectif de ce document non-engageant est de cadrer la négociation en vue de la signature du pacte d'associés.

Parmi les nombreux éléments fondamentaux abordés dans ce document, deux notions méritent particulièrement que l'on s'y attarde : la valorisation et la liquidité préférentielle, tant ces deux concepts cristallisent l'équilibre entre le risque, le potentiel de retour et la protection de l'investisseur.

○ La valorisation :

La valorisation d'une entreprise est l'estimation de sa valeur économique à un moment donné, généralement avant une levée de fonds (pre-money) ou après (post-money). Dans le cas des start-up, la valorisation est négociée et plusieurs méthodes de calcul de cette valorisation peuvent être utilisées :

- **La méthode des comparables** : comparaison avec des entreprises similaires, avec des valorisations observées lors de phases de financement similaires. Il faut dans le cas de l'utilisation de cette méthode être particulièrement vigilant à la date des levées prises comme comparables et au contexte économique dans lequel elles se sont déroulées. Par exemple, les valorisations de start-up étaient beaucoup

plus importantes en 2021 qu'en 2025 (rationalisation du marché post-covid).

- **La méthode DCF** (Discounted Cash Flows) ou rNPV (risk-adjusted net Present Value) : peu utilisée en early-stage mais gagne en pertinence à partir des phases cliniques avancées (les chances de succès sont moins spéculatives). La méthode rNPV consiste à projeter les flux de trésorerie futurs en les ajustant aux probabilités de succès.

Les VC peuvent aussi calculer la valorisation de l'entreprise de manière plus rétrospective (backsolving) : en tenant compte des besoins capitalistiques du projet et de la valorisation de sortie envisagée, afin de construire une equity-story cohérente, qui aura pour objectif de présenter une trajectoire claire de création de valeur aux futurs investisseurs ou acquéreurs potentiels de l'entreprise.

- **La liquidité préférentielle :**

La liquidité préférentielle est une notion souvent difficilement appréhendable par les entrepreneurs. Il s'agit d'une clause contractuelle donnant à certains investisseurs – détenteurs d'actions de préférence – le droit d'être remboursés en priorité en cas d'évènement de liquidité. C'est une protection du capital investi par le fonds : cette clause permet en effet à l'investisseur de récupérer tout ou partie de son investissement avant que les fondateurs et autres actionnaires ordinaires ne soient servis. Dans les cas où l'investissement apparaît comme particulièrement risqué, la liquidité préférentielle peut prévoir un multiple : un investisseur touchera par exemple deux fois son montant investi avant que les autres actionnaires ne commencent à percevoir un retour sur investissement.

Enfin, la liquidité préférentielle peut être de deux grands types :

- **Le non-participating :**

Dans le cadre d'une clause de liquidité préférentielle dite « non-participating », les investisseurs vont bénéficier du remboursement intégral (ou multiplié) de leur investissement initial et le reste sera divisé au prorata des % de détention des propriétaires des actions ordinaires, sans inclure les investisseurs ayant bénéficié des actions de préférence.

Dans la plupart des cas de non-participating, les investisseurs se ménagent le droit de convertir leurs actions de préférence en actions ordinaires : en cas de sortie très avantageuse, il est plus performant de détenir des actions ordinaires que des actions de préférence non-participating.

- **Le participating :**

La liquidité préférentielle dite « participating » permettra aux investisseurs de récupérer leur mise (éventuellement multipliée) mais aussi de récupérer le fruit de la cession au prorata de leur participation à la table de capitalisation.

La capped participating est une variante intégrant un plafond de rendement.

La mauvaise compréhension de ces clauses par les entrepreneurs, additionnée à l'éthique discutable de certains VCs, peut mener à de terribles déconvenues : les entrepreneurs ayant levé beaucoup d'argent et vendant leur entreprise sans en bénéficier financièrement sont nombreux.

Enfin, lorsque les entrepreneurs et les investisseurs seront en accord sur les termes du pacte d'actionnaires, ce dernier pourra être rédigé par les avocats, les audits réglementaires mis en œuvre et le décaissement des fonds effectué : le closing est réalisé.

○ **Le suivi des participations**

Une fois l'investissement réalisé, le fonds VC devient actionnaire de la start-up et s'implique activement dans sa réussite. L'investisseur apporte souvent plus que de l'argent : il siège au comité stratégique afin de peser sur les décisions relatives au développement de l'entreprise, il peut ouvrir son réseau, aider aux recrutements ou encore aider à trouver des synergies avec d'autres sociétés du portefeuille.

En fonction du développement de l'entreprise, l'investisseur pourra décider de réinvestir aux tours de table suivants, ou pas, dans le cas où il constaterait que l'entreprise n'a pas performé. L'objectif de l'investisseur sera alors de maximiser son exposition aux lignes du portefeuille ayant le plus performé et ayant le plus grand potentiel de cession et de valorisation. Accepter ses pertes et éviter l'« acharnement thérapeutique » fait ainsi partie du métier de l'investisseur en capital risque : « on coupe les branches mortes et on arrose les fleurs ». Les capacités au « suivi sélectif » seront évaluées par les LPs.

○ **L'organisation des exits**

L'objectif final du fonds est de sortir du capital de l'entreprise avec un profit substantiel. Plusieurs modalités de sortie sont possibles :

- La cession à un industriel, pour lequel, par exemple, la technologie développée par la start-up pourrait être d'une importance stratégique. C'est le cas de figure le plus fréquent dans le domaine de la santé.
- L'introduction en bourse – cette alternative à la cession déjà mentionnée plus haut n'a généralement pas les faveurs des investisseurs
- La sortie en secondaire consistera quant à elle à la revente des parts du fonds à un autre fonds lors d'un tour de table ultérieur.

Si la société échoue et que le projet s'arrête, le fonds perd la majeure partie de son investissement : en effet, en cas de liquidation, les banques et les potentiels créanciers seront remboursés en priorité. En France, certains investissements peuvent aussi être – sous certaines conditions – garantis par la BPI : en échange d'un pourcentage des plus-values en cas de succès, la garantie peut rembourser environ la moitié de l'investissement en cas d'échec.

○ **Les modalités de rémunération des équipes de gestion :**

Les modalités de rémunération des équipes de gestion sont assez communes pour tous les acteurs du marché : la société de gestion percevra 2% du montant total du fonds par an (avec généralement un maximum de 18% de la taille totale du fonds sur la totalité de la durée de vie du fonds). L'autre modalité de rémunération de l'équipe de gestion sera constituée du Carried Interest, correspondant à 20% des profits réalisés une fois que les investisseurs ont récupéré leur mise initiale enrichie d'un taux minimum de rendement annuel appelée hurdle rate.

Le schéma ci-dessous a pour objectif de synthétiser le fonctionnement global d'un fonds :

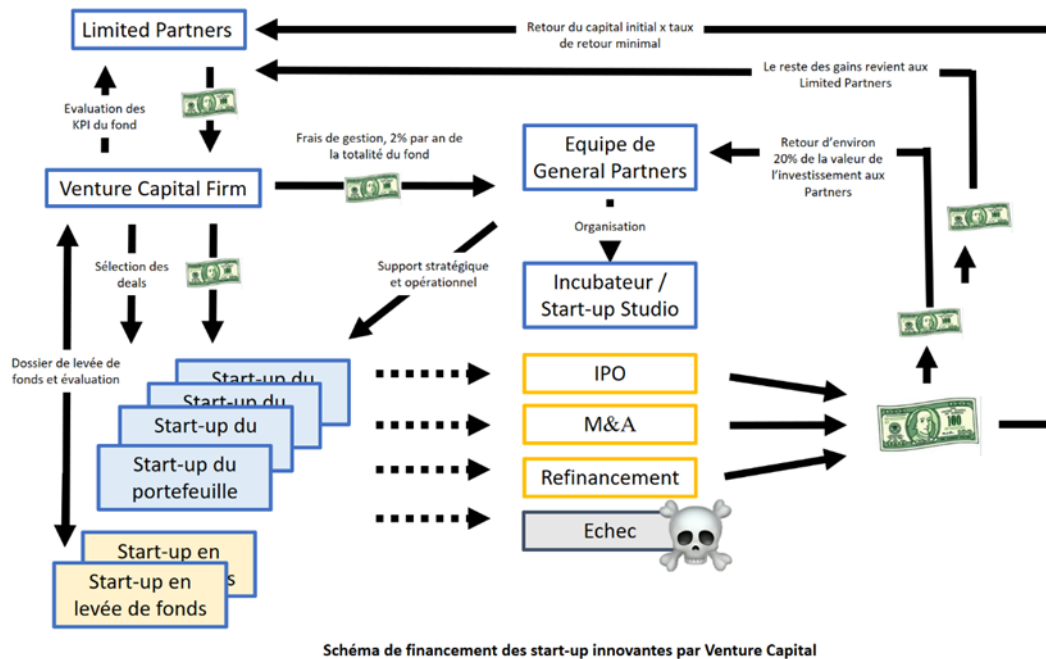


Figure 12 - Illustration du fonctionnement d'un fonds de Venture Capital

III] Les particularités des fonds de Venture Capital dédiés à la biotech

Les fonds de VC spécialisés en biotech fonctionnent selon les mêmes principes financiers mais ils opèrent dans un domaine aux caractéristiques intrinsèques très particulières, ce qui influencent leur stratégie d'investissement :

- **Un horizon d'investissement souvent très long :(21)**

Le développement d'un médicament ou d'une technologie de santé prend souvent 10 ans ou plus avant d'atteindre le marché. Ce qui pose parfois un problème de décorrélation avec l'horizon de temps financier du fonds. Les fonds biotech doivent être prêts à soutenir des sociétés sur une durée prolongée, parfois au-delà de la durée de vie nominale du fonds. Par conséquent, différentes stratégies doivent être mises en place afin de prolonger la durée de vie des fonds ou encore pour passer le relai à d'autres investisseurs ou à des fonds successeurs. Cette inadéquation entre le temps scientifique et le temps financier peut par ailleurs nous pousser à ouvrir un débat fondamental : est-ce que les fonds de VC sont les bons instruments pour financer des biotechs ?

- **Des besoins de capitaux particulièrement élevés :**

Le secteur de la biotech est très capital-intensif. Développer un nouveau traitement peut coûter plusieurs centaines de millions de dollars. Même dès le stade de start-up, les tours de table en biotech sont souvent plus importants que dans le numérique. Il n'est pas rare de voir des levées supérieures à 100M€ pour financer des essais cliniques coûteux. Ainsi, en plus des syndications entre fonds de VC, les VC biotech auront recours à d'autres montages financiers avec des acteurs industriels stratégiques par exemple ou en utilisant des financements non dilutifs issus de contrats de licence (un contrat de licence est un accord juridique par lequel une partie, le concédant, autorise une autre partie, le licencié, à utiliser un droit de propriété intellectuelle dans des conditions déterminées, sans pour autant transférer la propriété de ce droit).

A l'heure où l'IA – en permettant aux entreprises du numérique d'atteindre plus rapidement et à moindre coût la rentabilité – met en cause la survie du VC tech, il est certain que les entreprises de biotech auront toujours besoin de capitaux pour se développer (même si l'IA diminuera les coûts et les durées de développement des nouveaux médicaments).

- **Des profils de risque technologique et marché différents :**

La biotech incarne structurellement le modèle « haut risque et grandes récompenses ». Le risque d'échec scientifique est majeur : en effet, la majorité des candidats-médicaments échoue en cours de développement en raison de signaux de toxicité ou d'un manque d'efficacité. Ainsi, il sera important pour un fonds biotech de diversifier son portefeuille non seulement par nombre de sociétés mais aussi en pariant sur plusieurs approches scientifiques différentes afin d'augmenter ses chances qu'au moins l'une d'elles aboutisse.

En contrepartie du risque, un succès en biotech peut être extrêmement lucratif et le risque marché est très faible : en connaissant les règles d'accès au marché et de remboursement ainsi que les populations cibles, il est possible de l'évaluer de manière assez précise. Ainsi, cette asymétrie entre les risques technologiques élevés et les risques marchés relativement faibles rend les retours encore plus polarisés qu'en VC tech : quelques gros succès font l'essentiel de la performance du fonds, tandis que de nombreux projets sont abandonnés.

- **Des modalités différentes de valorisation :**

Dans le domaine de la tech, lorsqu'une start-up va vouloir se faire financer, elle a généralement déjà un produit en vente ou a minima un MVP (minimum viable product – la version la plus simple et fonctionne d'un produit, à même de répondre au besoin principal de l'utilisateur avec le minimum de fonctionnalités nécessaires) permettant d'obtenir des preuves de traction marché. Ainsi la valorisation pourra s'appuyer sur des métriques liées aux revenus ou au nombre de clients ou d'utilisateurs. Dans le domaine de la biotech en revanche, la start-up ne générera aucun revenu pendant des années, ainsi, la valorisation va dépendre de scénarios futurs très incertains.

La cinétique d'évolution de la valorisation va également être très différente : dans le domaine de la tech, la croissance de la valorisation est continue (puisque'elle évolue en même temps que le chiffre d'affaires de la société) alors que dans le domaine de la biotech, elle va évoluer par palier en fonction des jalons créateurs de valeur atteints par l'entreprise.

- **Des profils d'investisseurs différents :**

Investir dans la biotech demandera une certaine connaissance du développement d'un nouveau médicament et des capacités minimales de compréhension et d'analyse de la recherche scientifique : ainsi, les équipes d'investissement intégreront généralement d'anciens chercheurs, des médecins ou des pharmaciens auprès des équipes de financiers.

Contrairement à la plupart des VC tech classiques, l'écosystème du financement biotech est par ailleurs généralement bien connecté avec les centres de recherche académiques et les organismes de transfert de technologie, a fortiori dans le cas de VC early-stage. Ces connexions peuvent dans certains cas donner lieu à la création de Venture Builders : des fonds spécialisés dans la création de start-up à partir de recherches académiques. Ces modèles – de plus en plus en vogue – permettent au fonds d'être co-fondateur de l'entreprise, d'être majoritaire et de choisir le management.

Encore davantage que dans les autres segments du marché du VC, les fonds biotech acceptent un risque d'échec extrême et un délai de sortie potentielle particulièrement long. C'est à ce prix que des start-ups peuvent générer l'innovation capable de sauver des vies et de générer des retours financiers considérables en cas de succès. En finançant la transition entre des découvertes scientifiques – souvent académiques – et des produits commercialisables et en intervenant à un stade où ni les marchés boursiers ni les industriels ne se risquent, ces fonds de Venture Capital, au même titre que les entrepreneurs, sont un élément clé de toute la chaîne de création de valeur du médicament.

Réfléchir à la manière d'utiliser la technologie afin de permettre à ces fonds de maximiser leurs performances est fondamental pour permettre à davantage de LPs de financer ces fonds – et donc, in fine, pour permettre le financement plus massif du développement de nouveaux candidats-médicaments.

I. Les tendances récentes des fonds de Venture Capital Biotech

Le secteur de la biotech représente une part de plus en plus importante de l'industrie du Venture Capital, passant en 2018-2019 de 7 à 8% du total des montants investis à plus de 15% en 2023. Parallèlement, les secteurs traditionnellement dominants comme le logiciel sont descendus autour de 28% (alors qu'ils représentaient environ 39% il y a 5 ans).(22) Ces chiffres doivent toutefois être nuancés – en effet depuis l'émergence de l'IA générative, les start-ups développant des produits basés sur cette technologie sont plébiscitées par les investisseurs, notamment aux USA.

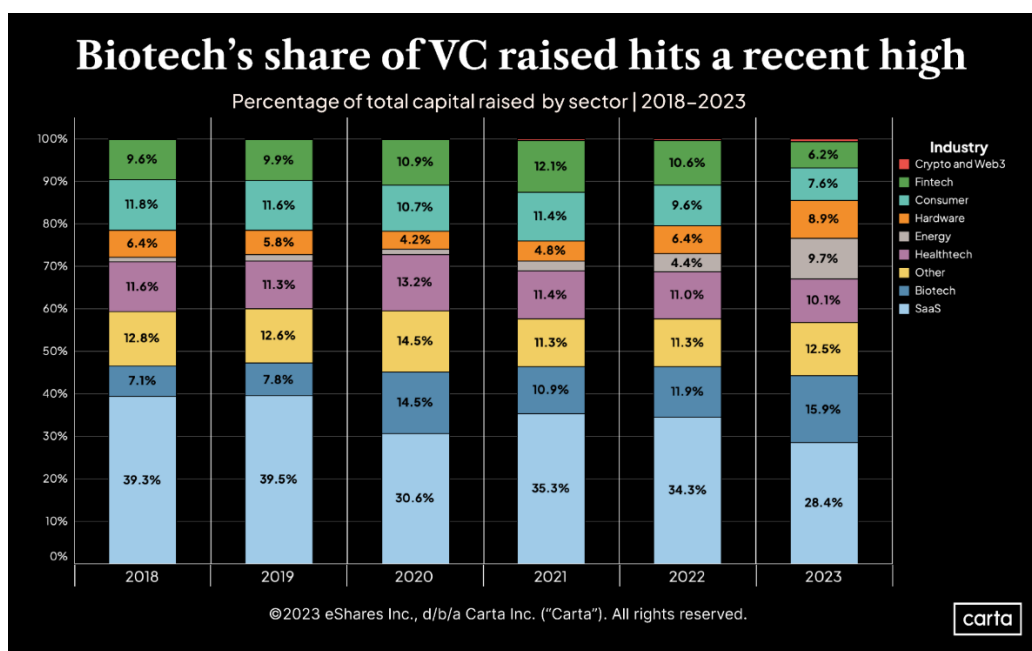


Figure 13 - Part de la biotech dans la totalité des financements VC

○ Des montants investis en baisse depuis 2021 puis en hausse en 2024 :

Comme déjà exprimé précédemment, après le record historique de 2021 (plus de 28 Mds de dollars investis dans les biotechs au niveau mondial) le financement VC en biotech a considérablement diminué jusqu'à un certain rebond en 2024. Ce regain s'explique par le retour de la confiance des investisseurs et le redémarrage des sorties par M&A (Mergers and Acquisitions, désigne les fusions et acquisitions, c'est-à-dire l'ensemble des opérations par

lesquelles des entreprises se rapprochent ou s'achètent entre elles) qui donne le rythme des investissements VCs.

- **L'évolution du nombre de deals et de la taille des tours :**

Le volume de transactions a légèrement diminué ces dernières années, tandis que la taille moyenne des tours augmente (les investisseurs s'intéressant à des entreprises plus matures). Par exemple, au premier trimestre 2025, on a compté 162 financements biotech dans le monde, soit une baisse de quasiment 10% par rapport à la même période en 2024(23). La sélectivité est plus forte et les « mega-rounds » (des tours de financement supérieurs à 100M\$) sont fréquents.

En parallèle, les jeunes biotechs en amorçage ont vu leurs valorisations s'envoler pendant la période COVID : la valorisation pré-money médiane d'une biotech en seed est passée de 7M\$ en 2020 à 20m\$ en 2023

- **Les rendements des fonds Biotech :**

Historiquement, les performances moyennes des fonds de capital-risque ont été volatiles et inférieures aux attentes. Il s'agit en effet d'un secteur d'investissement présentant une distribution très large des rendements : les différences entre le premier quartile des fonds et le dernier est très importante (de l'ordre de X3 à X4). En France, les fonds VC ont dégagé un rendement annuel net moyen d'à peine +2,6% sur la période 2005-2015. Les incitations fiscales visant à diriger du capital vers le VC n'y sont pas tout à fait étrangères : les bénéfices étant réalisés pour les LPs à l'entrée dans le fonds (sous forme de réduction d'impôts), il n'est pas nécessaire de produire une immense performance pour les ravir.

Aux Etats-Unis, en revanche, en raison de la structure de l'actionnariat des fonds, plus exigeant, de la proximité de l'innovation et de la maturité de l'écosystème, le TRI (Taux de rendement interne – calculé de manière à ce que la valeur actuelle nette des flux de trésorerie générés par un investissement soit égale à zéro) médian des fonds de VC a été d'environ 10,4% par an sur les 20 dernières années, avec une performance du premier quartile de 18,5% net par an.

Les fonds de biotech, quant à eux, présentent souvent des retours en « U » : encore plus que dans les autres secteurs, la majorité des investissements peut être perdue, mais une ou deux sorties extraordinaires peuvent propulser le rendement du fond à des niveaux très élevés. Le profil de risque plus élevé en biotech que dans d'autres secteurs produit des exigences plus importantes des LPs, qui exigent généralement dans le secteur des cibles de rendement brut au-delà de 20%.

- **Les tendances récentes des investisseurs biotech :**

Depuis 2022-2023, nous observons un certain réajustement dans l'appétit pour le risque des investisseurs en biotech. En effet, ces derniers se montrent plus prudents et privilégient des projets ayant franchi des étapes clés (preuve de concept clinique, approbations en vue, etc), ce qui a pour conséquence de rediriger les capitaux vers des entreprises plus matures et de rendre beaucoup plus difficile le financement des entreprises early-stage, qui doivent inventer de nouveaux modèles (utilisation de plateforme de crowdfunding, stratégie de génération de revenus via des business models hybrides, etc). Ces nouvelles stratégies ne sont cependant généralement pas suffisantes ni accessibles à toutes les entreprises. Ainsi, il convient, pour maintenir le développement de l'innovation en santé, de réfléchir à la manière de dérisquer précocement les opportunités d'investissement. La pandémie liée au COVID-19 a par ailleurs mis en lumière l'importance du secteur de la santé, incitant de nouveaux investisseurs généralistes à allouer une portion plus grande de leurs fonds vers les sciences de la vie – cette redistribution du

capital plaide en faveur du développement d'outils d'aide à la décision utilisables par des investisseurs non spécialisés.

Les prochaines années nous diront si le VC biotech retrouve une trajectoire de croissance durable, soutenue par les avancées scientifiques (thérapies géniques, CAR-T cells, ARNm, Immunothérapies, etc). Le futur à court terme devrait également nous renseigner sur l'adoption de l'IA par le secteur : est-ce que cette rupture technologique sera à l'origine d'une nouvelle étape dans l'histoire du VC biotech ?

D] Utilisation de l'Intelligence Artificielle dans le domaine de l'investissement : un état des lieux

Les raisons d'adopter l'IA au sein d'un fonds d'investissement sont nombreuses car aujourd'hui encore, le processus d'investissement demeure artisanal, inefficace et potentiellement biaisé.

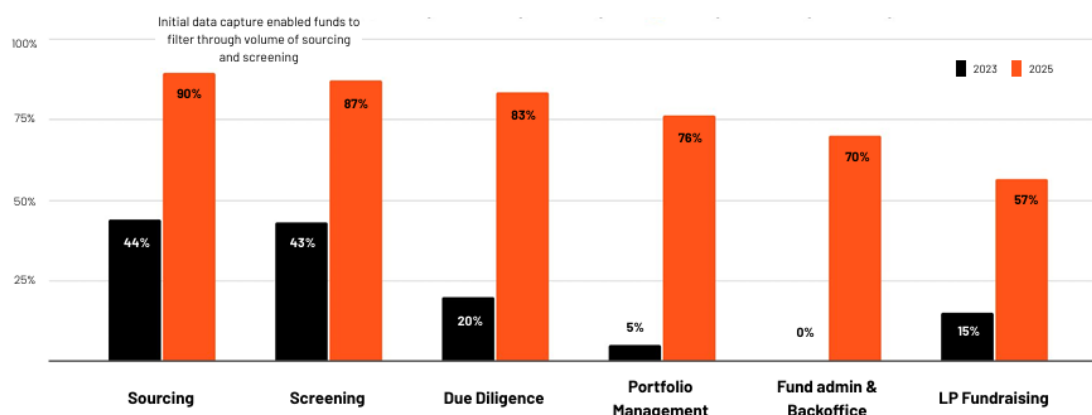
A chaque étape du cycle d'investissement, du sourcing à l'exit, persistent des inefficiences structurelles. Ce constat justifie de chercher à améliorer la performance, la transparence et l'équité des fonds d'investissement.

- Le sourcing repose encore très souvent sur des collectes manuelles d'information, des bases de données partielles et une logique réactive, là où l'IA peut permettre une approche proactive, continue et personnalisée.
- Le screening initial, quant à lui souffre d'un manque de standardisation et d'une forte subjectivité entre les membres de l'équipe d'investissement. L'IA peut permettre de réduire les biais cognitifs, d'homogénéiser les décisions, et à mieux documenter les arbitrages internes.
- Les due diligences, qu'elles soient commerciales, réglementaires ou technologiques mobilisent des ressources importantes et demeurent chronophages. L'IA a le potentiel pour automatiser l'analyse documentaire (via la lecture automatique des data-rooms), de comparer les éléments fournis à des données externes et de produire une première estimation de la qualité des actifs immatériels. L'analyse concurrentielle, souvent subjective, peut également être augmentée par des outils de veille automatisée ou de cartographie.
- Le closing est encore entravé par des formalités administratives manuelles, relatives notamment à la gestion des syndications. Des solutions d'IA peuvent accélérer ces processus, permettant de recommander les meilleurs co-investisseurs et de faciliter la coordination contractuelle entre les parties prenantes.
- La création de valeur et le suivi des entreprises du portefeuille est également une phase chronophage et souvent peu optimisée. L'IA peut enrichir la phase de post-investissement, notamment en analysant les indicateurs de performance des participations, en prédisant les besoins financiers, etc.
- Enfin, les processus de sortie, qu'ils passent par une revente industrielle ou un rachat secondaire, sont eux aussi perfectibles. L'IA peut être mobilisée pour cartographier les acquéreurs potentiels, réduire les coûts liés à la préparation des documents, etc.

Parmi ces différents cas d'usage, le panorama Data-Driven VC 2025(24) illustre que les VCs ont commencé à s'intéresser à l'utilisation de l'IA dans le sourcing et le screening puis ont cherché à l'utiliser dans d'autres parties du processus d'investissement.

Les prochaines iconographies ainsi que les analyses qui en sont faites sont largement issues de ce rapport extrêmement précieux.

Where DDVCs act in the value chain
 % of DDVC funds analyzed 2023 vs 2025



*Figure 14 - Cas d'usage de l'IA chez les VC data-driven (non-spécifique de la biotech) –
 Panorama Data-Driven VC 2025 – Andre Retterath*

En dépit de l'intérêt croissant pour l'intégration de l'IA dans les fonds de VC – comme l'illustre le succès de la newsletter « Data-Driven VC » rédigée par Andre Retterath, regroupant près de 35 000 lecteurs en 2025 – seulement 1% des fonds VC ont initié des démarches data-driven internes(25), portant le nombre de fonds pouvant être considérés comme « data-driven » à 235 en 2025.

La tendance s'accélère cependant grâce à plusieurs facteurs :

- **Une disponibilité de plus en plus importante des données sur les start-ups** (bases de données comme Pitchbook, métriques web, de popularité sur les réseaux sociaux, données de cartes bancaires permettant de détecter des signaux faibles – dans le domaine de la santé, de nombreuses données sont également disponibles : publications scientifiques, études cliniques, brevets, etc)
- **Une maturité de plus en plus importante** – portée par les LLMs – de nouveaux outils techs et no-code, qui sont de plus en plus accessibles et efficaces. Ces outils ne nécessitent pas forcément l'intégration d'équipes techniques en interne pour être déployés et permettent aux fonds d'expérimenter facilement. L'abaissement du coût d'expérimentation permet aux VC d'adopter une logique d'itération rapide.
- **La pression concurrentielle et le besoin de différenciation** : dans un marché saturé de fonds d'investissement et où les entreprises représentant de bonnes opportunités sont souvent sur-sollicités, la capacité à identifier rapidement les meilleurs projets devient crucial. La dynamique IA initiée par les fonds précurseurs pousse les autres acteurs à suivre pour ne pas décrocher.
- **Les attentes des Limited Partners**, qui en effet, demandent de plus en plus de transparence, de granularité et de rigueur dans les processus d'investissement. Pour les sociétés de gestion souhaitant lever de nouveaux véhicules (fonds successeurs par exemple), l'argument data-driven devient un atout marketing et stratégique.

Ainsi, le nombre d'investissements réalisés par des fonds que l'on peut qualifier de « data-driven » n'a cessé d'augmenter ces dernières années.

Panorama mondial de l'adoption de l'IA dans les VC

Une fois le constat fait de l'augmentation du nombre de fonds Data-Driven, il apparaît comme indispensable de déterminer précisément ce que signifie être un fonds data-driven et à quel niveau de développement sont les VCs se qualifiant de data-driven.

Ainsi, parmi les 124 répondants à l'étude de Data-Driven VC, 40% sont en train d'utiliser pleinement la solution IA / Data développée. Les 60% restant n'ont pas finalisé l'adoption de leur solution. Ces chiffres montrent que le mouvement vers l'IA est encore récent.

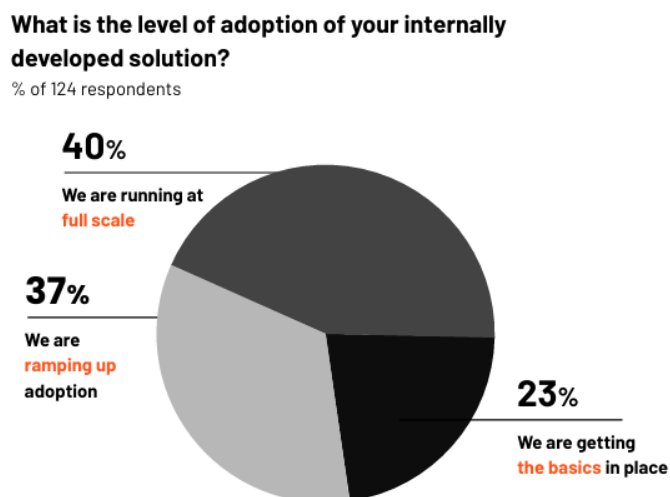


Figure 15 - Niveau d'adoption interne de la solution IA développée – Panorama Data-Driven VC 2025 – Andre Retterath

L'adoption des solutions IA – et notamment de celles qui sont basées sur des LLMs – peut s'expliquer par plusieurs facteurs, qui ont été étudiées de manière quantitative par l'étude Data-Driven VC :

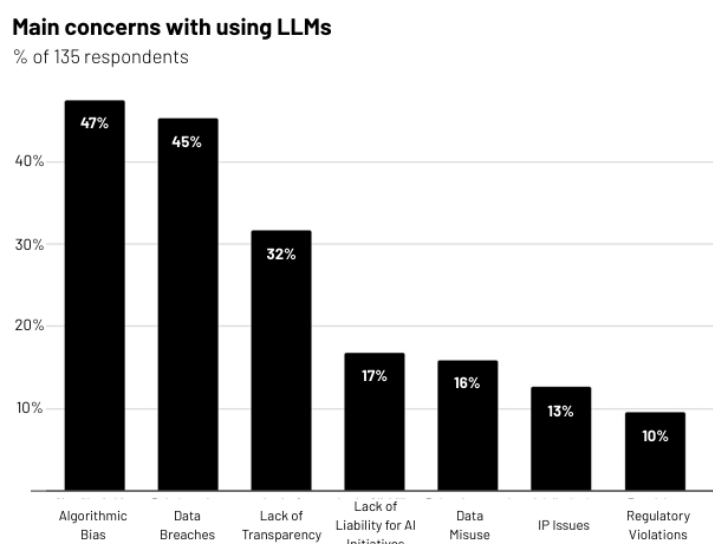


Figure 16 - Principales problématiques liées à l'usage des LLMs - Panorama Data-Driven VC 2025 – Andre Retterath

Parmi ces limitations, un certain nombre peuvent être solutionnées ou atténuées :

- La peur des biais algorithmiques :

Les LLMs s'entraînant sur des corpus massifs de textes, ils peuvent refléter des biais existants (géographiques, culturels, scientifiques, financiers, etc). Un fine-tuning ciblé, consistant à entraîner spécifiquement le modèle sur une base équilibrée (de publications scientifiques sélectionnées par exemple), peut éviter ces biais. Par ailleurs, une stratégie d'audit réguliers, mettant en place des tests internes pour identifier les biais peut également être mise en œuvre.

- La peur de la fuite des données ou de la mauvaise utilisation des données :

Dans le cadre d'une due-diligence, un fonds d'investissement est amené à travailler avec des données sensibles et hautement confidentielles, dont certaines peuvent être critiques pour l'entreprise. Il signe dans ce cadre un non-disclosure agreement l'engageant à ne pas partager ces données. Un LLM non sécurisé peut exposer à des fuites de données, notamment dans les cas où les données sont réinjectées dans les corpus d'entraînement.

De bonnes pratiques doivent par conséquent être mises en place : l'utilisation de LLMs on-premise (un modèle installé, hébergé et exécuté directement sur l'infrastructure information interne du fonds d'investissement), la séparation des données hautement sensibles de celles moins critiques ou publiques ou encore l'anonymisation des données avant ingestion (en ne mentionnant pas, par exemple, le nom d'un composé avant de soumettre ses résultats précliniques au modèle). Ces pratiques permettent par ailleurs d'être en accord avec la réglementation (RGPD par exemple) mais aussi avec la préservation de l'IP des entreprises concernées.

- La peur de l'effet « boîte noire » :

Les LLMs génèrent des réponses sans expliquer leur raisonnement et cela peut nuire à la crédibilité des outils développés et se basant sur ce type de modèles. Cette problématique, inhérente à ce type de modèles peut cependant être atténuée en utilisant des modèles offrant des références (citations, sources, etc) ou encore en couplant les LLMs à des moteurs symboliques.

- Le sujet de la responsabilité des décisions influencées par l'outil IA :

Les LPs peuvent reprocher aux GPs d'avoir délégué leur prise de décision en cas de mauvaise indication de l'outil d'IA. Il sera alors nécessaire de bien clarifier le rôle de l'outil et la gouvernance interne du fonds, ainsi que les procédures internes mentionnant que l'IA est un outil d'aide à la décision.

En résumé, il est indispensable, pour faciliter l'adoption d'un outil d'aide à la décision basé sur des LLMs, de traiter ces problématiques, principalement en adoptant une architecture technique sécurisée et en clarifiant les processus de gouvernance.

Une autre analyse intéressante portée par le rapport concerne cette fois la vision qu'ont les professionnels de la disruption permise (ou promise ?) par l'IA :

Une majorité d'entre eux situe aujourd'hui l'utilisation de l'IA au niveau du « Peak of Inflated Expectations », signe d'un fort engouement et, une fois encore, de la relative nouveauté de l'intérêt pour l'IA dans le Venture Capital. Près de la moitié d'entre eux (44%) voient cependant déjà glisser cette technologie vers le « Trough of Disillusionment », ce qui est le signe des difficultés opérationnelles rencontrées. A l'inverse, seuls 1% des VCs estiment que l'on a atteint

un plateau de la productivité. Cette répartition souligne que l'industrie est encore en grande partie en phase exploratoire, oscillant entre enthousiasme et scepticisme.

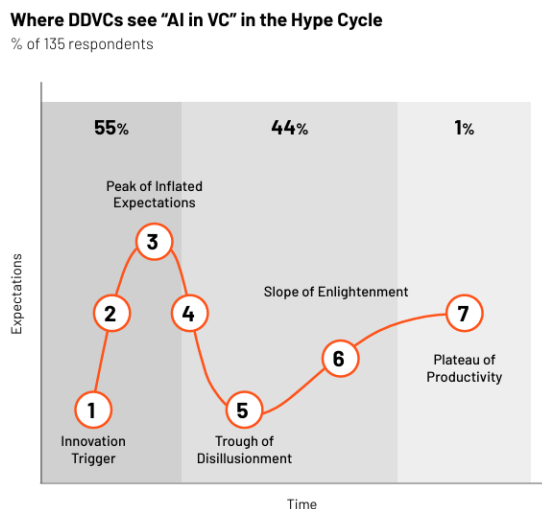


Figure 17 - Estimation du cycle de hype par les VC - Panorama Data-Driven VC 2025 – Andre Retterath

Quelques exemples de fonds utilisant l'IA de manière notable :

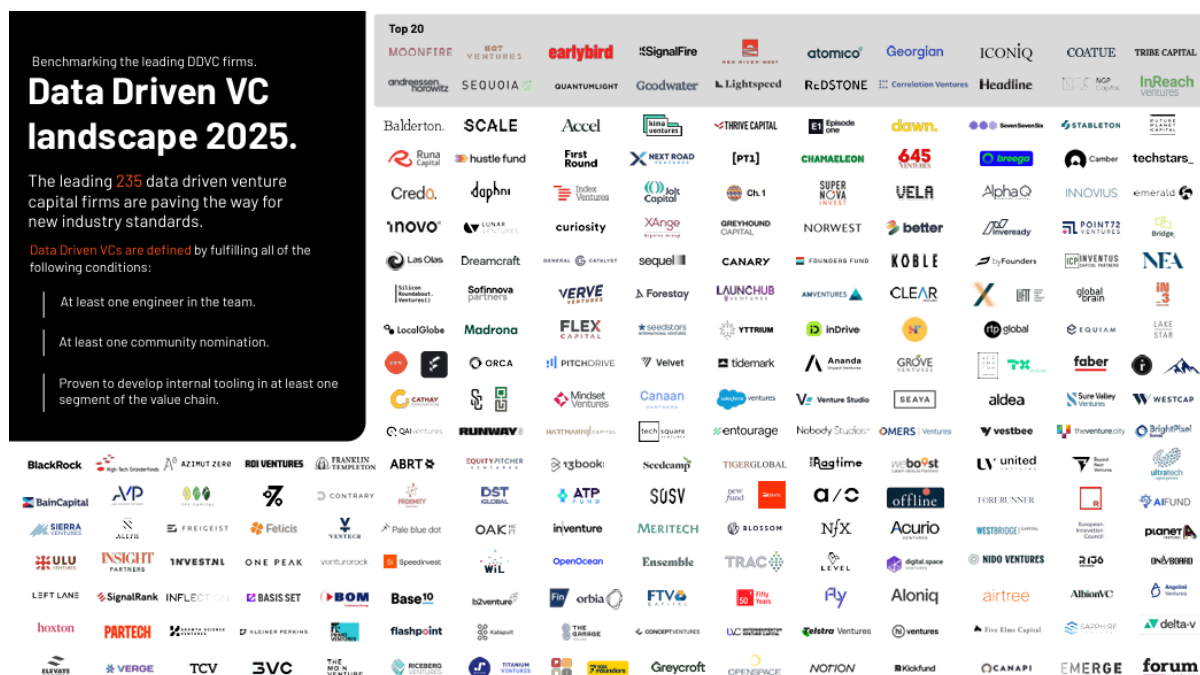


Figure 18 - Panorama des fonds VC data-driven - Panorama Data-Driven VC 2025 – Andre Retterath

Nous l'avons vu, et le panorama ci-dessus l'illustre parfaitement, les fonds data-driven sont de plus en plus nombreux. Plusieurs d'entre eux se sont positionnés à l'avant-garde en intégrant l'IA et les données au cœur de leur stratégie d'investissement. Ces cas pionniers illustrent la variété des approches possibles.

EQT Ventures : Ce fonds européen a développé une plateforme propriétaire nommée Motherbrain. Il s'agit d'un système d'IA qui agrège les données de plus de 10 millions d'entreprises

à travers le monde, issues de dizaines de sources structurées ou non structurées, dans le but d'identifier des signaux faibles annonçant une start-up particulièrement prometteuse. Motherbrain a permis à EQT Ventures de déployer plus de 100M\$ dans des start-ups qui n'auraient pas pu être détectées sans cet outil. L'objectif affiché par le fonds est de dégager du temps aux équipes d'investissement afin de leur permettre de se concentrer sur des tâches à plus forte valeur ajoutée.

SignalFire : Souvent cité comme étant l'archétype du VC data-driven, ce fonds a conçu en interne une plateforme baptisée Beacon AI, perfectionnée depuis plus d'une décennie et qui agrège un volume de données absolument considérable : plus de 650 millions de profils de personnes et plus de 80 millions d'organisations sont suivis en continu. Concrètement, Beacon analyse les tendances du marché, les mouvements de talents et fournit aux équipes d'investissement un « moteur d'intelligence » couvrant tout le spectre de leur activité. La particularité affichée de la plateforme de SignalFire est qu'elle est en mesure d'intervenir dans chaque étape du processus. Par exemple, elle est en mesure de repérer des signaux faibles indiquant des entrepreneurs à fort potentiel avant même qu'ils ne créent leur entreprise, en détectant des signaux faibles dans leur parcours ou leur réseau. En 2025, Signal Fire a levé plus d'1md\$, ce qui démontre la conviction que peuvent avoir les LPs dans cette approche très intégrée du VC data-driven.

InReach Ventures(26) : Parmi les fonds data-driven, InReach fait également figure de pionnier. En effet, le fonds lancé en 2015 utilise une plateforme IA propriétaire nommée DIG. Cette plateforme est structurée en 3 couches : le data layer dont le rôle est de collecter et d'enrichir des données publiques et propriétaires issues notamment de scrapping de sites internet et de réseaux sociaux, l'intelligence layer dont le rôle est d'utiliser de nombreux modèles de Machine Learning afin de détecter les start-ups émergentes, et enfin le workflow layer qui constitue un outil interne de gestion du dealflow, permettant à InReach Ventures d'analyser plusieurs milliers de start-ups par mois, contre quelques dizaines au maximum par un analyste humain. InReach se considère davantage comme une entreprise tech que comme un fonds de venture capital, précisant notamment que plus de 3M\$ ont été investis dans le développement de leur plateforme et que leur équipe compte davantage de profils tech que d'investisseurs.

Correlation Ventures(27) : Une société de gestion lancée en 2006 à San Francisco, gérant environ 500M\$ d'actifs et investissant de la seed à la série B. L'outil développé par Correlation Ventures repose sur une base de données propriétaire extrêmement complète, comprenant des milliers de tours de financement passés ainsi que sur un algorithme de scoring évaluant chaque opportunité selon des facteurs historiques. Ainsi, le modèle permet de prendre des décisions en moins de 24h. L'outil développé par Correlation Ventures propose également un « VC-Matching Tool » permettant d'identifier les investisseurs les mieux placés pour co-investir ou mener les tours suivants.

Nous pouvons conclure que l'état actuel de l'utilisation de l'IA dans les fonds de Venture Capital généralistes révèle un paysage relativement contrasté : d'un côté, une poignée de fonds visionnaires ont démontré la valeur ajoutée tangible d'une approche data-driven et de l'autre, la grande majorité de l'industrie en est encore aux premiers pas d'une transition digitale inéluctable.

Il apparaît de plus en plus clairement que les gagnants de demain – en particulier en ce qui concerne l'accès au capital des LPs – seront ceux qui sauront véritablement tirer parti de l'IA et construire une stratégie interne d'intégration de la technologie. Les prochaines années verront sans doute se généraliser des pratiques aujourd'hui innovantes, dans un secteur du capital-

risque qui, paradoxalement, aura mis du temps à appliquer à lui-même les révolutions technologiques qu'il a contribué à financer.

E] Etude de l'utilisation de l'IA dans les fonds d'investissement biotech

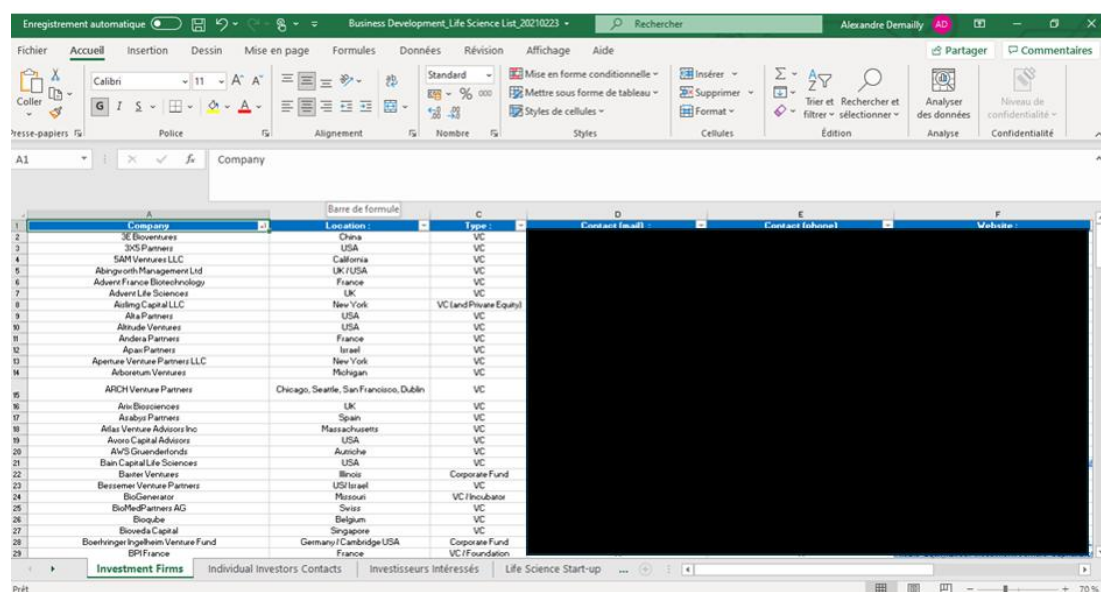
Avertissement : Les données recueillies dans cette partie datent du début de la rédaction de cette thèse, par conséquent, il est nécessaire de les analyser en ayant en tête qu'elles datent de 2021. Dans un écosystème évoluant rapidement et s'intéressant de plus en plus à l'IA, il est nécessaire de se rappeler qu'elles ne sont peut-être plus tout à fait exactes (l'émergence de l'IA générative et du no-code a rebattu les cartes, pour les plus petits fonds notamment).

Pour des raisons de temps, il était compliqué de mener une nouvelle étude – d'autant plus en travaillant désormais dans un fonds d'investissement concurrent.

Afin de savoir ce que pensent les investisseurs en venture capital spécialisés en biotech et en santé de l'usage actuel et futur de l'IA dans leur pratique, une enquête internationale a été menée dans le cadre de cette thèse.

I] Méthodologie et élaboration d'un questionnaire :

Dans un premier temps, une liste de 188 fonds d'investissement a été élaborée. Les noms des fonds ont été récupérés manuellement en parcourant les communiqués de presse des levées de fonds de start-ups santé. La visite des sites Web de ces fonds et l'utilisation d'outils de « Growth Hacking » spécialisés dans le recueil d'adresses électroniques a permis de recueillir les mails de 1441 collaborateurs de leurs équipes. La constitution de cette liste a été quasiment exclusivement manuelle, l'envoi de courriels personnalisés aux collaborateurs a cependant quant à lui été automatisé par publipostage. Cette liste a été constituée en grande majorité par des fonds de capital-risque mais également par quelques fonds corporate de grandes entreprises pharmaceutiques. Le monde du Venture Capital est réputé comme étant difficile d'accès, avec des acteurs souvent très occupés : pour recueillir un nombre significatif de réponses au questionnaire, le choix a donc été fait de contacter un nombre aussi important que possible de fonds et de leurs collaborateurs.



Company	Location	Type
3E Bioventures	China	VC
3GS Partners	USA	VC
SAM Ventures LLC	California	VC
Abingworth Management Ltd	UK / USA	VC
Advent France Biotechnology	France	VC
Advent Life Sciences	UK	VC
Arding Capital LLC	New York	VC (and Private Equity)
Alta Partners	USA	VC
Almudra Ventures	USA	VC
Anders Partners	France	VC
Apax Partners	Israel	VC
Aperture Venture Partners LLC	New York	VC
Arboretum Ventures	Michigan	VC
ARCH Venture Partners	Chicago, Seattle, San Francisco, Dublin	VC
Arvi BioSciences	UK	VC
Asaby Partners	Spain	VC
Atlas Venture Advisors Inc	Massachusetts	VC
Aven Capital Advisors	USA	VC
AVS Gruendertfonds	Austria	VC
Blain Capital Life Sciences	USA	VC
Blaire Ventures	Illinois	Corporate Fund
Bessemer Venture Partners	US / Israel	VC
BioGenerator	Missouri	VC / Incubator
BioMap Partners AG	Swiss	VC
Biocube	Belgium	VC
Biovada Capital	Singapore	VC
Boehringer Ingelheim Venture Fund	Germany / Cambridge USA	Corporate Fund
BT Finance	France	VC / Foundation

Figure 19 - Illustration de la base de données constituée (fonds)

Company	Sex	Firstname	Lastname	Contact	Contact Statut
3E Bioventures Capital	Mrs.				Email le 26/11/20
3E Bioventures Capital	Mr.				Email le 26/11/20
3E Bioventures Capital	Mr.				Email le 26/11/20
3E Bioventures Capital	Mr.				Email le 26/11/20
3E Bioventures Capital	Mrs.				Email le 26/11/20
3x5 Partners	Mr.				Email le 26/11/20
3x5 Partners	Mr.				Email le 26/11/20
3x5 Partners	Mr.				Email le 26/11/20
3x5 Partners	Mr.				Email le 26/11/20
3x5 Partners	Mr.				Email le 26/11/20
3x5 Partners	Mrs.				Email le 26/11/20
3x5 Partners	Mrs.				Email le 26/11/20
3x5 Partners	Mr.				Email le 26/11/20
SAM Ventures	Mr.				Email le 26/11/20
SAM Ventures	Mr.				Email le 26/11/20
SAM Ventures	Mr.				Email le 26/11/20
SAM Ventures	Mr.				Email le 26/11/20

Figure 20 - Illustration de la base de données constituée (investisseurs)

II] Le questionnaire

La première partie du questionnaire élaboré portait sur la caractérisation des fonds :

La première question visait à situer le répondant dans la hiérarchie habituelle d'un fonds mais aussi son rôle particulier, en différenciant notamment les Operating Partner (en charge de l'accompagnement des start-up plus que de l'investissement à proprement parler qui est un rôle qui se développe de plus en plus au sein des fonds dans la logique du CaaS pour « Capital as a Service » également appelé VC-platform) des Partners plus classiques. Y a-t-il une fracture générationnelle au sein des fonds dans le rapport à l'utilisation potentielle de l'Intelligence Artificielle ?

1] Respondent Information

1) You are:

- ☐ Analyst
- ☐ Associate / Senior Associate
- ☐ Principal
- ☐ Partner / Senior Partner
- ☐ Operating Partner
- ☐ General Partner
- ☐ Other: _____

La deuxième question portait sur la nature du fonds : s'agit-il d'un fonds de Venture Capital ou d'un fonds corporate ? Bien que plusieurs fonds corporate fassent partie des investisseurs les plus actifs en capital-risque dans la santé, très peu ont répondu au questionnaire.

2) What is the nature of your structure?

- ☐ VC
- ☐ Corporate Fund

La troisième question a pour objectif de déterminer l'intensité de l'activité du fonds en demandant combien de deals sont réalisés tous les ans en moyenne.

3) How many deals does your structure carry out per year on average?

La quatrième question s'intéresse à la quantité d'argent qui est sous gestion ainsi qu'au nombre de fonds dans lesquels sont réparties ces sommes : cette question a pour objectif de savoir si la taille d'un fonds est corrélée à l'intérêt qu'il peut porter aux nouvelles technologies et aux manières de réinventer ses missions.

4) How much money does your structure have under management? How many different funds are managed by your structure?

La cinquième question concerne le stade de développement auquel les fonds vont rentrer au capital des entreprises qu'ils soutiennent, de la même manière que précédemment, cette question a pour objectif d'identifier des différences dans le rapport à l'Intelligence Artificielle en fonction des caractéristiques de maturité des entreprises auxquelles s'intéresse le fonds d'investissement. Les entreprises plus matures permettant d'avoir accès à des données de qualité supérieure mais la plus-value la plus importante de la technologie étant apportée à un stade précoce, au moment où une start-up biotech a un risque important de ne pas passer à la phase de financement suivante.

5) At what stage of business development does your structure invest?

- ☐ Seed
- ☐ Serie A
- ☐ Late Stage
- ☐ Stage Agnostic

La sixième question s'intéresse aux verticales d'investissement des fonds, afin de déterminer si leur positionnement peut impacter leur appétence pour les technologies d'Intelligence Artificielle.

6) What are the investment verticals of your structure? (Several possible answers)

- ☐ Life Sciences / Pharmaceuticals
- ☐ Life Sciences / Diagnostics
- ☐ Life Sciences / Services ~~BtoB~~
- ☐ Medical Devices

La septième question porte sur les différents objectifs des fonds : en termes de « deal sourcing », de « deal closing » et également en termes de métriques financières, est-ce que le fonds atteint ses objectifs ?

7) Is your structure achieving its objectives?

· In terms of deal sourcing? ☐ Yes ☐ No

· In terms of deal closing? ☐ Yes ☐ No

· In terms of fund financial metrics (IRR, TVPI, DPI)? ☐ Yes ☐ No

La huitième et dernière question de cette première partie porte sur la sensation qu'ont les investisseurs concernant une amélioration possible de leurs performances financières. Le développement d'un système de prédiction basé sur les données n'aurait en effet que peu d'intérêt (uniquement le temps gagné permis par l'outil) si tous les investisseurs avaient la conviction d'être au maximum de ce qu'il est possible de faire en termes de sélection des start-up et de résultat financier.

8) Do you think these metrics could be improved? ☐ Yes ☐ No

La deuxième partie du questionnaire porte quant à elle sur les pratiques actuelles des investisseurs et leur rapport à la Data Science ou à l'Intelligence Artificielle :

La première question vise à appréhender le niveau de connaissance des investisseurs biotechs/healthcare des initiatives menées par les fonds d'investissement généralistes en matière de développement d'outils d'Intelligence Artificielle. Elle rejoint la cinquième question qui s'intéresse au niveau de connaissances techniques des investisseurs concernant le fonctionnement de l'Intelligence Artificielle. Dans l'optique du développement d'un produit, il est toujours intéressant de pouvoir constater la maturité d'un marché et le chemin restant à parcourir pour le défricher.

II) Information on the use of Data Science and Machine/Deep Learning Technologies

1) Are you aware of initiatives using Data Science and Machine/Deep Learning technologies in the field of Venture Capital?

☐ Yes ☐ No

5) What is your level of understanding of Data Science and Machine/Deep Learning techniques?

☐ I understand and practice.

☐ I understand and know the steps involved in building a Data Science model.

☐ I have read some articles on the subject.

☐ It's a field I don't know anything about.

La deuxième question porte sur l'utilisation de bases de données business spécialisées dans les Biotechs. Il est en effet courant d'utiliser dans les fonds d'investissement généralistes des bases de données business telles que Pitchbook, ou encore CB Insights. Il existe des bases spécialisées sur les Biotechs et entreprises du monde de la santé. Leur accès est cependant souvent onéreux et leur infrastructure n'est pas équipée d'outils d'analyse.

2) Do you use life sciences business databases?

☐ Yes ☐ No

La troisième question incite les investisseurs à se projeter dans le futur en leur demandant s'ils pensent que les technologies de l'Intelligence Artificielle feront partie du futur de l'investissement. Les réponses à cette question seront importantes pour prendre le pouls de l'écosystème du financement et pour savoir si les récentes avancées de l'Intelligence Artificielle (que nous illustrerons plus tard dans cette thèse) sont vues comme étant les prémices d'un mouvement de fond ou comme quelque chose qui restera marginal.

3) Do you think Data Science and Machine / Deep Learning technologies are part of Venture Capital's future?

☐ Yes ☐ No ☐ I don't know

La quatrième question demande aux investisseurs s'ils pensent que la technologie des données pourrait améliorer leurs retours financiers.

4) Do you think that these technologies will be able to improve the financial performance of Life Sciences / Healthcare VC-firms?

☐ Yes ☐ No ☐ I don't know

La sixième question s'intéresse au fait que les équipes d'investisseurs intègrent ou pas des Data Scientists. Les septième et neuvième questions s'intéressent aux champs explorés par les Data Scientists de l'équipe d'investissement. Cela permettra de déterminer s'il y a un marché pour un service externalisé et quels services et outputs sont particulièrement nécessaires à la décision des fonds.

6) Does your structure integrate Data Science skills, a team of Data Scientists?

☐ Yes ☐ No

7) If yes, on which themes does your team of Data Scientists or ML engineers work? (several answers possible)

- ☐ Deal Sourcing
- ☐ Screening Process
- ☐ Due Diligence / Scoring / Risk Assessment
- ☐ Supporting portfolio companies (talent sourcing, strategic decisions, etc.)
- ☐ Market research (benchmarking and market insights)
- ☐ Fund Raising

9) Assuming a Data Science tool dedicated to due diligence, which metrics would be the most useful for you in making an investment decision? (several answers possible)

- ☐ Probabilities of business success and exit (refunding, M&A, IPO)
- ☐ Valuation of the company
- ☐ Scores evaluating the quality of the company (founding team, product, intellectual property, product-market fit, etc)
- ☐ Competition Mapping
- ☐ Optimized use of the raised money
- ☐ Other: _____

La huitième question s'intéresse, en cas d'absence de data scientists au sein de l'équipe, aux raisons qui poussent le fonds à ne pas en intégrer. Il s'agit d'une question à choix multiples et les raisons proposées sont, notamment : un manque d'intérêt pour ce champ d'exploration, un manque de temps ou d'argent, des difficultés pour mener à bien le recrutement de Data Scientists.

8) If your structure does not have a Data Science team, what do you think are the reasons?
(several answers possible)

- ☐ We have not had any interest in Data Science so far and we have not taken the time to think about it yet.
- ☐ We don't think this would differentiate us from other funds / we don't believe in the contribution of Data Science in Venture Capital.
- ☐ It's too time-consuming
- ☐ It's too expensive.
- ☐ Difficulties in recruiting Data Science talent
- ☐ Other: _____
- ☐ I don't know

Les questions 10 et 11 vont, quant à elles, plus avant dans les réflexions internes de mise en place d'un outil en demandant aux investisseurs s'ils ont déjà travaillé avec des entreprises leur fournissant des services d'analyse des données et s'ils envisagent cette solution dans le futur.

10) Have you ever worked with companies allowing you to outsource Data Science skills and implement decision support tools specific to your investment thesis?

- ☐ Yes ☐ No

11) Depending on the development of the offer and its quality, do you envisage this solution in the future?

- ☐ Yes ☐ No ☐ I don't know.

L'ensemble du questionnaire adressé aux investisseurs ainsi que l'ensemble des réponses recueillies sont disponibles en annexe de cette thèse.

III] Les résultats du questionnaire

Plusieurs centaines de personnes ont été contactées et 60 réponses au questionnaire ont été reçues.

Les résultats obtenus ont montré que les investisseurs utilisaient extrêmement peu les techniques de Data Science et d'Intelligence Artificielle, bien qu'ils soient conscients d'un potentiel très important de ces techniques pour l'amélioration de leur pratique. Il est d'autant plus paradoxal que, malgré une prise de conscience croissante du potentiel de l'intelligence artificielle, les équipes de data scientists demeurent rares au sein des fonds, et que les initiatives d'intégration de ces technologies restent encore marginales. Certains acteurs de l'investissement ont rapporté le paradoxe que les fonds investissent dans des entreprises utilisant l'Intelligence Artificielle tout en n'utilisant pas eux mêmes cette technologie.

Avant d'entrer dans les détails des résultats obtenus, il convient de les nuancer en rappelant certaines limites. D'une part, ces résultats peuvent être entachés d'un biais de sélection important : il est en effet probable que les répondants soient majoritairement des investisseurs

déjà sensibilisés à l'intelligence artificielle, ce qui pourrait orienter les conclusions. D'autre part, le nombre relativement restreint de réponses pourrait limiter la portée statistique des résultats. À notre connaissance, cette enquête demeure néanmoins la plus importante à ce jour sur ce sujet, tous investisseurs confondus, et la seule étude spécifiquement menée auprès des acteurs spécialisés en santé et biotechnologies.

Sur la caractérisation des fonds d'investissement :

Les réponses à la première question révèlent une répartition relativement homogène des répondants, reflétant vraisemblablement la distribution habituelle des rôles au sein des sociétés d'investissement. La majorité des participants sont des Partners, des Associates et des Analysts, suivis par les Principals. Cette répartition constitue un bon indicateur de l'absence de biais majeur dans la composition de l'échantillon.

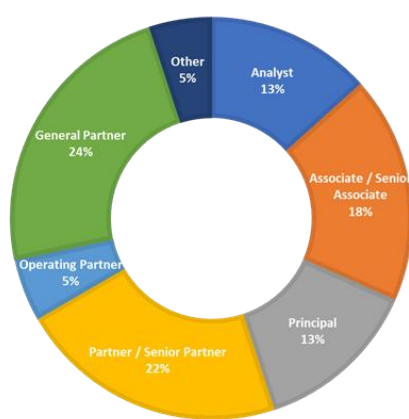


Figure 21 - Rôle du répondant

Les réponses à la deuxième question renseignent sur la nature des structures auxquelles appartiennent les répondants : on observe une prédominance nette des fonds de capital-risque par rapport aux fonds corporate. Cette disparité peut s'expliquer, en partie, par des besoins différenciés. En effet, dans le cas d'un fonds corporate, l'évaluation d'un investissement ne repose pas uniquement sur des critères financiers, mais implique également d'apprécier, à partir des données disponibles, le niveau de compatibilité stratégique et la plus-value potentielle qu'une start-up peut apporter au grand groupe.

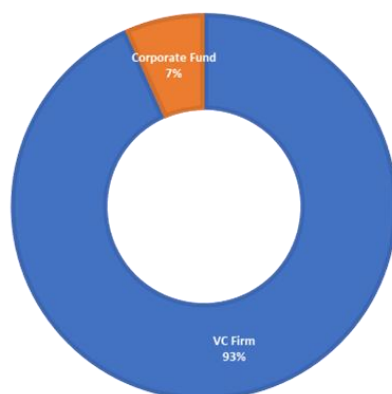


Figure 22 - Nature du fonds du répondant

Les réponses aux deux questions suivantes montrent que les investisseurs ayant participé à l'enquête proviennent d'équipes d'investissement variées, tant en termes d'activité que de taille. La moyenne du nombre de deals réalisés par an s'élève à 55,2, tandis que la médiane est de 5, traduisant une forte dispersion dans les pratiques d'investissement. En ce qui concerne les montants sous gestion, la moyenne est de 895,4 millions de dollars, avec une médiane à 400 millions de dollars.

L'immense majorité des répondants exercent au sein de fonds intervenant en phase Seed, en série A, ou adoptant une approche agnostique vis-à-vis du stade d'investissement. À l'inverse, les investisseurs spécialisés dans les phases late-stage sont très peu représentés dans l'échantillon.

Plusieurs hypothèses peuvent être avancées pour expliquer cette sous-représentation. D'une part, les start-ups en série B, C ou au-delà sont souvent considérées comme déjà partiellement dérisquées, réduisant potentiellement l'intérêt d'intégrer des outils d'intelligence artificielle dans les processus de due diligence. D'autre part, du fait de la courbe de mortalité des start-ups, leur nombre diminue mécaniquement à mesure qu'on progresse dans les stades d'investissement, ce qui peut se traduire par un écosystème moins dense au niveau late-stage.

Il n'est pas exclu non plus que cette différence de représentation soit liée à des caractéristiques structurelles du marché du financement lui-même — les fonds late-stage étant souvent moins nombreux, plus importants en volume financier mais moins actifs en nombre d'opérations — ou à des biais liés à la méthode de constitution de l'échantillon, notamment au moment du recueil des fonds sollicités.

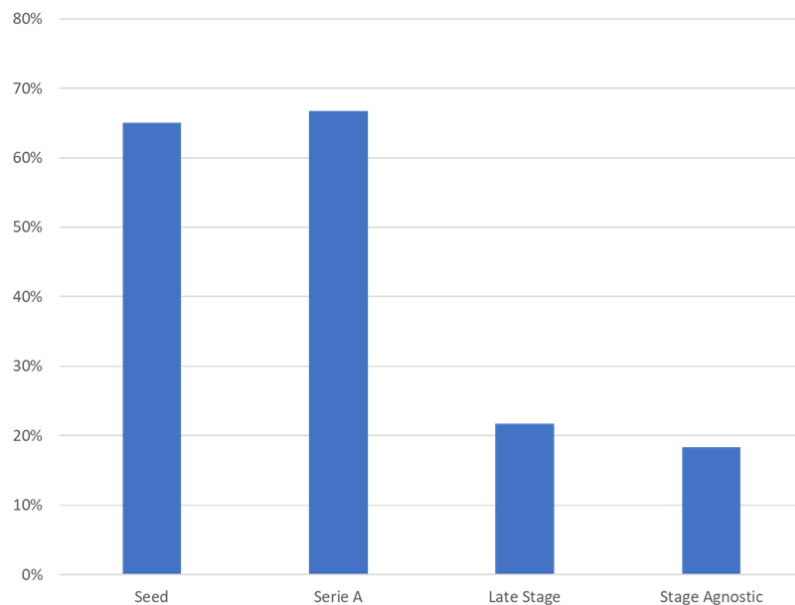


Figure 23 - Typologie d'intervention des fonds des répondants

En ce qui concerne les verticales d'investissement couvertes, les réponses montrent que la plupart des fonds interrogés interviennent sur un large spectre de thématiques, sans spécialisation exclusive. Ils déclarent ainsi s'intéresser à la fois aux entreprises développant des médicaments, des outils de diagnostic, des services BtoB en biotechnologie, ainsi qu'à des solutions numériques dans les domaines de la e-santé, de la m-santé ou de l'informatique médicale.

Cette diversité témoigne de la tendance croissante des investisseurs à adopter une approche multidisciplinaire, reflet de la porosité croissante entre les secteurs traditionnels de la santé, du numérique et des biotechnologies. Malgré cette transversalité, l'analyse des portefeuilles des fonds participants montre que les entreprises de biotechnologie, de dispositifs médicaux (Medtech) et de e-santé constituent les segments les plus représentés, confirmant leur rôle central dans l'écosystème de l'innovation en santé.

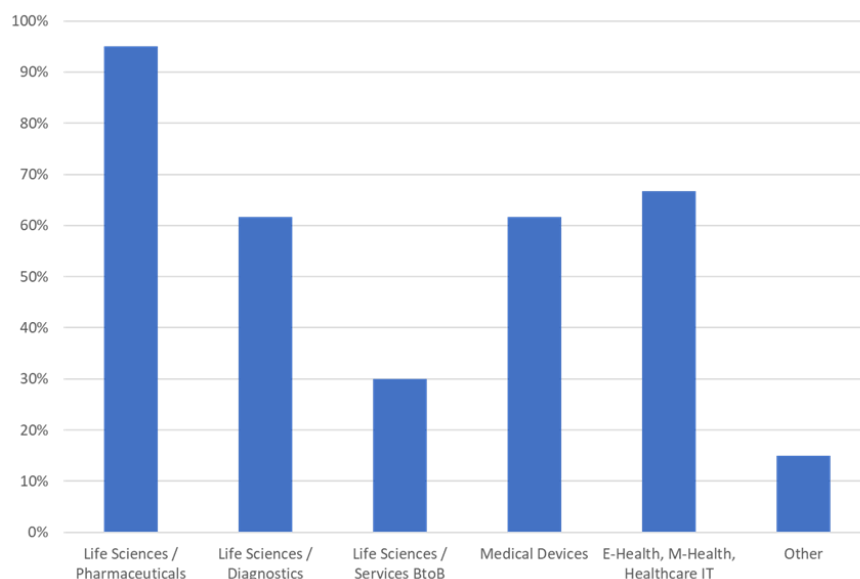


Figure 24 - Verticales des fonds des répondants

En matière de flux d'affaires, les réponses indiquent que la plupart des fonds atteignent leurs objectifs, tant en ce qui concerne le sourcing de dossiers que la réalisation d'investissements (deal closing). Cette observation suggère que les équipes d'investissement disposent de processus efficaces pour identifier et sélectionner un volume suffisant d'opportunités.

Il convient toutefois de nuancer cette efficacité quantitative par une réflexion qualitative : le fait d'atteindre un nombre cible d'investissements ne garantit pas nécessairement l'optimalité des opportunités sélectionnées. Un sourcing plus ciblé, mieux aligné avec les thèses d'investissement spécifiques à chaque fonds, pourrait permettre d'améliorer la qualité des dossiers analysés et, potentiellement, la performance globale des portefeuilles.

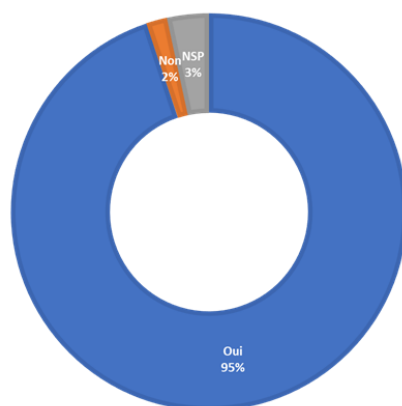


Figure 25 – Atteinte des objectifs en termes de deal sourcing

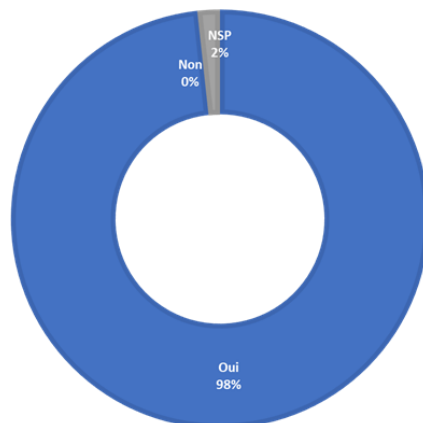


Figure 26 - Atteinte des objectifs en termes de deal-closing

Les réponses concernant les métriques financières des fonds apportent des enseignements particulièrement intéressants. D'une part, la majorité des fonds déclarent atteindre les objectifs qu'ils se sont fixés — objectifs que l'on peut supposer calibrés de manière réaliste, voire conservatrice. D'autre part, un nombre tout aussi significatif d'investisseurs estiment qu'ils pourraient améliorer leurs performances financières.

Ce décalage apparent entre atteinte des objectifs et perception d'un potentiel d'amélioration suggère que les critères de succès internes ne traduisent pas nécessairement un niveau d'ambition maximal. Il reflète peut-être aussi une lucidité sur les marges d'optimisation possibles, notamment dans la sélection des opportunités, l'accompagnement des participations ou le timing des sorties.

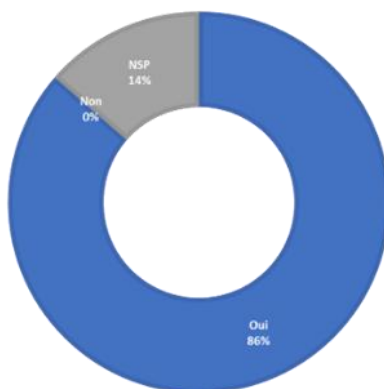


Figure 27 - Atteinte des objectifs financiers

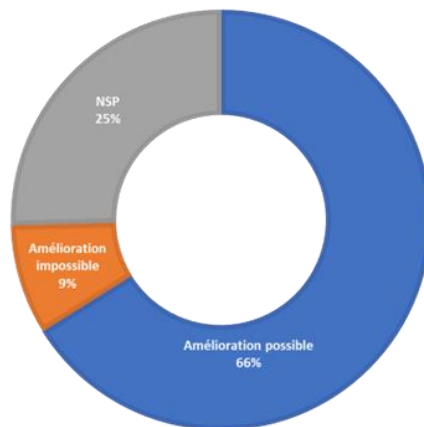


Figure 28 - Evaluation de l'amélioration possible des métriques financières

Sur le rapport des structures des répondants à l'Intelligence Artificielle :

Les réponses indiquent que la majorité des participants ont connaissance des initiatives menées dans le secteur du capital-investissement généraliste en matière d'outils d'aide à la décision fondés sur les données (data-driven). Ce constat, associé à celui selon lequel de nombreux répondants estiment qu'un tel outil pourrait contribuer à améliorer les performances financières des fonds, témoigne d'un niveau de sensibilisation élevé. Il s'agit là d'un signal encourageant quant à la relative maturité du marché sur ces enjeux technologiques.

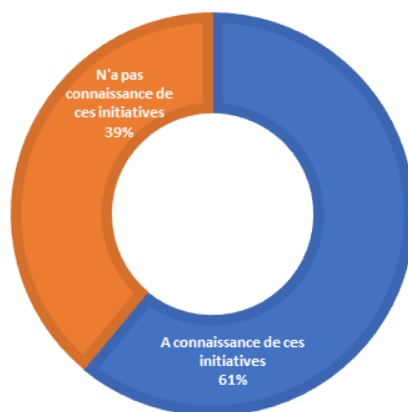


Figure 29 - Evaluation des connaissances en matières d'utilisation de l'IA en VC

Ce constat mérite toutefois d'être nuancé : il convient de s'interroger sur les barrières qui freinent encore les fonds dans le développement de tels outils en interne, alors même que de nombreuses initiatives émergent dans le capital-investissement généraliste.

Comprendre ces freins — qu'ils soient d'ordre technique, organisationnel, culturel ou économique — constitue l'un des objectifs de cette thèse. C'est également l'une des dimensions que nous avons explorées à travers le questionnaire adressé aux investisseurs.

La question suivante portait sur l'utilisation, par les fonds, de bases de données spécialisées dans le secteur des biotechnologies. De nombreuses entreprises de type data providers proposent en effet des bases régulièrement mises à jour, principalement à destination des investisseurs. Si une majorité des répondants déclare y recourir, ce n'est toutefois pas le cas de l'ensemble des fonds. Plusieurs hypothèses peuvent être avancées pour expliquer cette hétérogénéité. Il est possible, d'une part, que certains fonds jugent ces services insuffisamment utiles, dans la mesure où ils fournissent essentiellement des données brutes, sans outils d'analyse ni accompagnement à la décision. D'autre part, il n'est pas à exclure qu'une partie des acteurs du capital-risque biotech demeure encore relativement éloignée des approches fondées sur les données — qu'il s'agisse d'evidence-based investment ou de logique véritablement data-driven.

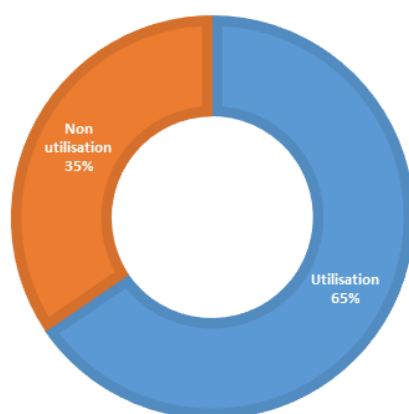


Figure 30 - Evaluation des bases de données Biotech

La question suivante portait sur la perception qu'ont les investisseurs du rôle futur de l'intelligence artificielle dans leur métier. Comme évoqué précédemment, une très large majorité d'entre eux estime que cette technologie jouera un rôle important dans la transformation de leurs pratiques professionnelles. Seuls 8 % des répondants expriment une opinion contraire, ce qui témoigne d'un consensus fort autour de l'idée que l'IA constitue un levier d'évolution significatif pour le capital-risque, y compris dans les secteurs spécialisés comme les biotechnologies ou la santé.

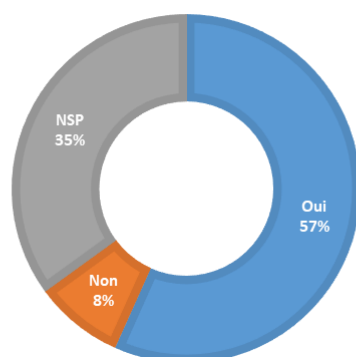


Figure 31 - L'IA fera-t-elle partie de l'avenir du VC ?

Seuls 17% des investisseurs répondants pensent que l'Intelligence Artificielle ne leur permettra pas d'améliorer les performances financières de leurs fonds.

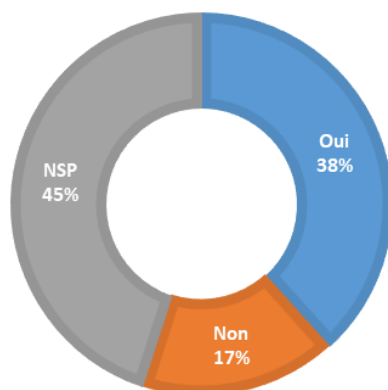


Figure 32 - Pensez-vous que l'IA permettra d'améliorer les perf. de votre fonds ?

La question suivante visait à évaluer le niveau de compréhension de l'intelligence artificielle parmi les investisseurs. Seuls 10 % des répondants déclarent ne disposer d'aucune connaissance dans ce domaine. Toutefois, la majorité admet ne posséder qu'une compréhension partielle, voire superficielle, des principes et des mécanismes sous-jacents à cette technologie.

Ce constat souligne l'importance que pourrait revêtir un effort de pédagogie dans toute démarche de diffusion ou de commercialisation d'outils d'intelligence artificielle à destination des professionnels du capital-risque. Une stratégie d'adoption efficace ne saurait faire l'économie d'un accompagnement explicatif permettant de clarifier les apports réels de l'IA, ses limites, ainsi que son intégration dans les processus de décision existants.

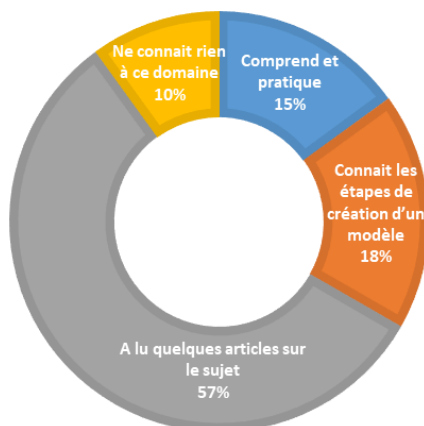


Figure 33 - Evaluation de la compréhension de l'IA par les VC

En dépit d'une compréhension relativement large des avantages que pourrait représenter l'utilisation de l'intelligence artificielle, et d'un niveau de connaissance globalement satisfaisant de la technologie, peu de structures ont franchi le cap de l'intégration opérationnelle. Seuls environ 15 % des répondants déclarent disposer de data scientists au sein de leurs équipes.

Ce décalage entre l'intérêt exprimé pour l'IA et la faiblesse des ressources internes mobilisées pour l'exploiter suggère l'existence de freins structurels ou organisationnels persistants, que nous explorerons dans les sections suivantes.

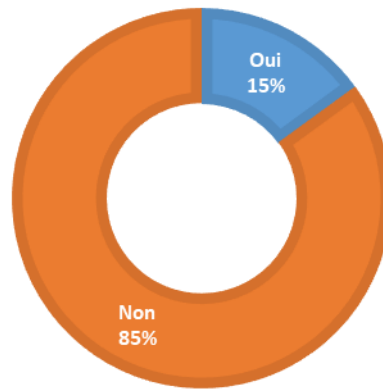


Figure 34 - Evaluation de la présence d'une équipe data au sein du fonds VC

La question suivante visait à identifier les freins au recrutement de spécialistes en data science au sein des fonds de capital-risque. Une telle intégration, bien que fréquente dans les sociétés de gestion intervenant sur les marchés cotés, reste marginale dans l'univers du Venture Capital.

Plusieurs raisons étaient proposées dans le questionnaire, mais aucune ne s'est nettement détachée : aucune n'a recueilli plus de 50 % des réponses. Ce résultat suggère que les obstacles à l'intégration de profils data sont pluriels et diffus, relevant sans doute de facteurs combinés tels que le coût, la culture d'investissement, l'absence de cas d'usage immédiats perçus, ou encore la difficulté à articuler compétences techniques et expertise sectorielle dans des structures souvent de taille réduite.

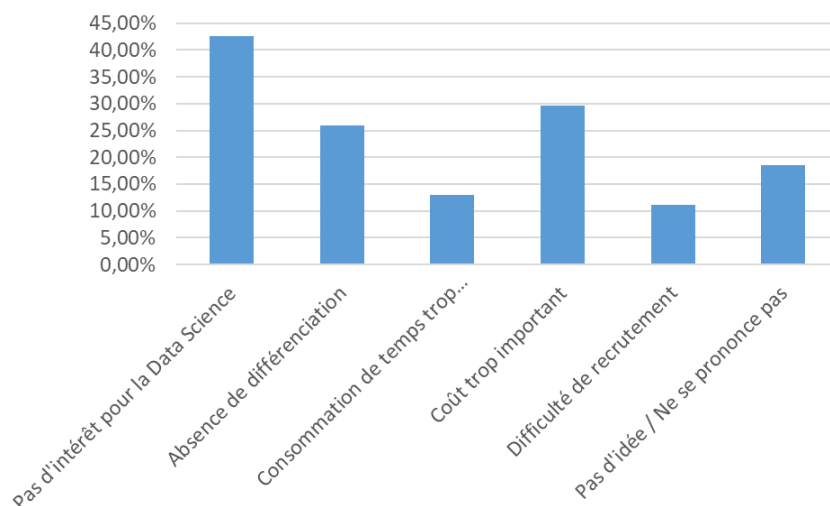


Figure 35 - Les raisons de l'absence d'une équipe data (par % de répondants)

La question suivante portait sur les missions confiées aux équipes de data science dans les fonds ayant intégré de tels profils. Deux types de tâches ressortent comme majoritairement attribuées, ayant recueilli plus de 50 % des réponses : la due diligence et la recherche de marché (market research), ce qui est cohérent avec les usages les plus attendus de l'intelligence artificielle dans l'analyse d'opportunités d'investissement.

D'autres fonctions, telles que le deal sourcing et le screening automatisé des dossiers, ont également été mentionnées de manière significative. Il convient toutefois de relativiser la portée de ces résultats, en raison du faible nombre de répondants concernés par cette question, ce qui limite la robustesse des conclusions que l'on peut en tirer.

Enfin, le terme de data scientist est ici employé au sens large : les équipes évoquées peuvent inclure des profils variés, tels que des data analysts, des machine learning engineers, ou d'autres experts en science des données.

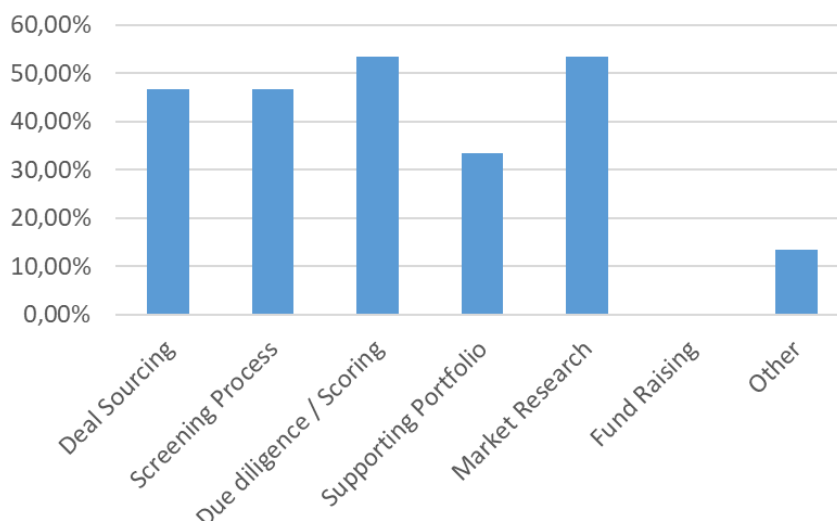


Figure 36 - Les domaines d'intervention des équipes data

Sur l'utilisation de solutions extérieures aux fonds d'investissement :

Le questionnaire constituait également une opportunité d'explorer la relation des investisseurs à l'égard de solutions externes d'aide à la décision. La première question de cette section portait sur l'expérience des fonds en matière de collaboration avec des entreprises spécialisées en data science appliquée à l'investissement.

Les résultats montrent qu'une très large majorité des répondants n'a jamais eu recours à ce type de prestataires. Ce constat s'explique en partie par le caractère encore émergent de ce segment industriel. À ce jour, il n'existe que très peu — voire aucune — entreprise véritablement spécialisée dans l'externalisation d'analyses data-driven à destination des fonds d'investissement en sciences de la vie. Le recours à des solutions externes d'analytics dans ce domaine reste donc marginal, voire inexistant.

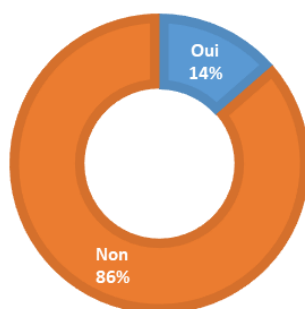


Figure 37 - Avez-vous déjà collaboré avec des entreprises proposant des services d'analyse data pour les VC ?

Nous avons également profité de ce questionnaire pour sonder l'intérêt des fonds à l'égard du développement futur d'un outil d'aide à la décision d'investissement. Il s'agissait ainsi de tester le positionnement du marché et d'évaluer la réceptivité des investisseurs à une telle solution, en tenant compte des besoins perçus, des freins potentiels et des conditions d'adoption envisageables.

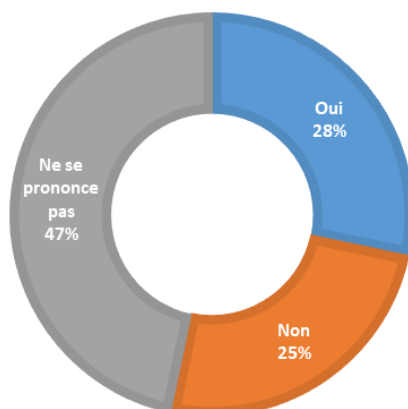


Figure 38 - Envisagez-vous, en fonction du développement de l'offre, d'utiliser des outils d'aide à la décision ?

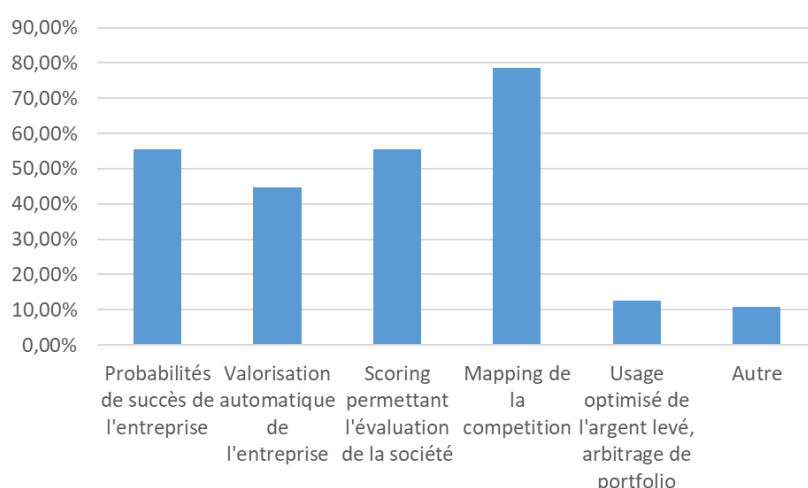


Figure 39- Quelles fonctionnalités d'un tel outil seraient les plus importantes ?

Nous avons également interrogé les VCs quant aux fonctionnalités qui seraient pour eux les plus importantes au sein d'un outil d'aide à la décision d'IA. Il est difficile de tirer une conclusion des réponses obtenues, tant un grand nombre de fonctionnalités semblent importantes à une majorité de VC. Une proposition de fonctionnalité de « portfolio management » visant par exemple à arbitrer les réinvestissement ne semble pas utile.

En conclusion, et bien qu'il soit nécessaire de garder en tête que ces données ont été récoltées il y a plusieurs années, il apparaît nettement que l'intérêt des VCs biotech pour l'IA est fort (d'autant plus qu'il s'est probablement renforcé depuis...) mais que les initiatives concrètes d'implémentation de systèmes d'aide à la décision étaient encore à ce moment là relativement embryonnaires.

F] IA et Propriété intellectuelle dans le contexte des VC Biotechs :

I] Introduction et enjeux de l'analyse de brevets dans les décisions d'investissement en biotech

A) Importance de l'analyse des brevets

Dans le secteur de la santé, et en particulier dans celui des Biotechs, la propriété intellectuelle joue un rôle important dans la prise de décision des investisseurs en capital-risque. En effet, en addition des savoir-faire gardés secrets au sein de l'entreprise – en ce qui concerne par exemple des méthodes de fabrication – les brevets constituent l'une des principales formes de protection intellectuelle des entreprises biotechs. Ils sont en effet en mesure d'offrir un avantage concurrentiel important durant une période de temps donné.

En échange de la publication des informations concernant son innovation, l'entreprise jouira en effet d'une protection légale contre la concurrence qui n'aura pas le droit de contrefaire son innovation. Cette protection constitue donc une barrière à l'entrée forte pour les concurrents et le monopole de marché pouvant être ainsi créé est particulièrement important dans le secteur de la biotech où les coûts de R&D sont élevés et où la rentabilité repose souvent sur une capacité à dominer des niches de marché. Les investisseurs sont par conséquent davantage attirés par des entreprises présentant des portefeuilles de brevets construits de manière à les protéger efficacement et durablement des nouveaux entrants. A ce titre, l'entreprise peut être propriétaire de ses propres brevets ou avoir signé des contrats de licences avec les organismes de recherche à l'origine des innovations.

Le portefeuille de brevets d'une biotech est également un élément essentiel de sa valorisation. En effet, cette valorisation est directement liée à la capacité de l'entreprise à générer des flux de revenus futurs et, comme discuté au paragraphe précédent, ces revenus futurs sont directement liés à la portée et à la qualité du portefeuille de brevets d'une entreprise. Par ailleurs, ces brevets offrent également des opportunités de monétisation supplémentaires, notamment à travers les licences (une entreprise « achète » le droit d'utiliser le brevet d'une autre entreprise dans un domaine spécifique – en général, le paiement se fait selon plusieurs modalités : upfront, milestones et royalties). Pour les investisseurs, cela signifie que même avant la mise sur le marché d'un produit final, une entreprise biotech peut avoir des flux de revenus tangibles issus de partenariats et de licences. Ce potentiel de génération de revenus intermédiaires peut rehausser la valorisation de l'entreprise – à l'inverse, si ce sont les principaux actifs de la société qui sont licenciés et sur lesquels s'opère le partage de la valeur future, alors la valorisation de l'entreprise diminuera. Enfin, si l'aventure de la start-up se passe mal, les brevets quant à eux conserveront leur valeur et pourront potentiellement être cédés à une autre société afin d'assurer un retour minimal aux investisseurs.

Ainsi, les brevets d'une entreprise sont examinés selon différents critères : leur qualité, leur solidité juridique, leur couverture géographique et leurs capacités à soutenir les ambitions commerciales de l'entreprise. Des études en liberté d'exploitation sont également menées préalablement à l'investissement si elles n'ont pas été réalisées par l'entreprise. Ces analyses ont pour objectif de démontrer qu'aucun brevet extérieur ne pourrait contrarier l'utilisation des brevets de l'entreprise. Il est rare que les fonds d'investissement disposent de l'expertise en

interne et du temps nécessaire durant la due-diligence pour explorer tous ces critères en détails. Ainsi, ces différentes analyses est souvent externalisées et confiées à des cabinets spécialisés, à la pertinence variable.

B) Défis spécifiques liés au domaine de la biotech

Le secteur de la biotech - et des sciences de la vie au sens large – présente des défis très importants en ce qui concerne l'analyse des brevets. Il s'agit en effet d'un domaine nécessitant, pour réaliser des analyses pertinentes, des connaissances scientifiques relativement poussées. Les brevets en biotech contiennent souvent des descriptions longues et détaillées de procédés scientifiques, de structures chimiques complexes ou de résultats expérimentaux. La complexité technique des concepts et de la terminologie employée rend l'analyse complexe et chronophage. Par ailleurs, la vitesse d'évolution et d'innovation dans le domaine des biotechs, rend difficile l'évaluation de la capacité des brevets à rester pertinents à long terme.

Outre la complexité des brevets inhérentes à la biotech, une des autres difficultés du secteur repose sur l'évolution rapide de ce domaine technologique, avec l'émergence constante de nouvelles techniques et découvertes. Ce phénomène devrait par ailleurs s'amplifier avec l'augmentation de l'utilisation de l'IA dans les laboratoires de recherche. Analyser la capacité d'un brevet à rester pertinent à long terme est particulièrement difficile dans un domaine aussi dynamique.

Enfin, il est fréquent en biotech que les brevets couvrent des aspects différents mais interdépendants d'une même technologie, ainsi l'utilisation de cette technologie peut nécessiter la mise en place de licences entre différents acteurs. Ce phénomène de chevauchement, appelé « patent thicket »(28) rend l'analyse encore plus complexe : il est en effet nécessaire d'évaluer le risque de dépendance à d'autres portefeuilles de brevets et que la stratégie de l'entreprise permettra d'éviter des conflits juridiques.

II] Applications de l'IA dans les analyses de brevets

Les brevets, documents souvent volumineux et complexes bénéficient tout particulièrement des avancées récentes de l'IA générative. Contrairement aux méthodes d'IA plus classiques (recherche booléenne, analyses statistiques, NLP, etc) qui pouvaient atteindre leurs limites dans le traitement des brevets, les modèles génératifs permettent une compréhension contextuelle plus riche.

A) Analyses sémantiques et extractions d'informations

L'IA générative bouleverse le cadre des techniques de traitement automatique du langage, traditionnellement utilisées pour extraire des informations des brevets. En effet, des modèles de langage type LLM peuvent rendre possible des recherches par similarité de sens et des résumés automatiques. Plusieurs outils intègrent déjà ces analyses sémantiques, afin de renforcer, par exemple, des moteurs de recherche de brevets. C'est notamment, le cas d'Octimine(29), un outil développé par Dennemeyer, incluant plusieurs fonctionnalités :

- Analyse de paysage de brevets ;
- Monitoring : afin de garder un œil sur la concurrence, de détecter de nouvelles tendances technologiques dans le secteur d'innovation d'intérêt et également de détecter les publications de nouveaux brevets proches de ceux existants dans notre portefeuille ;
- L'une des autres fonctionnalités porte sur l'évaluation des risques juridiques

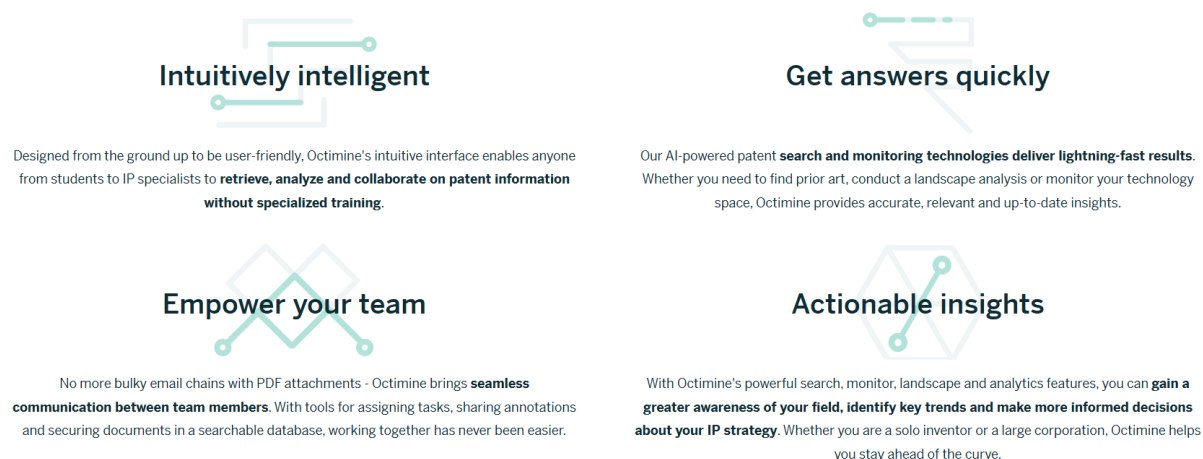


Figure 40 - Caractéristiques de l'outil Octimine

L'IA générative peut également prendre la forme d'un chatbot spécialisé pour interroger la base de brevets. De telles interfaces (disponibles par exemple dans des outils comme PatentPal(30)) réduisent la barrière technique pour l'utilisateur : un ingénieur ou analyste peut poser des questions complexes (par ex. « Quels brevets existent dans la famille X concernant la protéine Y ? ») et obtenir rapidement une synthèse contextualisée. Cette approche contraste avec la recherche manuelle par mots-clés et requiert moins d'expertise en requêtage technique tout en améliorant la précision.

B) Cartographie des technologies et analyse des tendances

Dans les méthodes classiques, la cartographie technologique se fait souvent via des réseaux de citations, des clusters de co-occurrence de termes ou des matrices de fonctions techniques. L'IA générative apporte une nouvelle dimension : en associant l'apprentissage profond et la génération de représentations, elle produit des visualisations plus dynamiques et prédictives. La génération de ces cartes interactives où les clusters de brevets apparaissent de manière plus naturelle qu'avec les méthodes traditionnelles peut permettre de prédire les futures combinaisons technologiques.

Des sociétés comme LexisNexis (31)proposent par exemple ce type de services.

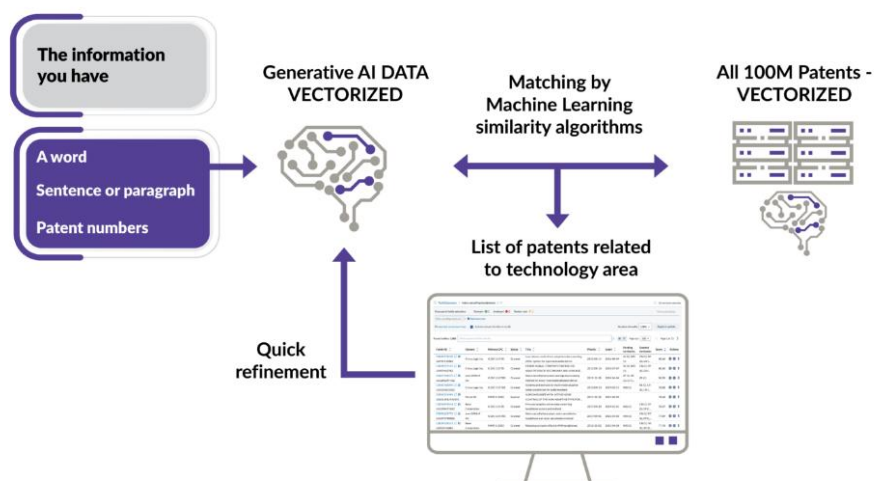


Figure 41 - Schéma de fonctionnement de l'outil de cartographie et d'identification de l'état de l'art de LexisNexis (issu du site web de la société)

C) Evaluation de la qualité des brevets

L'évaluation traditionnelle de la qualité des brevets repose sur des critères quantitatifs (nombre de revendications, citations reçues, etc) et sur l'expertise métier (jugement de la clarté des revendications, de l'innovation réelle, etc). Il s'agit d'une analyse chronophage reposant sur une longue expérience humaine des brevets.

L'IA, grâce à sa capacité à intégrer le contenu technique peut déduire de l'inventivité d'une revendication et de la solidité du brevet mieux qu'un simple comptage de citations. Ces systèmes génèrent des données « permettant d'évaluer la qualité des brevets et de prédire leurs trajectoires d'innovation futures ».(32)

D) Détection des conflits et risques juridiques

La détection des conflits juridiques potentiels est cruciale pour les investisseurs biotech. Classiquement, cela implique des Freedom-to-operate analysis, une analyse permettant de vérifier que l'on ne risque pas enfreindre le droit de propriété intellectuelle d'un tiers en procédant à la commercialisation, à l'importation ou à la fabrication d'un produit.

Les IA permettent d'automatiser le screening des brevets, d'extraire automatiquement les revendications pertinentes afin d'identifier le potentiel de blocage et d'enfin hiérarchiser les risques en fonction de la proximité technique et juridique des brevets détectés.

E) Simulation de stratégies de propriété intellectuelle

L'IA générative peut également servir d'outil prospectif pour la planification stratégique. Par exemple, elle peut modéliser différents scénarios d'évolution de portefeuille brevets et de réactions concurrentielles. On parle alors de passer d'une démarche réactive à une démarche proactive : l'entreprise peut, grâce à l'IA, simuler les conséquences de différents choix stratégiques (nouveaux segments à breveter, collaborations à explorer, risques de contentieux) et ainsi optimiser sa politique de PI face à la concurrence.

III] Intégration de l'IA & bénéfices et limites de son utilisation par les VC dans l'analyse des brevets biotech

Les cas d'usage et les outils mentionnés plus haut permettent de démontrer la pertinence et la possibilité de l'utilisation de l'IA dans le domaine de la propriété intellectuelle.

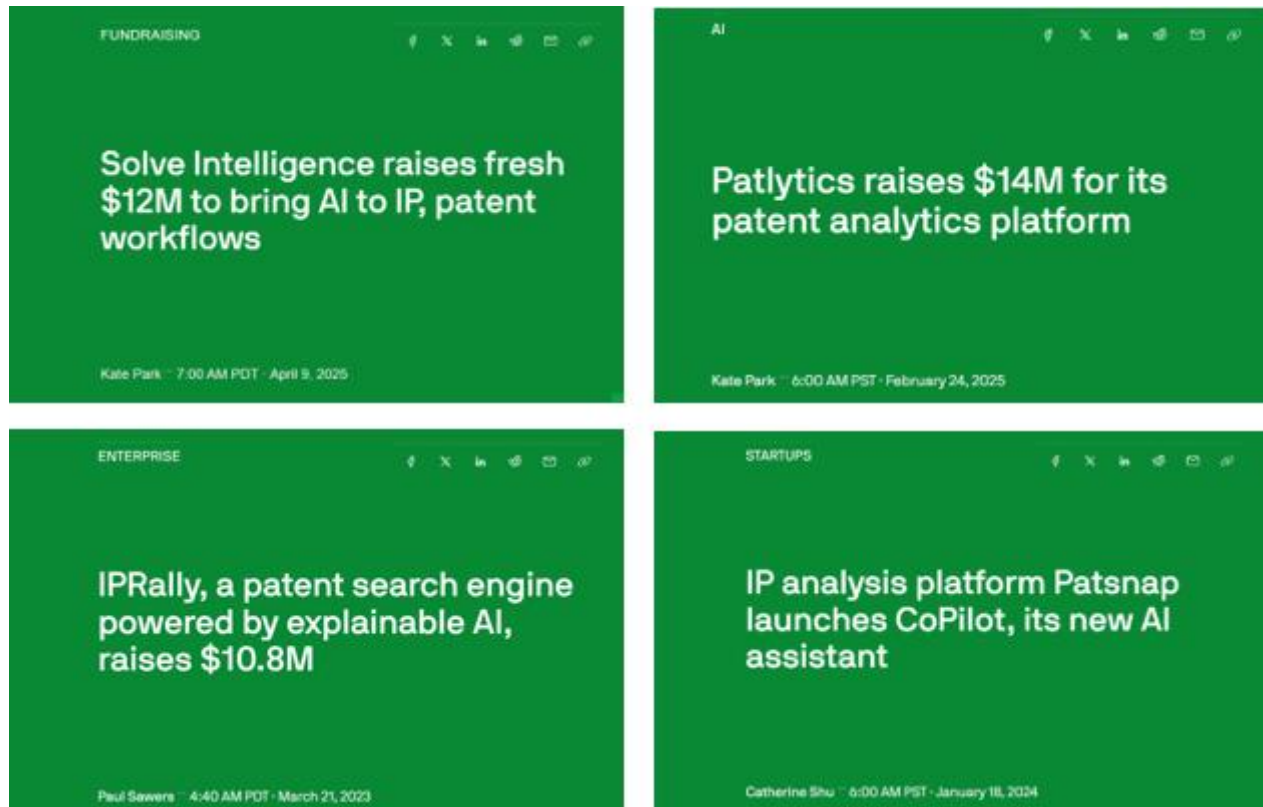


Figure 42 - L'utilisation de l'IA en IP devient de plus en plus populaire

Les nombreuses levées de fonds récentes de start-ups développant des outils d'IA travaillant sur des documents de propriété intellectuelle illustrent l'incontestable dynamisme du secteur.

A ce titre, nous pouvons mentionner :

- Solve Intelligence (33) : Une start-up britannique dont le produit développe plusieurs fonctionnalités dont un éditeur de documents en ligne intégrant un copilote IA spécialisé en IP aidant les professionnels à améliorer la qualité des brevets rédigés. Ce copilote peut baser son style de rédaction sur les brevets précédemment déposés par l'entreprise afin d'adapter son style de rédaction. Le copilote sera par ailleurs en mesure d'analyser l'état de l'art et de suggérer des amendements afin de surmonter les rejets potentiels lors du processus d'analyse du dépôt de brevet par les autorités. L'ambition est de créer une solution tout-en-un pour les professionnels de la rédaction de brevets.
- Patlytics(34) : Développant un outil déjà utilisé par de très grandes entreprises techs ou de grands cabinets d'IP, la start-up permet de rédiger automatiquement des drafts de brevets personnalisés en générant des descriptions détaillées des innovations. La société propose également une fonctionnalité de cartographie « brevet-produit » permettant de détecter de manière pro-active et systématique les opportunités de licence. L'outil aide par ailleurs à gérer stratégiquement les portefeuilles de brevets en identifiant ceux d'entre eux qui sont les plus précieux afin de prendre les décisions les plus éclairées possible.

- IPRally(35) : Une start-up finlandaise fondée en 2018 spécialisée dans la recherche de brevets assistée par IA. La start-up a notamment développé un assistant IA permettant aux utilisateurs de poser des questions en langage naturel sur un brevet ou un ensemble de brevets, obtenant ainsi des analyses techniques détaillées produites automatiquement. La start-up permet d'évaluer la valeur de portefeuilles de brevets existants, d'identifier les lacunes technologiques d'orienter des stratégies R&D.
- Patsnap : (36) A l'instar de certaines autres sociétés mentionnées, Patsnap a développé un assistant conversationnel permettant aux utilisateurs de rechercher rapidement des brevets et de poser des questions spécifiques sur les technologies et les entreprises. Le point de différenciation par rapport à d'autres assistants conversationnels spécialisés réside dans le fait d'intégrer des documents hors IP, comme des articles scientifiques par exemple. L'outil permet d'analyser des séquences biologiques et des structures chimiques, facilitant la recherche d'antériorité et l'évaluation de la brevetabilité dans le domaine des sciences de la vie. Patsnap a également développé PatSnap Discovery, un outil d'analyse de marché et de veille technologique, permettant d'explorer les tendances émergentes, d'analyser les stratégies des concurrents et d'identifier des opportunités de partenariat ou d'investissement.

Pour l'heure, aucun de ces outils ne cible spécifiquement les VCs – ils sont en effet davantage destinés aux grands groupes et aux cabinets spécialisés – mais chacun d'entre eux peut être utilisé par des investisseurs dans le cadre de ses travaux et analyses. Plus intéressant encore, ces outils peuvent servir, à différents degrés à inspirer les fonds de VC dans le développement de leurs propres outils sur mesure.

Ainsi, un outil d'analyse IP dédié aux VCs pourrait parfaitement intégrer les fonctionnalités suivantes :

- Participation au screening des sociétés : L'évaluation des sociétés du deal flow pourrait intégrer un score IP composé de sous-scores portant sur la force des brevets, les risques de contre façons et les tendances dans le domaine technologique. En plus du score, généré, ces outils pourraient permettre de générer des résumés des brevets de manière à gagner du temps sur l'analyse.
- En phase de due-diligence, la réalisation d'une étude de liberté d'exploitation automatisée pourrait permettre aux fonds d'économiser du temps et de l'argent en évitant de devoir passer par des cabinets spécialisés.
- Identification de partenaires de licence potentiels et des meilleures stratégies de développement, lorsque les sociétés sont en portefeuilles. Les outils d'IA pourront également accompagner les sociétés du portefeuille dans leurs stratégies de life-cycle management de la propriété intellectuelle.
- Enfin, de manière plus prospective, un outil d'IA pourrait également être utilisé pour prédire la valorisation des brevets

Les bénéfices de l'utilisation d'un tel outil seraient pour les fonds de Venture Capital assez évidents :

- Automatisation des tâches chronophages et réduction des délais de due-diligence. Le temps est souvent un élément critique dans le monde du venture capital.
- Meilleure identification des opportunités et des risques, en permettant une cartographie fine des technologies et des « white spaces », qui sont des lacunes dans le paysage des

brevets où il n'y pas ou peu de dépôts de brevets. Ces territoires inexplorés peuvent conduire au développement de nouvelles inventions.

- Gain en objectivité des analyses
- Renforcement de l'accompagnement stratégique du fonds de Venture Capital auprès de ses participations (via la mise à disposition de l'outil ?)

A terme, en combinant la meilleure sélection des deals et en permettant aux fonds d'apporter un accompagnement stratégique plus objectif à leurs participations, il est envisageable qu'un outil de ce type permette d'améliorer la performance du portefeuille.

Les limites de l'utilisation de cet outil :

D'un point de vue pratique, le développement et l'utilisation d'un tel outil nécessite un investissement initial et des coûts de maintenance qui peuvent apparaître comme relativement élevés, en termes de logiciel mais également d'infrastructure.

La mise en œuvre d'un outil d'IA dans le domaine de la propriété intellectuelle rencontre également la barrière de l'expertise, les fonds ne disposent généralement pas en interne d'expert en data science ni en propriété intellectuelle.

Quelques défis plus techniques peuvent également exister, concernant notamment la capacité à recueillir des données d'entraînement fiables. Les modèles d'IA sont en effet entraînés sur des données historiques. Si ces données sont biaisées ou incomplètes, les prédictions et analyses générées peuvent l'être également. Par exemple, une surreprésentation de certaines technologies ou régions peut conduire à des recommandations inappropriées.

De la même manière, des données trop anciennes dans un secteur évoluant rapidement comme celui des biotechnologies pourraient ne pas refléter certaines avancées, rendant certaines analyses obsolètes ou inexactes. Concernant l'âge des données utilisées, il existe par ailleurs une limitation propre aux données de propriété intellectuelle, qui ne sont rendues publiquement accessibles en France ou en Europe que 18 mois après la date de dépôt ou la date de priorité la plus ancienne.

Un autre enjeu résidera dans l'infrastructure à mettre en place pour valoriser les données propriétaires des participations historiques du fonds d'investissement.

Enfin, il existe également une limite de nature juridique : quid de la responsabilité si l'outil venait à délivrer des informations erronées ? En cas d'erreur — par exemple si l'IA omet un brevet bloquant en vigueur dans le cadre d'une FTO — les conséquences financières peuvent être majeures : contentieux pour contrefaçon, suspension du développement d'un produit, voire invalidation de la stratégie de sortie. Dans un tel cas, la question de la responsabilité se pose à plusieurs niveaux : celle de l'équipe d'investissement, qui a intégré l'outil sans validation juridique indépendante ; celle du fournisseur de l'IA, dont les conditions générales limitent souvent drastiquement la garantie sur les résultats ; et enfin celle de la startup elle-même, si elle a fourni des données incomplètes ou trompeuses. Par conséquent, il est à prévoir que l'IA sera positionnée au sein des fonds comme un outil de pre-screening avant le recueillir la validation d'un professionnel extérieur.

En conclusion, nous pouvons affirmer que les technologies d'IA commencent à atteindre un niveau de maturité suffisant pour être utilisées dans le domaine de la propriété intellectuelle. Les bénéfices qu'elles peuvent apporter aux fonds d'investissement sont relativement évidents mais aucun outil existant n'est adapté à ce stade aux VC. Par conséquent, pour cette raison et pour

préserver la confidentialité des données sensibles, le développement d'outils propriétaires est une voie devant être explorée par les fonds d'investissement. Les limites inhérentes aux fonds et mentionnés plus haut nous poussent à penser que le développement de ces outils seront limités aux fonds les plus volumineux.

G] IA & Données pré-cliniques

I] Introduction

L'analyse des données pré-cliniques générées par les biotechs constitue un élément critique de la due-diligence menée par les investisseurs en capital-risque. Elle permet en effet, pour des biotechs early-stage, d'évaluer la robustesse scientifique et la translationnalité d'un candidat médicament, avant toute décision d'investissement dans le développement du projet.

La rigueur des études menées constitue aussi un gage de crédibilité pour les fondateurs.

II] Les données disponibles

Il existe plusieurs bases de données qui centralisent (et structurent !) des données pré-cliniques, pour autant, aucune ne présente une exhaustivité suffisante à ce jour.

Ces bases de données (ChEMBL, eTOX, ToxCast, etc) constitueront un excellent point de départ, pour autant, elles seront rarement suffisantes. Il sera par conséquent généralement nécessaire d'utiliser directement les publications scientifiques et de mettre en place un processus d'extraction automatisée.

Une des limites à l'accès des données sera constituée des différents barrières au téléchargement mises en place par les éditeurs des revues scientifiques.

III] Les critères d'évaluation de la pertinence des études menées et des données générées

A) Evaluation de la pertinence de la cible thérapeutique(37)

Lorsqu'un candidat médicament va s'intéresser à une nouvelle cible thérapeutique, il va être nécessaire d'évaluer la pertinence du choix de cette cible et d'évaluer le rôle de cette dernière dans la physiopathologie de la maladie. Il existe différentes catégories de cibles : d'un côté des cibles nouvelles et ne disposant que de quelques données bibliographiques, et de l'autre, des cibles déjà éprouvées et validées par un produit approuvé ou en phase de développement clinique avancé.

Les méthodologies diffèrent d'un fonds à l'autre et d'un investisseur à l'autre, mais l'une des voies d'analyse peut être de s'intéresser à la génétique. Existe-t-il des maladies génétiques associées à la cible ? Quelle est la méthode de transmission génétique de cette pathologie ? Il est aussi possible de s'intéresser aux cibles en amont ou en aval de la cible d'intérêt dans la voie biologique. Il sera également possible de s'intéresser aux autres cibles explorées dans la littérature dans le cadre de la pathologie donnée.

Il faudra également évaluer la « druggabilité » de la cible, dans la continuité de l'analyse génétique, évaluer l'impact fonctionnel connu de la cible. Est-ce que, dans ce contexte, il faudra induire un gain de fonction ou une perte de fonction (activer ou inhiber la cible) ? En fonction de la localisation de la cible, quelle sera la modalité d'intervention à privilégier (petite molécule, anticorps, édition génique, etc) ? Faudra-t-il intervenir de manière localisée ou systémique ?

B) Evaluation de la pertinence de la molécule

Pour garantir l'effet souhaité, il faudra vérifier que la pharmacocinétique de la molécule sélectionnée lui permet d'atteindre la cible avec une exposition suffisante, sur une période définie. Il faudra interagir avec la cible de manière adéquate, en déclenchant une réaction au sein de la voie biologique sélectionnée. Enfin, il faudra valider les faisceaux de preuves permettant de penser que la modification de la voie biologique permettra, in fine, de modifier la trajectoire clinique du patient.

Dans ce travail de due-diligence pharmacologique, l'existence et la mise en exergue de biomarqueurs sera fondamentale. Il y aura, à ce titre trois types de biomarqueurs pouvant être exploités : les biomarqueurs proximaux, permettant de mesurer l'engagement direct de la cible, les biomarqueurs distaux, permettant de mesurer les effets en aval de la cible : modification de la cascade de signalisation, altération de métabolites circulants, etc. Enfin, les biomarqueurs validés cliniquement qui seront quant à eux corrélés à la pathologie chez l'homme.

La mesure de la cohérence scientifique, entre le mécanisme d'action revendiqué et les preuves expérimentales passera notamment par le degré de validation externe (peer-review, réplification indépendante des résultats).

C) Evaluation de la pertinence translationnelle

La pertinence translationnelle désigne la capacité des résultats précliniques (in vitro, in vivo ou in silico) à prédire de manière fiable l'efficacité, la sécurité et le comportement d'un traitement chez l'être humain. Il s'agit du degré de confiance que l'on peut accorder au fait que les observations faites dans les modèles animaux se traduiront par des résultats cliniques pertinents.

L'évaluation de la pertinence translationnelle repose notamment sur :

- **L'évaluation du choix des modèles in vitro :** est-ce que les modèles cellulaires sont pertinents ? Est-ce que des techniques innovantes ont été utilisées (organoïdes, organs-on-chip, etc.) ?
- **L'évaluation du choix des modèles animaux :** est-ce que les espèces sont adaptées ? Combien de modèles ont-ils été utilisés ? Est-ce que les souches utilisées sont pertinentes d'un point de vue de la biologie du patient ? (il existe par exemple des modèles animaux humanisés génétiquement modifiés pour exprimer des récepteurs humains)
- **L'évaluation du choix des endpoints des études précliniques :** peuvent-ils trouver un équivalent chez l'homme ? (Exemple : mesure de la taille des tumeurs chez les animaux vs critères RECIST chez le patient atteint d'un cancer).
- **L'évaluation de la qualité de la modélisation PK/PD :** quelle est la qualité des données d'entrée du modèle ? Le nombre d'animaux intégrés dans chacun des groupes est-il suffisant ? Est-ce que les mesures plasmatiques ont été réalisées à des intervalles pertinents ? Est-ce que le modèle choisi est pertinent (nombre de compartiments, hypothèses, etc) ?

La méthode utilisée par l'investisseur en capital, qui n'est évidemment pas un spécialiste de toutes les pathologies, pour évaluer la pertinence des choix de développement pré-clinique sera de s'appuyer sur la littérature : est-ce que les modèles utilisés sont les plus courants ? Si ce n'est

pas le cas, en quoi sont-ils meilleurs ? Une expertise externe sera quasiment systématiquement nécessaire pour analyser ces points avec précision.

D) Robustesse statistique

L'analyse de la robustesse statistique des études pré-cliniques permet de déterminer que les résultats observés ne sont pas le fruit du hasard, qu'ils sont reproductibles et que les conclusions que l'on peut en tirer sont fiables.

Les points à vérifier sont les suivants :

- La taille des échantillons et puissance statistique
- La randomisation et l'insu afin d'éviter respectivement les biais d'allocation et les biais de mesure
- La justification des tests statistiques
- La reproductibilité interne et externe : est-ce que les tests ont été répliqués dans des conditions identiques avec des opérateurs différents ? Est-ce que les tests ont été répliqués par des laboratoires indépendants ?

Il est toutefois nécessaire de noter que la plupart des investisseurs en capital se reposent sur les peer-reviewers des revues à comité de lecture en ce qui concerne ces considérations statistiques, l'intégration de ces sujets à la due-diligence étant relativement rare.

E) Adhérence aux guidelines

Le suivi des guidelines est indispensables à la bonne réception des données pré-cliniques générées par les agences réglementaires comme la FDA et l'EMA (par exemple dans le cadre d'un dépôt d'« Investigational New Drug », une demande d'autorisation soumise à la FDA par le sponsor afin de pouvoir lancer des études cliniques).

Les guidelines à suivre sont bien connues, et concernent essentiellement la sécurité des candidats médicaments :

- ICH S6(R1) – Safety Studies for Biotechnological Products
- ICH M3(R2) – Non clinical Safety Studies for the Conduct of Human
- ICH S7A/B – Safety Pharmacology
- OCDE Guidelines

Outre le fait de permettre de réduire le risque de rejet des résultats par les autorités, l'adhérence aux bonnes pratiques démontrera la capacité de l'équipe à faire face aux enjeux réglementaires.

IV] Les apports possibles de l'IA

L'IA peut être utilisée de différente manière en ce qui concerne les données pré-cliniques des projets à analyser :

- **Pour l'extraction et la structuration des données**

A partir des articles publiés par la cible ou des documents fournis, il est possible de mettre en place un outil permettant de « lire » les PDF, de faire de la reconnaissance de caractères pour les textes fournis au format image et de faire de la reconnaissance d'entités (gènes, cibles, lignées cellulaires, modèles animaux, doses, etc).

- **Pour la réalisation des revues de littérature**

Une revue de la littérature préclinique est essentielle pour situer les nouvelles données dans le contexte des travaux existants et pour identifier de potentielles lacunes méthodologiques. Les revues permettent d'agréger les résultats de multiples études pour évaluer la cohérence et la puissance des nouvelles preuves. Dans ce cadre, l'IA pourra permettre de définir un protocole de recherche (critères d'inclusion et d'exclusion des études, bases interrogées, mots-clés utilisés, etc), et pourra évaluer la qualité méthodologique de chaque étude, ce dernier point faisant parfois défaut aux revues de littérature menées par des opérateurs humains.

- **Pour l'évaluation critique des données et la mise en place d'un scoring**

Une fois les données générées par l'entreprise cible extraites, une fois la revue de la littérature menée, des modèles d'IA permettent d'évaluer de manière critique les données précliniques à travers des grilles de lecture intégrant les standards expérimentaux (la randomisation, la taille de l'échantillon, la reproductibilité, etc) et peuvent générer un scoring de robustesse, de pertinence translationnelle ou de différenciation, nécessaires à la priorisation des opportunités.

V] L'émergence des agents IA scientifiques

Récemment, nous avons assisté à l'émergence d'agents IA scientifiques. Il est important de s'intéresser à ces systèmes tant ils sont porteurs de potentiel pour les VC et tant ils pourraient être facilement adaptables à un contexte de due-diligence et d'investissement.

Un agent IA(38) est un programme logiciel autonome capable de percevoir son environnement, de formuler des plans et de prendre des actions pour atteindre des objectifs définis. Contrairement à un simple modèle de génération de texte (GenAI), l'agent IA dispose souvent de mémoire (pour conserver le contexte), de droits d'accès à des outils externes (API, bases de données, etc) et de mécanismes d'auto-vérification.

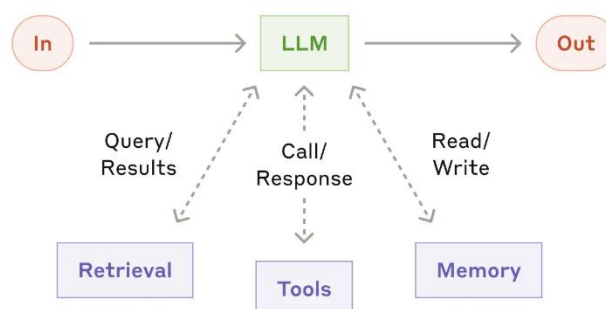


Figure 43 - Schéma de fonctionnement d'un agent IA

- **Google AI Co-Scientist(39)**

L'AI Co-Scientist a été lancé par Google en début d'année 2025. Il s'agit d'une plateforme d'agents IA spécialisés collaborant pour générer, critiquer et affiner des hypothèses de recherche et des protocoles expérimentaux, principalement dans le domaine de la recherche biomédicale.

L'AI Co-Scientist repose sur une coalition d'agents aux rôles définis :

- Génération de nouvelles hypothèses de recherche à partir d'un objectif formulé en langage naturel
- Réflexion sur la solidité des hypothèses vis-à-vis de la littérature existante
- Priorisation des idées et des hypothèses en ce qui concerne la faisabilité et l'impact
- Les agents « évolution » et « proximity » vont enrichir les propositions sélectionnées en itérant sur les retours précédents
- Enfin, l'agent « meta-review » compile dans un rapport final les hypothèses, les analyses bibliographiques et les protocoles expérimentaux suggérés

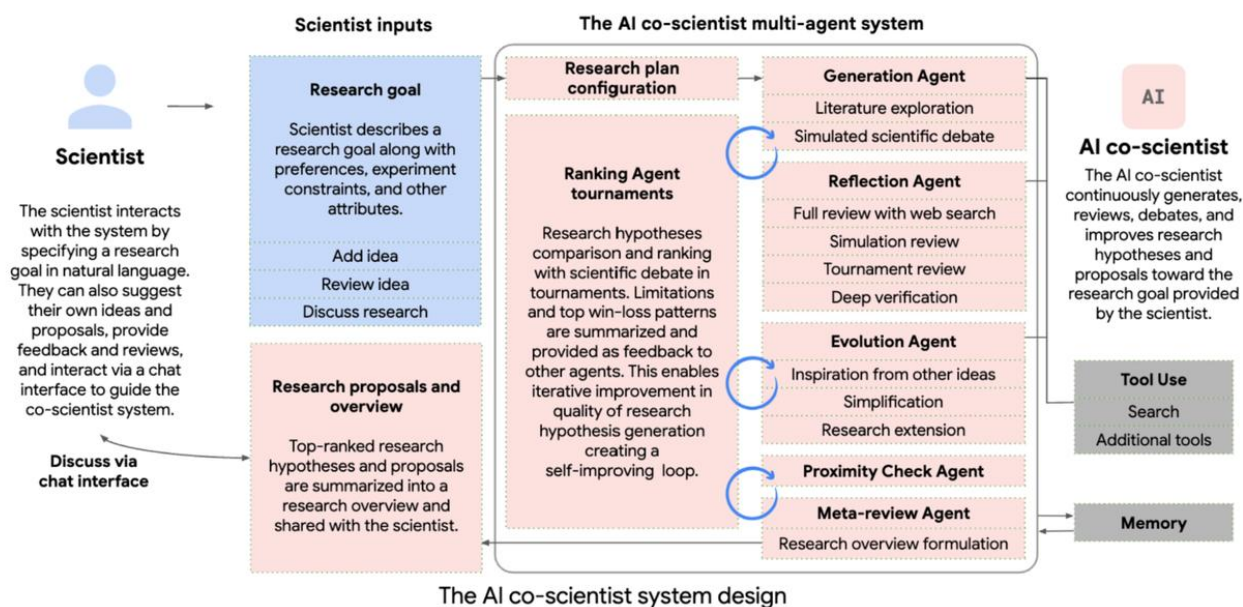


Figure 44 - Schéma de fonctionnement du système multi-agents de Google (issu de la publication pré-citée)

Les premiers cas d'usage de ce système multi-agents concernent notamment la découverte de nouvelles cibles thérapeutiques, le repositionnement médicamenteux, la planification d'essais cliniques (permettant de réduire de 30% la durée de conception) et la synthèse documentaire.

- **FutureHouse Platform(40)**

FutureHouse est une initiative à but non lucratif visant à automatiser la recherche scientifique, notamment dans les domaines de la biologie et des sciences complexes, en développant des agents IA capables de générer, critiquer et affiner des hypothèses de recherche, ainsi que de proposer des protocoles expérimentaux détaillés.

La plateforme comprend quatre agents principaux, chacun ayant une spécialisation distincte :

- Crow est un agent polyvalent pour la recherche documentaire, capable de répondre à des questions scientifiques en s'appuyant sur des sources fiables et académiques
- Falcon est spécialisé dans les revues de littérature approfondies, avec des accès à des bases de données scientifiques (articles en libre accès)
- Owl est conçu dans l'objectif de vérifier si une recherche spécifique a déjà été réalisée, si une hypothèse a déjà été explorée, afin d'éviter des duplications redondantes et de gagner du temps

- Phoenix est axé sur l'élaboration de protocoles expérimentaux

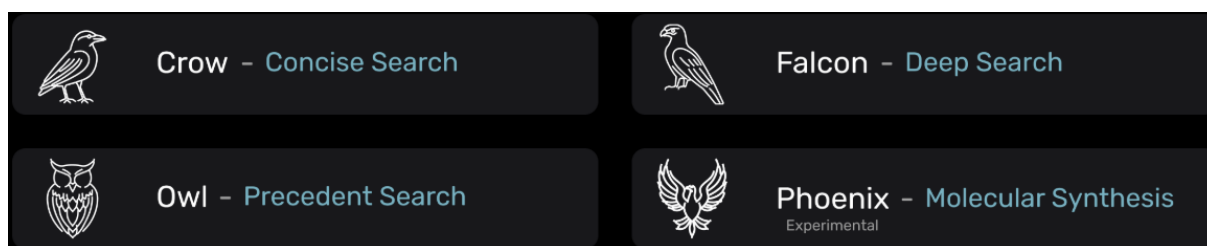


Figure 45 - Les 4 agents autonomes de FutureHouse

FutureHouse a récemment communiqué sur la découverte d'un nouveau médicament.(41) En effet, la plateforme Robin, développée par FutureHouse a permis d'identifier un candidat médicament pour la DMLA de forme sèche. Il s'agit d'un inhibiteur ROCK, le ripasudil, utilisé pour traiter le glaucome mais jamais investigué dans cette nouvelle indication. L'ensemble des étapes de génération d'hypothèses, de planification expérimentale et d'interprétation des données ont été menées par Robin, tandis que les humains ont simplement exécuté les expérimentations.

- **Autres systèmes multi-agents**

Les systèmes ayant pour objectif d'industrialiser la recherche scientifique ont le vent en poupe et plusieurs équipes de chercheurs ont proposé leur propose système. C'est le cas notamment de Ghafarollahi & Buehler(42) qui ont proposé un système de SciAgents permettant l'auto-génération et le raffinement d'hypothèses en science des matériaux ou encore de Su et al.(43), ayant développé les Virtual Scientists, des agents autonomes conçus pour fonctionner de la même manière qu'une équipe de chercheurs et qui ont démontré leur supériorité aux chercheurs humains dans leur capacité à générer de nouvelles idées de recherche.

Comme discuté plus haut, ces systèmes – très récents et en évolution constante – pourraient constituer d'excellentes bases à un système développé par un fonds d'investissement biotech.

H] Utilisation de l'IA pour prédire le succès des études cliniques

Comme déjà vu précédemment, le travail de due-diligence scientifique et technique d'un investisseur en capital-risque consiste à évaluer la pertinence des données générées mais aussi à essayer de se projeter dans le futur de la société pour en évaluer les chances de succès en matière scientifique et technique, mais aussi en matière financière (est-ce que l'entreprise sera suffisamment armée pour se financer jusqu'à un exit ?) ou commercial (est-ce que le produit proposé par l'entreprise rencontrera son marché ?).

Une part importante de ce travail, concernant des start-up développant des thérapeutiques, consiste à essayer de prédire, sur la bases des données déjà générées, les chances de succès des phases cliniques suivantes.

Différents modèles d'IA permettant de prédire le succès d'une étude ont d'ores et déjà été utilisés en recherche académique ou pour la R&D de grandes entreprises pharmaceutiques. En effet, cette technologie est utilisée pour réallouer le capital R&D en fonction des chances de succès.

En revanche, leur utilisation dans la finance d'investissement est plutôt novatrice, cette approche ayant été mise en avant à la fin des années 2010 par le professeur Andrew Lo du MIT. Ce dernier propose d'intégrer le score probabiliste de succès (POS) aux modèles de projection financière de l'entreprise mais également du portefeuille des fonds.

La prédiction du succès des études cliniques peut avoir lieu à deux moments : après l'obtention des résultats de Phase II ou après les résultats de Phase III.(44) Dans l'état actuel des choses, ces prédictions intéressent donc les fonds late-stage mais il est également possible, nous le verrons, d'utiliser des données plus précoces – rendant la technologie utilisable par des fonds early-stage.

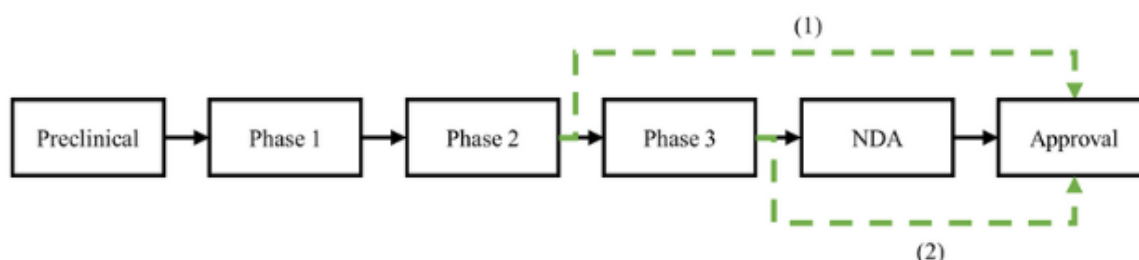


Figure 46 - Stades d'utilisation d'un outil de prédiction des études cliniques

I] Les données utilisables : quelles données pour nourrir les algorithmes ?

Les modèles prédictifs dans le domaine des études cliniques reposent sur un maillage de sources hétérogènes.

- ClinicalTrials.gov et EudraCT sont des registres publics permettant d'obtenir les protocoles, les critères de jugement, les vitesses de recrutement et les résultats finaux des études.
- Citeline TrialTrove ou Clarivate Cortellis sont des bases de données privées et payantes permettent d'agréger les données publiques à d'autres informations plus difficilement accessibles, comme les informations réglementaires.

- DrugBank et PubChem vont permettre d'intégrer au modèle des données sur les structures et caractéristiques physico-chimiques des molécules. En couplant ces données à celles de bases comme l'Expression Atlas ou GTEx, il est possible d'alimenter des modèles de toxicité.

II] Evolution des modèles d'IA et performances publiées

Un grand nombre de modèles ont été publiés jusqu'à présent concernant la prédiction des résultats des études cliniques. Chacun d'entre eux peut inspirer un fonds de Venture Capital souhaitant implémenter son propre système de prédiction. En voici les plus marquants :

- Le projet ALPHA du MIT de 2019(45)

Mené par le professeur Andrew Lo, ce projet a développé un modèle combinant du machine learning avec des méthodes d'imputation statistiques pour gérer les données manquantes et améliorer les prédictions. En se basant sur une collaboration avec Informa Pharma Intelligence, le projet a développé une base de données intégrant jusqu'à 140 variables par étude clinique. Le modèle a été validé sur un corpus de données historiques et est mis à jour trimestriellement.

Il est mise à disposition gratuitement en open source.

- Le projet HINT (Hierarchical Interaction Network) de 2022 (46)

Modèle de deep learning se basant sur un réseau d'interaction hiérarchique ayant pour objectif de modéliser les interactions complexes entre le médicament, la maladie et le design de l'essai. Les données utilisées par ce modèle proviennent d'un peu plus de 17 000 études cliniques et sont multimodales : les caractéristiques du médicament (notamment sa structure moléculaire), les informations sur la maladie ciblée (Code CIM-10, description physiopathologique, etc), les critères d'inclusion et autres données sur le design de l'étude.

Les performances du modèle HINT ont surpassées celles des modèles basés sur le machine learning et sur les inférences statistiques afin d'atteindre un score PR-AUC d'environ 0,90 en phase III.

- Le projet ChemAP datant de 2024(47)

Ce modèle de deep learning se différencie par sa capacité à prédire l'issue des études cliniques (et la capacité d'un candidat médicament à obtenir l'approbation réglementaire) avant même la génération des premières données cliniques, en ne se basant par conséquent que sur des données physico-chimiques relatives à la structure moléculaire. La performance du modèle sur un jeu de données de validation atteint une AUROC d'environ 0,78. Sur des données réelles, sa performance est d'environ 0,69.

Un certain nombre d'acteurs industriels ont également développé des modèles visant à prédire le succès des études cliniques en utilisant différentes approches, parmi ceux-ci, nous pouvons mentionner : Intelligencia(48), Unlearn.AI (49), qui utilise des jumeaux numériques, ou encore, plus proche de nous, en France, nous pouvons également citer la start-up Botdesign qui utilise une nouvelle technologie algorithmique.

III] Quelle (future) place des agents IA dans la prédiction des résultats cliniques ?

Dans le domaine de la prédiction des résultats cliniques, l'agentisation – le recours à des agents IA capables d'observer, de raisonner et d'interagir avec des outils externes, généralement organisés en écosystème, chaque agent étant optimisé pour une fonction précise – ouvre plusieurs horizons : l'industrialisation du pipeline de données nécessaires à l'alimentation des modèles, l'adaptation de ce pipeline de données en temps réel et l'intégration des données prédictives directement dans les modèles financiers.

Ainsi, en passant d'une architecture mono-modèle à un écosystème d'agents, nous pourrions améliorer les performances des modèles de prédiction et rendre leurs conclusions plus pertinentes et utilisables.

Organisation possible d'un écosystème d'agents :

- **Ingestion et curation des données :** l'objectif de l'agent sera ici d'aspirer ClinicalTrials.gov et EudraCT, de repérer les doublons et de remplir les valeurs manquantes par des données synthétiques afin de constituer une base de données structurée et utilisable. Ainsi, chaque jour, nous pourrions travailler avec un data set le plus à jour possible.
- **Analyse sémantique :** l'agent dédié à l'analyse sémantique permettra d'extraire les endpoints, les bras de traitement, les critères d'éligibilité via du NLP et de reclassifier les indications avec un thesaurus MeSH (il y a parfois un problème dans la constitution des bases de données car une même pathologie ou une même intervention peut être nommée avec plusieurs noms différents)
- **Enrichissement de la base de données :** Cet agent a pour objectif d'enrichir la base de donnée en la connectant aux API de bases de données existantes (DrugBank, etc...)
- **Modélisation auto-ML :** Cet agent aura pour mission de tester différents modèles de Machine Learning sur le dataset (Random forest, GBoost, Graph Neural network, LLM, etc) et de comparer leur performance (AUC – l'aire sous la courbe ROC, représentant la probabilité qu'une étude clinique positive reçoive un score supérieur à une étude clinique négative) & PR-AUC – correspondant à la capacité du modèle à détecter les positifs sans générer trop de faux positifs, d'autant plus utile dans une population dans laquelle nous avons beaucoup d'études négatives). A titre d'indication, voici généralement les ordres de grandeurs typiques que nous retrouvons dans les publications portant sur la prédiction du succès clinique.

Phase évaluée	AUC-ROC couramment rapportée	PR-AUC couramment rapportée
Phase I → II	0,70–0,80	0,55–0,65
Phase II → III	0,75–0,85	0,60–0,75
Phase III → Approval	0,80–0,90	0,70–0,85

Figure 47 - Performance du modèle de prédiction du succès des études cliniques

- **Génération de la POS :** Cet agent calculera la probabilité de succès de l'étude d'intérêt, d'un score de safety et d'un score d'efficacité de la molécule, en fonction notamment de

ses caractéristiques physico-chimiques. Il sera en mesure d'évaluer l'incertitude et de constituer différents scénarios et études de sensibilité.

- **Surveillance en temps réel** : Cet agent guettera les communiqués d'arrêt d'étude, de mise à jour des protocoles afin de pouvoir mettre à jour la base de données en temps réel et ajuster la POS. Il fonctionnera en étant connecté aux flux RSS de la FDA, de l'EMA mais aussi à l'API de ClinicalTrials.gov par exemple.
- **Connexion à la prise de décision financière** : Cet agent rendra utilisable les probabilités de succès générées en les intégrant à différentes métriques financières couramment utilisées dans des fonds d'investissement :

Le rNPV – la valeur actuelle nette ajustée du risque. Il s'agit d'une variante de la NPV classique qui permet d'intégrer la probabilité d'aboutissement d'un projet incertain, phase par phase.

$$rNPV = \sum_{t=0}^T \frac{(CF_t \times p_t)}{(1+r)^t} - I_0$$

- CF_t : Flux de trésorerie attendu à l'année t.
- P_t : probabilité cumulée qu'à l'année t le projet ait franchi avec succès toutes les étapes précédentes (POS)
- r : taux d'actualisation
- I₀ : investissement initial ou cumulé jusqu'au temps t

De manière générale, des POS typiques sont utilisées en fonction de la phase de développement de la thérapeutique. Mais ces probabilités ne tiennent pas compte des particularités du produit. Par conséquent, elles dégradent la pertinence et la précision de l'indicateur final.

Le TRI – ou taux de rentabilité interne (équivalent à l'Internal Rate Return, en anglais) est le rendement annuel composé qu'offrent les flux futurs, compte tenu du calendrier et des montants investis. Il doit être calculé pour chaque investissement du fonds d'investissement et dans divers scénarios. Mathématiquement, il représente la valeur actuelle nette d'un projet, c'est-à-dire le taux auquel la somme actualisée des flux de trésorerie futurs est exactement égale à l'investissement initial.

Au total, les apports concrets pour la prédiction des résultats cliniques d'un écosystème d'agents autonomes sont nombreux :

- **La robustesse** : en cloisonnant les tâches, chaque agent exploite des jeux de données spécifiques et limite l'effet « boîte noire » et favorise l'explicabilité des résultats.
- **L'adaptativité** : en effet, les POS sont dynamiques et mises à jour au fur et à mesure de l'avancée des essais et de la publication des résultats.
- **La scalabilité** : un grand nombre de programmes peut être monitoré et analysé de manière simultanée

Plusieurs équipes de recherche ont travaillé sur la pertinence d'utiliser des écosystèmes d'agents autonomes pour prédire le succès d'études cliniques, la publication « CT-Agent : Clinical trial Multi-Agent with LLM-based reasoning »(50) est un exemple d'utilisation d'agents à cette fin.

Les auteurs sont en effet partis du constat que malgré leurs performances, les LLM peinent à exploiter les bases de données médicales pour prédire le succès d'une étude clinique. Ils ont

donc proposé une architecture combinant GPT-4 avec 5 agents spécialisés, apparaissant dans le schéma ci-dessous :

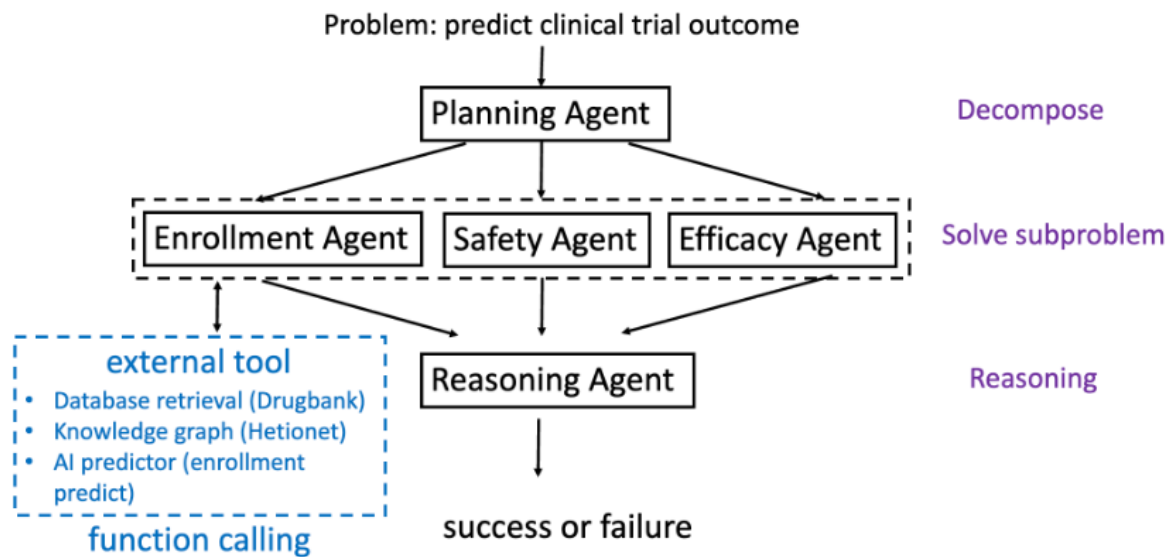


Figure 48 - Schéma de fonctionnement du système multi-agents de prédiction du succès de l'étude clinique(50)

- **Un Planning Agent** dont le rôle est de décomposer le problème général en sous-problèmes
- **Un Efficacy Agent** qui va permettre d'évaluer l'efficacité du médicament
- **Un Safety Agent** qui estimera les risques d'effets indésirables
- **Un Enrollment Agent** qui aura pour objectif de prédire la faisabilité du recrutement des patients
- Et enfin, **un Reasoning Agent**, dont la mission sera d'agrégier les résultats et de les pondérer afin

Il s'agissait du premier système multiagents réellement orienté vers la prédiction du succès d'une étude clinique.

IV] Bénéfices et limites

Nous l'avons vu, l'application des technologies d'intelligence artificielle au domaine de l'analyse des données pré-cliniques et des données cliniques pour les VCs a le potentiel de transformer les pratiques de due diligence des fonds. Cette évolution, probablement inéluctable, offre des avantages significatifs tout en présentant certaines limites.

En matière de bénéfices, il est bien sûr nécessaire d'évoquer :

- **Le volume des données analysées :**

L'IA peut ingérer des milliers de rapports pré-cliniques et cliniques en quelques heures. Cela prendrait des jours ou des semaines à un analyste humain. Dans les faits, et pour des raisons de contraintes de temps, les analyses réalisées par les fonds restent généralement en surface. L'IA accélère par conséquent le « time to insight » des équipes de VC.

- **L'aide à la décision d'entrée en due-diligence (en positionnant au plus tôt l'analyse technique)**

Dans un fonds d'investissement traditionnel, la sélection des projets à diligenter doit se faire en un temps limité, souvent sans maîtrise complète des aspects scientifiques. En accélérant l'évaluation technique et en la positionnant dès les premières étapes, l'équipe peut lancer les analyses plus tôt, examiner un plus grand nombre de dossiers et éviter de mobiliser inutilement ses ressources sur des opportunités non viables. Finalement, cette approche améliore la performance globale du fonds en concentrant les efforts sur les dossiers à fort potentiel et en optimisant le temps de l'équipe d'investissement.

- **La réduction des biais cognitifs et la précision des analyses statistiques**

En permettant d'appliquer systématiquement les mêmes critères à tous les dossiers, l'IA limite les effets d'ancrage et de halo propres aux évaluations humaines. Cette uniformisation des critères d'analyse permet aussi de comparer plus facilement et automatiquement différents projets du dealflow, ou même de comparer aux concurrents des entreprises analysées si ces derniers ont des données publiées.

- **Optimiser un portefeuille en temps réel**

Un système intégrant l'IA étant en mesure de calculer en temps réel la probabilité de succès de chaque actif et de chaque entreprise, il permet d'optimiser un portefeuille en permanence.

- **Des limites généralement connues des modèles d'IA**

De manière générale, les modèles permettant de déterminer la qualité des données précliniques et les probabilités de progression clinique d'une molécule sont touchées par les mêmes limites que tous les autres modèles d'IA :

- Vigilance quant à la qualité des données d'entrée
- Interprétabilité du modèle : il faut éviter de désigner des modèles d'IA « boîte noire »
- Éviter la surconfiance (des erreurs sont possibles). Il faudra former les utilisateurs aux forces et aux faiblesses des modèles d'IA.
- Nécessité de maintenance et de validation régulière. Il faudra, en ce qui concerne la prédiction notamment, mettre en place un niveau de validation du modèle important en utilisant des dataset « gold standard » et réaliser des benchmarks périodiques pour mesurer la performance et recalibrer le cas échéant les différents modèles.

I] Conclusion et Perspectives

I] Le no-code comme moteur d'accélération de l'intégration de l'IA au sein des fonds d'investissement

Le no-code désigne l'ensemble des plateformes visuelles qui permettent de concevoir des applications, des workflows ou des intégrations sans écrire de code traditionnel : chaque fonctionnalité est encapsulée dans un bloc que l'on relie aux autres blocs par « drag-and-drop » - technique consistant à déplacer un élément dans une application ou d'une application à l'autre dans une interface graphique – concrètement, cela revient à transformer la programmation en assemblage de briques, ouvrant la création d'outils IA aux professionnels qui ne sont pas développeurs.

Les outils utilisés peuvent être divisés en quatre catégories :

- **Outils d'orchestration et d'automatisation** : ces outils sont le cœur de l'architecture : ils déclenchent les workflows, routent les données et gèrent les bugs et erreurs. Ces plateformes vont être équipées de centaines de connecteurs permettant d'activer des applications variées et offrir une vue graphique des scénarios mis en œuvre.
Exemples : Make, Zapier, N8n, Pipedream
- **Outils de base de données** : ces outils permettent de stocker les données et de les rendre disponibles.
Exemples : Airtable, Notion DB
- **Front-end & Dashboards** : ces outils ont pour objectif de transformer les données brutes en interfaces actionnables par les utilisateurs. Ils vont permettre de créer des applications internes, à l'aide de composants drag-and-drop.
Exemples : Retool, Glide, Softr, Appsmith
- **Outils AI plug-ins** : ces outils servent à manipuler et enrichir les données. Ils vont par exemple permettre d'extraire des données d'un CSV et de les analyser avec un LLM.
Exemples : Parabola, Bubble avec plug-ins Open AI

Les outils no-code présentent différents avantages :

- Rapidité de prototypage
- Interopérabilité native entre les différents outils, grâce au large choix de connecteurs
- Modalités de gouvernance intégrée (afin notamment d'être en cohérence avec la réglementation RGPD)

Le no-code peut donc s'imposer comme un formidable catalyseur dans l'intégration rapide de l'IA au sein des fonds d'investissement, en abaissant radicalement la barrière technique qui séparait jusqu'ici les équipes en charge du deal-flow et de l'investissement des innovations algorithmiques.

En effet, en s'affranchissant de la nécessité d'écrire du code, les analystes peuvent désormais prototyper des workflows alimentés par des LLM, des extracteurs de données ou des moteurs de scoring en quelques heures, là où un projet traditionnel mobilisait développeurs et DevOps pendant des semaines. Ainsi, le cycle d'innovation et d'itération entre l'idée et la réalisation est considérablement réduit et quasiment continu.

La conformité à la réglementation de sécurisation des données personnelles peuvent souvent être des freins au développement d'outils d'IA en interne. Les plateformes no-code intègrent nativement le versioning, la traçabilité des appels API et les obligations RGPD, sans présenter de surcharge technique.

D'un point de vue du modèle économique, le no-code peut également présenter des avantages dans le cadre d'une étape de transition pour un fonds d'investissement vers un modèle data-driven : en effet, les coûts de CAPEX sont très faibles dans le cadre du déploiement d'un système de no-code, remplacés par des coûts d'OPEX pouvant être flexible : l'équipe de gestion paiera à l'appel API et les coûts de chaque agent autonome pourront être coupés si l'agent n'apporte pas satisfaction ou ne présente plus d'utilité.

Comme mentionné plus haut, le no-code pourra être la base d'une étape de transition d'un fonds vers un modèle davantage data-driven, cependant, les limitations inhérentes à ces outils et les coûts pouvant être sur le long terme relativement exponentiels, pousseront la plupart des fonds à recruter dans un second temps une équipe technique afin de développer des outils propriétaires.

Voici, pour rendre ces éléments plus concrets, un schéma de système multi-agents pouvant être mis en place en no-code afin d'accompagner les équipes d'investissement dans leur due-diligence :

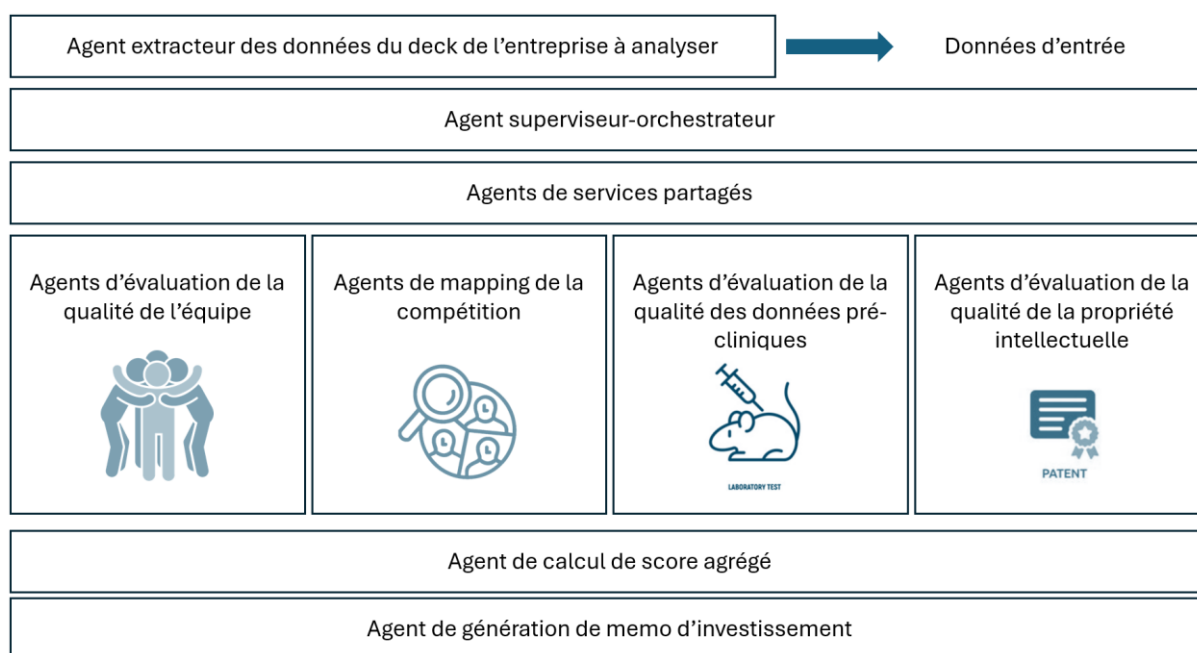


Figure 49 - Proposition d'un système multi-agents pouvant être déployé en no-code

L'agent superviseur-orchestrateur aura pour mission de préparer le plan d'exécution, de déclencher les micro-agents, de surveiller les délais, de relancer les agents tombés en erreur et d'envoyer des messages d'alerte en cas d'impossibilité de relance des agents.

Les agents de services partagés constituent la fondation technique commune sur laquelle s'appuient les micro-agents verticaux (Team, Compétition, Pré-clinique, IP). Leur mission principale est de mutualiser tout ce qui serait coûteux, redondant ou critique à dupliquer dans chaque pipeline. Parmi ces agents, nous pouvons imaginer :

- Un Connector Hub qui aura pour objectif de centraliser et de gérer les appels API et les robots de scraping afin de garantir un accès unifié aux sources externes
- Un data normalizer et deduplicator qui uniformisera les formats, unités et nomenclatures des données, qui fusionnera les doublons issus de différentes sources afin d'assurer la cohérence des données
- Un RGPD-Filter qui aura pour rôle de journaliser chaque requête de sortie, de masquer ou de pseudonymiser les données personnelles (scrapées sur LinkedIn par exemple), cet agent est indispensable à la conformité réglementaire et permettra d'auditer l'outil global.

Les agents en charge de l'évaluation de la qualité de l'équipe vont avoir plusieurs rôles :

- Le « talent crawler » va scraper LinkedIn afin de recueillir les données professionnelles des fondateurs
- Le « calculateur d'impact académique » va analyser, via l'interrogation de l'API Semantic Scholar, l'H-index de l'équipe et le nombre de citations sur les articles publiés. Ces éléments seront un proxy de la capacité de l'équipe – ou du moins d'un de ses membres – à générer de la science de qualité.
- Le « calculateur de score d'exit » : en utilisant des données PitchBook par exemple, cet agent aura pour objectif d'établir un score de prédiction d'exit.
- Le « réputation risk spotter » : cet agent utilisera des recherches en ligne et dans les médias spécialisés pour déterminer l'existence de potentiels red-flags réputationnels sur l'un des membres de l'équipe. Actuellement, différentes vérifications sur les fondateurs sont effectuées par les fonds via des outils spécialisés (recherche d'exposition politique, de mandats de recherche, etc) mais ces vérifications sont toutefois relativement limitées.
- Enfin, nous pouvons imaginer un « synthétiseur de score d'équipe » qui, en utilisant un LLM, permettra de comparer le score de l'équipe par rapport à ceux des concurrents (l'identification des entreprises concurrentes reposera sur un autre pool d'agents autonomes)

Un mapping concurrentiel sera mis en place par un groupe d'autres agents :

- Le « market scanner » va exploiter un large éventail de sources gratuites et ouvertes pour identifier les concurrents de la société cible : des bases de données d'études cliniques : ClinicalTrials.gov ou la WHO ICTRP (International Clinical Trials Registry Platform), les bases de publications scientifiques (PubMed, bioRxiv), les bases de données économiques ou financières (Pitchbook ou CB Insights par exemple), les sites d'actualité et de veille marché (Evaluate Pharma, Fierce Biotech, Endpoint News, etc), les bases de brevets (Lens.org propose par exemple une API gratuite pouvant être utilisée).
- L'« entity linker » aura pour fonction essentielle d'identifier précisément et de relier entre elles les différentes entités citées dans les documents et les données récupérées. L'objectif sera de relier des noms légèrement différents vers une même entité mais aussi de relier des innovations (identifiées dans des publications scientifiques) aux entreprises qui les développent et commercialisent.
- Le « similarity embedder » aura pour fonction de classer les concurrents identifiés selon leur proximité technologique et scientifique.
- Le « funding-stage analyzer » qui déterminera le stade de financement et la solidité financière des concurrents, en

Les agents du pool dédié aux données pré-cliniques :

- Le « paper retriever » aura pour rôle de récupérer automatiquement toutes les publications scientifiques pertinentes associées à la technologie analysée en interrogeant via leurs API les bases documentaires comme PubMed ou bioRxiv. Il filtrera les résultats selon des critères bien définis (date, auteurs, mots-clés, etc)
- L'« experiment extractor » va extraire les données expérimentales essentielles (le design expérimental, le modèle mis en œuvre, les résultats quantitatifs, les doses, les durées, etc) directement depuis l'article scientifique. Ces données seront transformées en format structuré et exploitable.
- Le « stats checker » validera la robustesse statistique des résultats expérimentaux rapportés dans la publication, en réalisant notamment l'extraction des valeurs clés : taille des échantillons, moyennes, écarts-types, intervalles de confiance, etc. L'agent validera les p-values et la significativité.
- L'agent de validation du modèle qui aura pour objectif d'évaluer la pertinence translationnelle des modèles pré-cliniques utilisés vis-à-vis de la pathologie humaine ciblée. Un modèle LLM scorera la pertinence du modèle en fonction de la littérature scientifique de référence.
- L'agent de vérification de la conformité GLP : Vérifier la conformité déclarée des études pré-cliniques aux Bonnes Pratiques de Laboratoire (GLP) en recherchant de manière automatisée les mentions « GLP », « Good Laboratory Practice », etc.
- L'agent d'évaluation de la robustesse et de la reproductibilité des données expérimentales générées : analyse des intervalles de confiance ou des écart-types rapportés, vérification de la possibilité de biais et finalement attribution d'un score.

Les agents dédiés à l'analyse de la propriété intellectuelle :

- L'agent « Patent Retriever » aura pour rôle d'identifier et de récupérer l'ensemble des brevets pertinents déposés par la société analysée en interrogeant les API de bases ouvertes (Lens.org, WIPO, etc) et en stockant de manière structurée les données d'intérêt des brevets récupérés.
- L'agent « Patent Structurer » sera à même de structurer automatiquement les données extraites des brevets bruts afin de les mettre en forme de manière tabulaire ou vectorielle.
- L'agent que nous pourrions appeler « Scientific Relevance Matcher », l'un des plus important du schéma, aura pour rôle d'identifier les liens entre les brevets et la technologie scientifique analysée, afin de déterminer les chevauchements technologiques, les manques ou les zones blanches.
- L'agent « freedom-to-operate analyzer » aura pour mission d'évaluer la liberté d'exploitation de la technologie, en analysant notamment les revendications bloquantes, l'estimation de la portée des brevets tiers, etc. Cet agent ne pourra évidemment pas se substituer à une analyse approfondie d'un cabinet expert en propriété intellectuelle, mais il pourra orienter de manière précoce une due-diligence.
- L'agent « Legal Status Checker » aura pour objectif de suivre la validité juridique des brevets (en cours, abandonné, délivré, opposé, etc)
- L'agent « Patent Landscape Builder » pourra, à l'aide des données concernant la technologie développée par l'entreprise et des données récupérées via les API des bases de données publiques, construire une vue d'ensemble de l'environnement brevet. Les tâches de cet agent seront : la cartographie des familles de brevets, le clustering par thématique et enfin l'analyse temporelle et géographique des dépôts. Des scores d'évaluation des brevets (de leur qualité, de leur risque d'entrer en concurrence avec la

propriété intellectuelle de l'entreprise cible par exemple), pourront être calculés et intégrés à la cartographie.

II] Mise en place de la transformation data-driven d'un fonds de capital risque : par où commencer ? Quelle feuille de route ?

Outre les aspects technologiques de la transformation d'un fonds traditionnel en fonds data-driven, il est important de tenir compte des aspects culturels, humains et organisationnels.

Lorsqu'un fonds traditionnel va décider de s'orienter vers un fonds davantage data-driven, un certain nombre de questions vont se poser à lui (par exemple : faut-il choisir développer des outils en interne ? Utiliser des outils existants ? Ou passer par un prestataire extérieur ?) et les réponses à apporter ne sont pas toujours évidentes.

C'est pourquoi, dans un souci de rendre ce travail de thèse le plus ancré dans le réel possible, nous allons détailler une feuille de route possible de transformation data-driven :

A) Alignement initial et préparation

La première étape sera de définir la vision et le ou les chefs de projet : dès le départ, la direction du fonds doit en effet sponsoriser l'initiative, participer à clarifier les objectifs et dégager du budget pour expérimenter, acheter des abonnements SaaS, etc. Il sera crucial de désigner un chef de projet interne (idéalement un profil combinant appétence pour la tech et connaissance du métier d'investisseur qui pilotera la feuille de route au quotidien. Ce référent sera également à même de faire le lien entre l'équipe d'investissement et les éventuels intervenants extérieurs. Lors de cette première phase, le référent data-driven pourra interroger l'équipe sur ses besoins.

Une autre étape importante sera celle de la communication interne à destination de l'équipe d'investissement afin de la sensibiliser aux enjeux et bénéfices attendus de la démarche data-driven, pour obtenir l'adhésion de chacun. Il faudra souvent insister sur le fait que la data vient augmenter l'expertise et l'intuition de l'investisseur et pas les remplacer. Des réunions de travail amorceront le changement culturel nécessaire.

Durant ce temps de préparation de la transformation data-driven, il sera nécessaire d'organiser des formations, notamment à l'utilisation d'outils no-code. L'objectif étant que le responsable data-driven et quelques autres membres de l'équipe puisse prototyper de simples automatisations. C'est à ce moment-là qu'il pourrait être pertinent de décider si un free-lance externe est nécessaire pour avancer sur les premières automatisation no-code. Ces projets, accompagnés par un professionnel extérieur seront également une excellente opportunité pour l'équipe d'apprendre en faisant.

L'objectif de cette première étape sera de livrer au management de la société de gestion un plan d'action validé avec des objectifs clairs, des rôles assignés, une sélection d'outils no-code et des équipes engagées. Un mapping des outils disponibles « sur étagère » et pouvant être utilisés ou intégrés à un outil développé en interne sera réalisé. (exemple : Finthesis, outil de visualisation de données financières pouvant être intégré à un dashboard plus général). Un premier inventaire des données disponibles sera également réalisé (contacts, deals passés, sources d'information utilisées, etc) ce qui servira pour l'étape suivante.

B) Prototypage no-code des premiers cas d'usage

Cette étape de prototypage aura pour objectif de développer un prototype fonctionnel couvrant les besoins de base d'un cas d'usage (le sourcing des start-up par exemple). L'idée est d'obtenir rapidement un système opérationnel à tester en conditions réelles. Ce pilote servira de preuve de concept, facilitant l'adhésion des équipes.

Pendant le prototypage, le responsable data-driven impliquera activement le reste de l'équipe pour qu'ils se familiarisent avec le nouvel outil. L'objectif est que toute l'équipe adopte le pilote pour saisir et consulter les informations de l'outil au quotidien, afin d'avoir des retours d'utilisation concrets.

L'évaluation du pilote – en quelque sorte un MVP sur un cas d'usage donné – sera également importante : il faudra à cette fin organiser une session de retour d'expérience de l'équipe et mesurer les premiers bénéfices.

L'objectif de cette deuxième étape sera donc de livrer à l'équipe un MVP développé en no-code par le responsable data-driven, avec ou sans l'aide d'un free-lance externe. Il s'agira, si ce MVP est adopté, d'une première victoire dans le processus qui prouvera la valeur de la démarche et qui servira de fondation pour aller plus loin.

C) Extension, itérations et décision « make or buy »

Cette troisième étape aura pour objectif d'apporter des améliorations itératives du prototype et de l'enrichir progressivement en fonctionnalités. Avec l'augmentation des volumes et des types de données utilisées, il sera nécessaire de mettre en place des procédures de nettoyage et de dédoublement des données.

Cette étape sera aussi l'occasion de commencer à documenter le système et de formaliser la modification des processus métiers. Cela professionnalise l'approche et favorise l'onboarding de nouveaux utilisateurs du système IA (dans le cas de l'arrivée d'un nouveau collaborateur par exemple)

Enfin, c'est à ce moment du projet data-driven que l'équipe de gestion aura suffisamment de recul pour pouvoir répondre à la question stratégique de la suite du développement : faut-il continuer en no-code interne ? Faut-il en externaliser une partie ? Ou faut-il recruter un ingénieur data en interne ? Plusieurs facteurs seront à même de guider la réflexion :

- a. **La charge de travail et la complexité technique** : si le volume de données et d'automatisations commence à dépasser ce qu'un non-technicien peut maintenir (risque d'erreurs, besoins de scripts complexes), il peut être temps d'envisager le recours à un spécialiste. Un ingénieur data pourrait être pertinent.
- b. **Le budget disponible** : en fonction de la taille du fonds, le recrutement d'un ingénieur spécialisé et exclusivement dédié à l'élaboration d'un outil IA peut être un investissement trop lourd. Une alternative est de collaborer avec un expert externe sur des missions précises.
- c. **La pérennisation de la solution** : si la solution no-code mise en place permet déjà de répondre à une grande partie des besoins, alors il ne sera pas indispensable d'engager des ressources externes ou de faire des recrutements.

Quel que soit la décision prise, il faudra alors mettre à jour la feuille de route.

D) Structuration à moyen terme et changement d'échelle

Cette étape aura pour but de dépasser le stade du prototypage et d'ancrer la solution dans la durée et dans la vie quotidienne des équipes de gestion. En fonction des choix établis à l'étape précédente, elle pourra prendre différentes dimensions opérationnelles.

Ainsi, le fonds pourra mettre en place une infrastructure plus robuste pour remplacer les outils no-code utilisés durant la phase de prototypage : un data warehouse dans le cloud, des applications front-end codées en python, etc.

En parallèle de cette pérennisation technique, il sera pertinent d'ancrer la transformation culturelle en continuant à investir dans la formation des équipes et dans le suivi des KPIs.

En résumé, l'architecture mise en place sera hybride avec pour objectif de tirer le maximum d'outils SaaS no-code existants en début de parcours (afin de tester différentes options rapidement et à coût réduit, pour itérer rapidement) puis de construire progressivement des outils interne lorsque le volume et la complexité des données et des cas d'usage le rend nécessaire. Ce déploiement en plusieurs phases permet d'acculturer les équipes au changement data-driven avant d'investir dans des ressources humaines et une infrastructure technique.

Afin d'évaluer la pertinence du processus de mise en place data-driven, il sera important de définir dès le début et avec l'équipe des KPIs qui seront suivis dans le temps. Ces KPIs seront propres à chaque fonds et fonction des objectifs définis, mais nous pouvons citer à titre d'exemple :

- Augmentation du volume de deal-flow, en comparant le deal-flow outbound (issu d'un sourcing pro-actif) avec le deal-flow inbound (demandes entrantes spontanées)
- Taux de conversion par étape : est-ce que le fait de mieux filtrer les opportunités par la data permet de diminuer les échecs tardifs faisant perdre du temps à l'équipe.
- Réactivité et suivi : délai moyen de réponse aux fondateurs par exemple, afin de mesurer les gains en matière de productivité de l'équipe d'investissement
- Taux d'utilisation de l'outil par l'équipe : cette mesure permettra de savoir s'il faut accentuer la pédagogie autour de l'équipe à travers de nouvelles réunions d'information ou de nouvelles journées de formation par exemple

En cas de difficulté à convaincre l'équipe, il peut être pertinent de réfléchir à mettre en place des « quick-wins » : des petits outils facilement déployables et ayant un impact immédiatement palpable sur la productivité de chacun.

III] Un VC Data-Driven, et après ?

Les différents outils présentés dans cette thèse, de scoring ou de prédiction auront pour objectif d'informer et d'améliorer le processus décisionnel, tout en conservant une intervention humaine significative et en laissant aux humains un rôle décisionnaire. Ainsi, un VC Data-Driven présente une organisation très similaire aux VC traditionnels, impliquant des équipes d'investissement, des équipes de back-office et des comités de sélection des projets et d'investissement.

Bien que, comme nous l'avons montré dans cette thèse, le niveau d'avancement dans l'évolution vers un modèle data-driven soit encore très hétérogène d'un fonds à l'autre, des voix avant-gardistes essaient d'ores et déjà de dessiner un des futurs possibles de l'investissement en capital risque : le quant-VC.

A) L'émergence des nouveaux VC Data-Driven : les Quant-VC

Le quant-VC, sur le modèle des Hedge Funds quantitatifs – des fonds dont les stratégies de trading reposent sur des modèles mathématiques et des algorithmes visant à détecter des incohérences ou des tendances de marché, dont le fonds pourrait profiter, et souvent entièrement automatisés – reposera principalement voire exclusivement sur des modèles algorithmiques pour prendre des décisions d'investissement en réduisant l'intervention humaine à un rôle de surveillance ou d'ajustement des modèles.

Dans la continuité de ce qui est mis en place par les VC Data-Driven, l'objectif des Quant-VC sera de maximiser la scalabilité, la rapidité et l'objectivité des prises de décision.

La comparaison entre les modèles Data-Driven actuels et modèles Quant-VC, possiblement mis en place à plus long terme, peut se faire suivant plusieurs points :

Tableau 1 - Différences entre Data-Driven VC et Quant-VC

Caractéristique	Data-Driven VC	Quant-VC
Rôle de l'investisseur humain	Prise de décision d'investissement	Intervention humaine minimale
Utilisation des données	Soutien à la décision humaine	Prise de décision automatisée
Technologies	Outils d'analyse reposant quasiment exclusivement sur des modèles de ML et sur des LLM	Modèles algorithmiques avancés (jumeaux numériques pour prédire le succès des études cliniques par exemple)
Scalabilité	Améliorée	Très élevée
Flexibilité	Adaptable aux spécificités de chaque investissement	Peu flexible

Un certain nombre d'éléments doivent toutefois tempérer la pertinence du modèle Quant-VC :

- Le modèle quantitatif fonctionne principalement sur des marchés très liquides, permettant des achats et des ventes très fréquents (problématique de la valorisation relativement subjective des start-up)
- Le nombre des investissements des fonds ne sont pas suffisants pour pouvoir entraîner des modèles prédictifs fiables.

Par ailleurs, la transition vers un modèle Quant-VC, du fait des implications organisationnelles et technologiques très fortes, sera beaucoup plus complexe que celle des VC traditionnels vers les modèles data-driven. Ainsi, il est à prévoir que ce modèle, bien qu'intéressant, ne soit jamais largement adopté, en dehors de fonds réalisant un très grand nombre de petits investissements en développant une stratégie de diversification importante, peu adaptée au monde de la santé où les besoins capitalistiques sont importants.

B) Quelle stratégie d'évaluation des approches data driven ?

Nous l'avons vu, de plus en plus de fonds de venture capital adoptent des stratégies d'intégration de l'intelligence artificielle à leurs processus. Un grand nombre d'entre eux utilisent par ailleurs l'argument de leurs outils développés en interne comme un argument afin de lever des fonds auprès de leurs LPs.

Cependant, à ce jour, les études évaluant la performance des fonds data-driven comparativement aux fonds ayant des pratiques plus traditionnelles sont rares. Nous pouvons cependant à ce titre mentionner l'étude de Melissa Crumling,(51) publiée en fin d'année 2024 et démontrant que les VC data-driven ont une capacité de sourcing en dehors des grands centres technologiques plus approfondie que les VC traditionnels et que les entreprises ainsi sourcées ont plus de chance de mener à une IPO que leurs équivalents issus d'autres zones géographiques.

Plusieurs facteurs peuvent expliquer cette absence de données exhaustives et d'autres études pouvant conduire à une évaluation plus systématique :

- **La confidentialité des données de performance :** les taux de rendement interne des fonds d'investissement et les multiples sur le capital investi sont généralement considérés comme des informations sensibles. Les fonds ne publient pas souvent ces données. Il est toutefois probablement possible de les reconstituer partiellement et approximativement en se basant sur les comités de presse d'investissement dans les sociétés du portefeuille et en suivant les performances de sortie de celles-ci.
- **L'hétérogénéité des approches data-driven,** qui englobent une grande variété d'utilisation de technologies rendent relativement difficile la catégorisation et la comparaison des fonds, qui ne communiquent pas tous en détails sur leur utilisation des technologies. Une manière, encore une fois approximative, d'évaluer leur degré de maturité sur le sujet, consiste à analyser leur historique d'offres d'emploi des fonds : ont-ils recruté des ingénieurs data ? Si oui à partir de quand ?
- **La temporalité des retours sur investissement.** Les fonds de VC ont en effet des horizons de retour sur investissement longs, souvent supérieurs à 7 – 10 ans, ce qui rend insuffisant le recul sur les performances des fonds data-driven.
- **La grande hétérogénéité des performances des fonds de VC.** Les rendements des fonds de capital-risque varient considérablement en fonction de nombreux facteurs, tels que la stratégie d'investissement, le secteur ciblé, la région géographique, la taille du fonds et l'expérience de l'équipe de gestion. Cette variabilité rend difficile l'isolement de l'impact spécifique de l'approche data-driven sur la performance globale.

En dépit des difficultés, la poursuite de la démocratisation des approches data-driven, et notamment la continuation de la montée en puissance de leur popularité auprès des LPs, passera nécessairement par la mise en place de davantage d'études méthodologiquement solides permettant de valider le premium de performance des fonds IA-powered.

Ainsi, une étude quantitative de ce type – peut-être à mener dans la continuité de l'évaluation qualitative présentée dans ce travail de thèse ? – pourrait reposer sur la méthodologie suivante :

- Identification et catégorisation des fonds : en se basant sur des critères comme la communication auprès de l'utilisation d'outils d'IA ou sur le recrutement d'ingénieurs, comme mentionné plus haut
- Collecte de données : en utilisant des bases de données telles que Pitchbook pour recueillir des informations sur les investissements, les sorties.
- Calcul des performances des fonds : Afin de calculer les taux de réussite des entreprises et les approximations des multiples et des TRI.

C) Vers un co-pilote des VC Biotech ?

Le travail de cette thèse s'intéressait spécifiquement à l'utilisation de l'IA dans la prise de décision d'investissement. C'est dans ce domaine que la majorité des efforts d'intégration de l'IA ont porté jusqu'à présent au sein des fonds de VC.

Cependant, l'IA pourrait être utilisée à bien d'autres fins par les équipes de gestion, afin de constituer un co-pilote à 360°, permettant de gérer le dealflow mais aussi le portefeuille, en suivant de près l'évolution des différentes participations du fonds :

- **Veille stratégique, réglementaire et concurrentielle récurrente :**

L'objectif est de construire une veille continue et automatisée sur les concurrents des sociétés du portefeuille (nouvelles études, levées de fonds, accords de licencing, etc), sur les évolutions réglementaires, avec une certaine transversalité sur les sociétés et enfin sur les tendances technologiques pertinentes. Les données utilisées seront issues de sites publics et d'API de bases de données (PubMed, ClinicalTrials.gov, EMA, FDA) mais aussi de revues scientifiques et de newsletters. L'idée est de générer des alertes en temps réel en cas d'événement critique détecté, permettant d'informer les décisions prises en comité stratégique.

- **Analyse automatique de la trésorerie et du respect du business plan :**

L'objectif est de comparer les prévisions du BP à la réalité observée : cash runway, burn mensuel, prochaines levées de fonds, et, en fonction du statut commercial de l'entreprise : la marge brute et les revenus. Cela peut être réalisé de manière assez facile en retirant les données chiffrées des comptes-rendus de board à l'aide de techniques d'OCR ou de NLP. Ces données seront ensuite structurées et rapprochées des objectifs du BP stockées en base de données. Le calcul de l'écart est ensuite mesuré et un score de respect du BP est calculé.

Le résultat pourrait potentiellement être présenté sous forme de graphiques. Les avantages de cet outil sont nombreux : gain de temps sur l'analyse financière manuelle, détection des dérapages silencieux et meilleure anticipation des besoins de refinancement.

- **Recommandations stratégiques générées par IA :**

L'objectif est de fournir à l'investisseur des pistes d'action ou de vigilance à partir de l'analyse des documents et des données des sociétés en portefeuille. Les données auront plusieurs sources différentes : les documents de data room, les CR de comités stratégiques, les emails échangés entre l'équipe de gestion et les dirigeants et les données de veille externe.

Les avantages d'un tel outil seraient notamment de faire émerger de nouvelles idées potentiellement non envisagées par les membres du comité stratégique, d'aider à la prise de recul, de générer de la matière pour les discussions stratégiques et enfin de contribuer à la standardisation du suivi des participations.

L'idéal serait de pouvoir générer 2 à 3 recommandations contextualisées, ainsi qu'une justification pour chaque proposition. Cette fonctionnalité devrait être proactive : envoyer un mail à l'investisseur lorsque la recommandation a été jugée pertinente par l'agent IA chargé de la vérification des recommandations.

- **Chatbot Q&A sur le portefeuille :**

L'objectif de ce chatbot serait de permettre aux membres de l'équipe d'investissement de poser n'importe quelle question sur la société en portefeuille, sans avoir besoin d'aller consulter les

documents contenant les réponses. Ce chatbot reposerait sur les documents relatifs à l'entreprise contenus dans le drive du fonds d'investissement (documents issus de la data room et CR de Comités stratégiques, principalement).

- **Gestion documentaire intelligente :**

L'objectif serait de permettre aux membres de l'équipe de retrouver immédiatement un document dans la data room ou dans le drive du fonds d'investissement. Cet outil reposerait sur les documents présents dans le drive, sur les métadonnées des documents et sur le contenu texte des documents. Ce mécanisme de gestion intelligente reposerait sur une indexation des documents dans une base de données vectorielle et sur un LLM. Cette solution, en complément du chatbot susmentionné permettrait un gain de productivité important : la recherche documentaire représente une charge de travail importante dans un fonds, sans valeur ajoutée.

- **Génération automatique du reporting :**

Le reporting est une tâche chronophage au sein d'un fonds d'investissement : il s'agit, à intervalle régulier d'agréger et de résumer les données issues des entreprises en portefeuille afin de tenir au courant les LPs sur l'avancée du portefeuille.

En fournissant le template du reporting standard du fonds, et en se basant sur les résultats des agents IA cités précédemment : analyse de la courbe de cash, de la poursuite de la roadmap des sociétés et de l'émergence de potentiels concurrents, il est possible de générer des documents de reporting entièrement automatisés. En plus du gain de temps important qui lui est inhérent, cette fonctionnalité permettrait d'améliorer l'archivage et la traçabilité.

Exemple de la présentation graphique que pourrait prendre un tel outil de co-pilotage d'un fonds : (généré par IA)

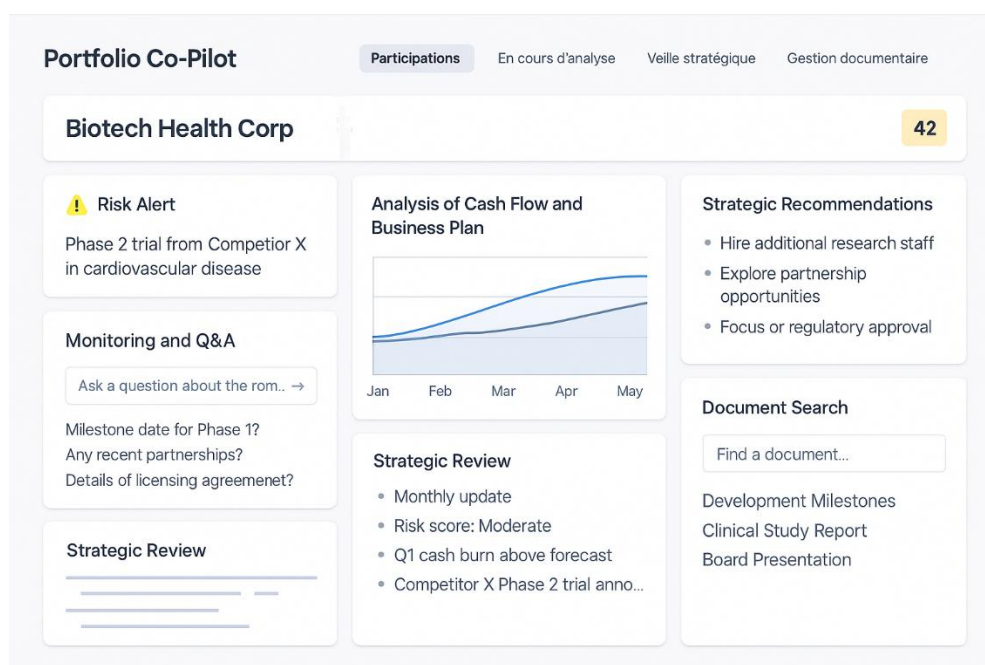


Figure 50 - Exemple d'un tableau de bord VC data-driven (généré par IA)

La mise en place d'un tel outil – inédit dans le domaine du VC Biotech – combiné aux outils d'aide à la décision précédemment détaillés, pourrait représenter un axe de différenciation fort pour des LPs, en recherche de performance et de visibilité sur les investissements réalisés.

Sam Altman, CEO d'OpenAI, prédit que l'IA générative permettra bientôt à un individu seul de bâtir une licorne⁽⁵²⁾ (start-up valorisée plus d'un milliard de dollars). Et si, demain, les plus grands fonds d'investissement n'étaient plus composés d'équipes... mais d'une seule personne, dont les capacités seraient amplifiées par l'intelligence artificielle ?

Outre ces évolutions organisationnelles, l'implémentation de l'IA dans un nombre croissant de fonds de VC tendra naturellement à concentrer ces derniers – les petits fonds n'auront probablement pas les moyens de lutter contre les performances assurées aux grands fonds par des systèmes algorithmiques complexes – et de lisser les performances de tous les fonds : nous l'avons vu plus tôt dans cette thèse, les performances de l'industrie du VC sont particulièrement hétérogènes : le premier quartile des fonds multiplie par 3 les fonds des LPs quand le dernier quartile leur rend moins d'argent qu'ils n'en ont investi.

Le lissage de ces performances, ou en d'autres termes, la suppression d'une partie du risque par les analyses IA, devrait amener davantage de capitaux vers le Venture Capital, ce qui sera très positif pour l'innovation. Le domaine spécifique de la santé, par ses particularités et du fait des nombreuses données disponibles, peut particulièrement tirer son épingle du jeu dans le futur du Venture Capital Data-Driven. Il faut en tout cas le souhaiter, tant les patients ont besoin de nouvelles thérapeutiques et tant les start-ups qui sont à l'origine du développement d'un grand nombre de ces candidats médicaments ont un besoin criant de fonds.

J] Annexe : L'intégration de l'IA aux Venture Builders Biotech

Les Venture Builders, également appelés « start-up studios », représentent une approche particulière de l'innovation entrepreneuriale, fondée non pas, comme les fonds de VC, sur la sélection d'entreprises existantes à financer, mais sur la création proactive de start-ups, à partir d'idées, de technologies ou de besoins identifiés.

Contrairement aux fonds de VC, qui interviennent à posteriori dans des processus entrepreneuriaux exogènes, les Venture Builders adoptent une logique endogène, intégrée et itérative.

Le modèle des Venture Builders présente des atouts particuliers d'un point de vue financier. En effet, une étude indique que les start-ups issues de Venture Builders atteignent le financement de série A en moyenne en 25,2 mois contre 56 mois pour les start-ups traditionnelles(53). De plus, ces start-up affichent un taux de succès supérieur de 30% et un taux de rendement interne moyen de 53%, comparé à 21,3% pour les start-ups classique. Plusieurs facteurs peuvent expliquer cette surperformance, notamment :

- **La mutualisation des ressources et des coûts :** le venture builder dispose d'équipes mutualisés réparties sur plusieurs projets. Cela réduit le coût marginal de création d'une start-up par rapport à un entrepreneur seul. Les workflows (séquence structurée d'actions ou d'étapes permettant d'accomplir un processus métier de manière fluide, logique et efficace) sont également standardisés et mutualisés, ce qui permet également de réduire les coûts et les temps de développement, à ce titre, nous pouvons notamment mentionner des bonnes pratiques en matière de dépôt de brevets, de design d'études précliniques, de demandes de subvention, etc.
- **La détention de parts de la société à coût très bas :** le venture builder prend généralement une part significative du capital (généralement plus de 70%) lors de la création, avec un coût initial limité (en partie sous forme de travail et de ressources en nature).
- **L'accès privilégié aux subventions et financements publics :** En professionnalisant la recherche de financements non-dilutifs, les venture builders obtiennent pour leurs sociétés un taux de succès supérieur à celui des start-ups isolées dans l'accès aux fonds publics, ce qui permet de cofinancer le démarrage de chaque projet avec le moins de capital privé possible.
- **La réduction du taux d'échec global du portefeuille :** Le modèle du Venture Builder permet de réorienter les projets en échec vers des verticales adjacentes et de redéployer le capital humain sur d'autres projets, ce qui aura pour conséquence de diminuer le taux de perte et de lisser la performance du portefeuille.

Les modes de gestion du risque seront différents entre les deux modèles VC et VB :

- Le Venture Builder construit des projets pour maîtriser le risque à la racine, avec un haut niveau de contrôle, dans une logique d'industrialisation de la création de start-ups.
- Le fonds de VC sélectionne les projets les moins risqués en s'appuyant sur des outils analytiques, un réseau d'experts et une logique probabiliste de portefeuille.

Par conséquent, les utilisations possibles de l'IA sont légèrement différentes entre les deux modèles : là où l'IA sert principalement à améliorer la sélection et le suivi des participations dans

les fonds, elle devient pour le Venture Builder un véritable outil de conception, de structuration et d'exécution.

Voici la manière dont des agents IA pourraient être implémentés dans le cadre d'un Venture Builder Santé :

Afin d'aider à l'idéation des projets :

- **L'agent scanner d'opportunités**, dont l'objectif sera d'identifier automatiquement des opportunités scientifiques issues de publications, de brevets ou de conférences. Son implémentation reposera sur un orchestrateur no-code (Make ou n8n), sur les API de Pubmed ou de Lens.org pour sourcer les données et enfin sur l'API de ChatGPT pour les analyser. La mission de cet agent sera de réaliser du scouting technologique en continu afin de générer davantage d'idées de start-ups.
- **L'agent d'évaluation précoce de la pertinence translationnelle**, dont l'objectif sera d'analyser les modèles utilisés dans les publications scientifiques et de déterminer leur robustesse et leur pertinence. Nous avons déjà évoqué ce type d'agents plus haut dans cette thèse.
- **L'agent d'analyse automatique du paysage concurrentiel** qui aura pour fonction de mapper les projets concurrents (à partir des publications, des brevets et des publications) autour d'une cible, d'un mécanisme ou d'une indication. L'agent sera également en mesure de scorer les différenciations potentielles des projets.
- **L'agent de scoring multidimensionnel de « fit-venture » ou agent de priorisation** dont l'objectif sera de déterminer un score de pertinence entrepreneuriale en évaluant objectivement et de manière systématique le potentiel d'un projet biotech en vue de la création d'une start-up. Cet agent permettra de comparer différents projets entre eux, de prioriser les ressources du Venture Builder et d'industrialiser le processus de sélection. Les données d'entrée portant sur le projet pourront être renseignées par l'utilisateur ou directement par les autres agents. Les axes selon lesquels les projets pourraient être analysés sont nombreux : robustesse scientifique, potentiel translationnel, qualification du besoin médical, IP, attractivité du marché, faisabilité de la roadmap, potentiel d'attractivité pour les VCs (exits potentiels, milestones, etc), accès au financement non-dilutif, etc.

Afin d'aider à structurer et à accélérer les projets sélectionnés :

- **L'agent de génération automatique de roadmap de développement** dont l'objectif sera de générer une roadmap de développement optimisée : à partir des données intrinsèques du projet (indication visée, mécanisme d'action, état actuel du projet, préférences stratégiques) et de sa base de connaissance, l'agent sera à même de produire un plan détaillé projeté dans le temps avec des tâches, des processus, des durées, des jalons, etc.
- **L'agent de revue de la littérature** dont la fonction sera de synthétiser automatiquement les publications clés sur un sujet donné pour justifier le rationnel d'un projet. Cet agent permettra de prendre les décisions les plus éclairées possibles mais également d'être une source pour la rédaction automatisée de dossiers d'autorisation d'études précliniques ou cliniques par exemple.
- **L'agent de rédaction de dossiers de financement** qui aura pour objectif de rédiger des pré-drafts pour la BPI, l'EIC Accelerator, etc.

- **L'agent de design de plans expérimentaux** dont l'objectif est de générer rapidement des protocoles précliniques expérimentaux adaptés à la validation des hypothèses scientifiques du projet. Cet agent pourra générer des plans expérimentaux structurés (sélection des modèles, nombre de sujets, tests les plus adaptés, etc) et recommander des partenaires et des CROs.

De manière transversale, il pourra aussi être intéressant d'implémenter un knowledge agent qui constituera une mémoire organisationnelle augmentée. Le capital cognitif d'une structure financière, qu'elle soit un fonds de VC ou un Venture Builder, est un actif hautement stratégique, construit au fil du temps et des expériences accumulées des équipes. Pour autant, cette intelligence collective est rarement formalisée ou interrogeable. L'intégration d'un agent à même de gérer la connaissance peut présenter une innovation majeure en termes de scalabilité. Concrètement, le knowledge agent se composera d'un moteur sémantique RAG connecté à la base documentaire de la structure. Un RAG (Retrieval-Augmented Generation) est une architecture d'IA combinant deux fonctions : la recherche dans une base documentaire vectorisée et la génération de réponses à partir d'un modèle de langage comme GPT-4. Cela permet d'avoir des réponses fiables et personnalisées au contexte de l'entreprise.

Il agira comme un copilote transversal et permettra la recherche intelligente de précédents (erreurs à éviter, apprentissages et méthodologies précédemment utilisées), la génération de réponses à partir des données capitalisées et l'auto-alimentation des autres agents IA.

A terme, l'IA pourrait transformer le modèle actuel des Venture Builders afin de permettre de créer une start-up santé en quelques semaines, en automatisant une grande partie des étapes

A ce jour, et après de nombreuses recherches sur le sujet, il apparaît qu'actuellement aucun Venture Builder Santé n'explore publiquement la possibilité d'automatiser le company building à l'aide de l'intelligence artificielle. Le « Venture Builder AI-Powered » reste un modèle à inventer, combinant, au service du développement de nouveaux médicaments, la puissance de l'IA et les avantages structurels inhérents au modèle du Venture Builder.

K] Annexe 2 : L'utilisation de l'IA pour sourcer et scorer des assets inemployés ou disponibles à la vente

De la même manière que les systèmes d'IA sont en mesure de sourcer et d'évaluer des start-ups, nous l'avons montré dans cette thèse, ils peuvent également être utilisés pour détecter des actifs à céder ou non exploités par l'industrie pharmaceutique et pouvant pourtant potentiellement tout de même être utilisés pour traiter des patients et créer de la valeur.

Créer des start-ups à partir d'assets inemployés ou disponibles à la vente offre des projets déjà partiellement dérisqués et réduit considérablement le délai d'accès au marché.

Les différents types d'assets :

Qu'est-ce que les « ghost assets » ?

Définition :

Un ghost-asset⁽⁵⁴⁾ est un candidat médicament ou une technologie de laboratoire dont le développement a cessé alors qu'il restait juridiquement protégé (brevets, données réglementaires, etc). L'actif continue d'exister sur le plan des droits de propriété intellectuelle mais plus personne ne l'exploite. La plupart des ghost assets sont issus de sociétés de petite taille ayant fait faillite faute de financement suffisant.

Les ghost assets, représentant souvent de l'espoir pour les patients et dont le développement a été abandonné pour des raisons qui ne sont pas de nature scientifique, peuvent être repris et valorisés de 3 manières différentes :

- Lorsque la société est en liquidation : les brevets et les packages de données réglementaires et cliniques sont des actifs confiés au liquidateur qui doit leur trouver preneur pour rembourser les créanciers de l'entreprise. Il s'agit du cas de reprise d'actif le plus simple, mais aussi celui dans lequel il y a le plus de concurrence.
- Lorsque la société a été liquidée et qu'aucun repreneur ne s'est manifesté, le brevet sera radié pour non-paiement des annuités. Il y aura une période durant laquelle il sera possible de « réveiller » les brevets en reprenant le paiement des annuités. En fonction des juridictions, il sera alors possible de contacter le liquidateur et de proposer une somme d'argent pour reprendre l'ensemble de brevets radiés et données réglementaires et cliniques.
- Enfin, encore plus tard après la liquidation de la société, le brevet sera irrémédiablement tombé dans le domaine public et les données ne seront plus récupérables. En cas de souhait de poursuite du développement de l'asset, il sera alors nécessaire de reconstituer de la propriété intellectuelle (en brevetant une nouvelle indication, une formulation ou un combo de molécules par exemple) et un package réglementaire (en se basant sur les publications autour de l'asset). Il s'agit d'une voie beaucoup plus complexe et qui nécessite la plupart du temps de refaire des études cliniques afin de confirmer les données « reconstituées ».

Qu'est-ce que les « shelved assets » ?

Définition :

Dans l'industrie pharmaceutique, il existe des actifs autour desquels de la propriété intellectuelle a été générée mais qui n'ont jamais été commercialisés, ou, le cas échéant, pas pleinement valorisés. Il s'agit notamment de candidats-médicaments abandonnés en cours de recherche, de brevets et technologies non-licenciées. L'industrie pharmaceutique est, du fait de l'attrition inhérente à sa R&D, l'une de celles qui génèrent le plus grand nombre d'actifs abandonnés. On estime qu'environ 80% des composés entrant en phase clinique ne seront in fine jamais commercialisés. Ces molécules sont généralement protégées très tôt dans leur développement, ce qui aboutit à une certaine prolifération des composés « sur étagère » bénéficiant d'une part d'une protection intellectuelle sous la forme d'un ou de plusieurs brevets, et d'autre part d'un package de données pré-cliniques et cliniques. Il s'agit donc d'un gisement d'innovation important et pourtant encore relativement peu exploité.

Les raisons de l'abandon d'un actif :

Il existe différents facteurs pouvant expliquer l'abandon du développement d'un actif, de nature scientifiques, stratégiques, réglementaires, économiques ou encore organisationnels.

- **La cause principale de l'abandon d'une molécule est son insuffisance d'efficacité.** Ainsi, une revue de la littérature portant sur le repositionnement médicamenteux⁽⁵⁵⁾ a montré en 2022 que la majeure partie des échecs en développement clinique était liée à une absence ou à une insuffisance d'efficacité.
- **La deuxième raison la plus courante d'abandon relève de considérations stratégiques et de gestion de portefeuille :** les thérapies dont les projection de ventes sont jugées modestes par rapport à leur coût de développement par rapport à d'autres actifs du pipeline seront abandonnées. On estime qu'environ un quart des molécules abandonnées ne le sont ni pour des raisons d'efficacité ni pour des raisons de safety.⁽⁵⁶⁾ Parmi ces raisons stratégiques, il est intéressant d'en différencier un certain nombre :
 - Les perspectives de marché : faible taille de la population cible, concurrence existante intense, payeurs limités. Ces considérations sont toujours relatives aux autres actifs du portefeuille. Ainsi, un marché jugé faible pour une grande entreprise pharmaceutique pourra permettre à une plus petite entreprise de débiter une activité rentable.
 - Fusions-acquisitions et redondances : Lors d'une consolidation industrielle, les portefeuilles de deux entreprises seront fusionnés et des redondances pourront alors apparaître et conduiront à l'abandon de certaines molécules, mêmes si les données déjà générées montrent qu'elles sont sûres et actives.
 - Les erreurs d'appréciation sont aussi possibles : il s'agira de mauvaises décisions de stopper le développement de molécules. Ces jugements peuvent se baser par exemple sur des effets de classe (l'échec d'une autre molécule ayant les mêmes caractéristiques ou les mêmes mécanismes d'action) ou au contraire sur l'avancée rapide d'un concurrent (une pharma peut couper un programme de développement si elle estime qu'un concurrent arrivera sur le marché avant son produit).
- Les problèmes de sécurité liés à la toxicité du médicament et aux effets indésirables graves constatés au cours des études cliniques sont également une cause d'arrêt possible du développement des produits.
- Enfin, des problèmes liés aux décisions de développement en tant que telles peuvent également conduire à l'abandon de programmes de développement : par exemple, si le cadre réglementaire évolue de manière défavorable, une entreprise peut décider de ne

pas poursuivre un projet. Dans la littérature, on rapporte quelques cas où les autorités réclamaient plus de données avant de délivrer l'autorisation de mise sur le marché, ce qui a contribué à geler le développement de certains candidats médicaments, faute de temps ou de budget.

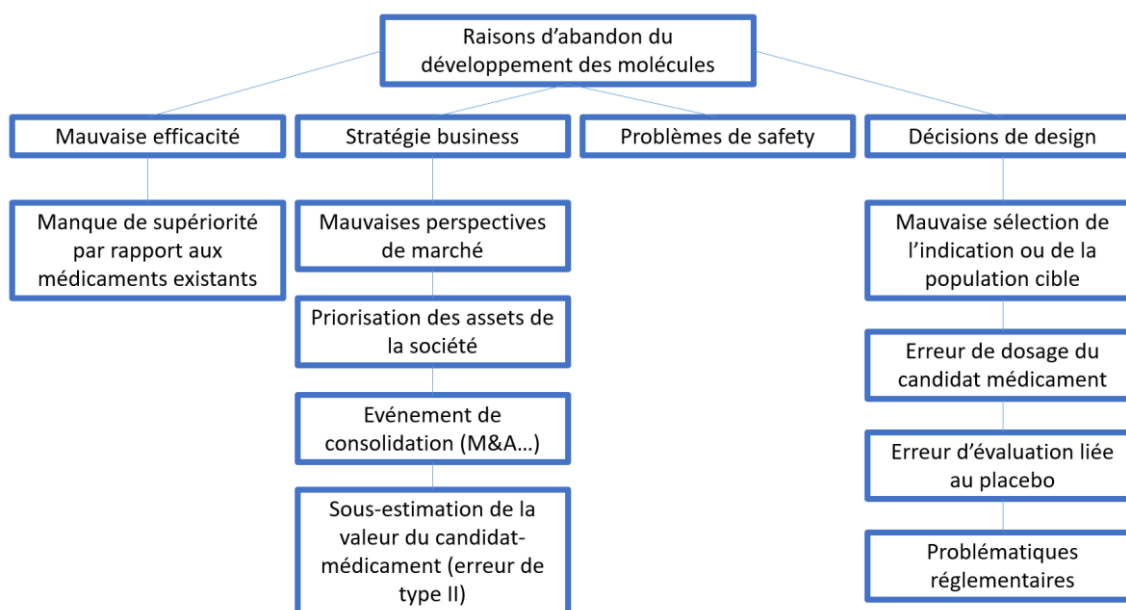


Figure 51 - Les raisons de l'abandon du développement d'un candidat médicament

Comment recréer de la valeur autour d'un shelved asset ?

Les méthodes pour créer de la valeur en partant d'un shelved asset peuvent s'appuyer sur un grand nombre de leviers réglementaires, scientifiques, industriels ou commerciaux. Elles pourront être plus ou moins pertinentes en fonction des raisons de l'abandon du développement du produit.

- **Repositionnement vers une nouvelle indication**

L'objectif sera d'analyser le mécanisme d'action et de faire émerger des indications mieux alignées avec celui-ci. D'un point de vue réglementaire, des données seront réutilisables et, en fonction de l'état d'avancement du développement clinique de l'asset, il est possible que seules une « bridging study » et une étude de phase II dans la nouvelle indication soient nécessaires. Ce repositionnement impliquera

- **« Orphanisation » et niches réglementaires**

Cibler une maladie rare permettra de transformer le profil financier de l'asset, à travers différents mécanismes : prolongation des exclusivités d'exploitation, réduction des coûts de développement (crédit d'impôts US, subventions en Europe ou encore utilisation de cohortes plus petites dans les études cliniques) ou encore des avantages commerciaux : prix premium, pénétration rapide via les associations de patients et accès compassionnel.

- **Exploration de nouvelles géographies**

Un projet abandonné dans les géographies classiques occidentales peut révéler son potentiel dans d'autres zones du monde : par exemple, les maladies métaboliques ou infectieuses ont souvent une prévalence et une sévérité plus élevées en Amérique latine, en Asie du Sud-Est ou

en Afrique du Nord. Il sera également possible de mettre en œuvre un modèle « Emerging-Market Brand » avec un packaging local, une chaîne logistique régionale et un accès au marché différencié.

- **Amélioration galénique**

Une reformulation galénique peut parfois permettre de dépasser les problématiques ayant entraîné l'abandon du développement de l'asset. Cela peut permettre d'améliorer la PK (dans le cadre par exemple du développement d'une forme à libération prolongée par exemple), l'adhérence au traitement ou encore la commodité de la prise et la qualité de vie du patient.

Cette reformulation peut également être l'occasion de déposer de la propriété intellectuelle additionnelle et de prolonger la durée d'exclusivité.

- **Développement d'une stratégie impliquant des biomarqueurs**

L'objectif de cette stratégie sera de cibler une sous-population qui sera sélectionnée en fonction de la présence d'un biomarqueur. Cette stratégie peut permettre de passer d'un résultat clinique négatif à un résultat positif sur une sous-population. Les biomarqueurs permettront de réduire l'hétérogénéité et la taille de l'échantillon de patients.

Cette stratégie permettra potentiellement de développer un test de diagnostic compagnon et des revenus additionnels.

- **Optimisation de la production industrielle et du CMC**

L'objectif de cette stratégie sera de diminuer le coût de production, si ce point pose problème, afin d'améliorer la marge brute.

- **Captation d'incitations financières**

Le fait de loger un asset dans une start-up plutôt que dans un grand groupe pharmaceutique permettra de diminuer les besoins capitalistiques pour le développer du fait, notamment, de l'obtention de subventions. En effet, les autorités peuvent encourager la relance d'actifs via différentes modalités : le priority review voucher (PRV), dédié aux maladies tropicales négligées ou pédiatriques, les subventions, les crédits d'impôts, etc... Le développement d'un actif pouvait ne pas être rentable lorsqu'il était porté exclusivement par une entreprise et le devenir lorsqu'une part importante des coûts du développement seront pris en charge par des financements non-dilutifs.

Qu'est-ce que les « divestitures assets » ?

Définition :

Les divestitures assets sont des médicaments faisant l'objet d'une divestiture : une opération par laquelle une entreprise cède un actif non stratégique à une autre entreprise, dans l'objectif de recentrer son activité principale, d'améliorer son efficacité opérationnelle ou de se conformer à des exigences réglementaires. Il s'agit généralement d'actifs commercialisés et matures.

Des acteurs sont désormais spécialisés dans l'acquisition d'assets en divestiture. Nous pouvons à ce titre présenter Juvisé Pharmaceuticals qui a été fondée en 2008 en France et qui se positionne dans les domaines de l'oncologie, de la cardiologie, des maladies neuro-psychiatriques et de la gastroentérologie. Ce modèle, impliquant une restructuration de la chaîne de production en Europe et l'expansion internationale permet à Juvisé d'injecter de la valeur grâce

à une expertise marketing et opérationnelle forte. La stratégie est donc de capitaliser sur des produits en fin de vie chez les big pharmas pour en prolonger et maximiser la rentabilité.

Une proposition adaptée à un Venture Builder

Un Venture Builder, tel que décrit dans l'annexe 1, est un outil permettant, dans le domaine de la biotech, de développer des start-ups à partir d'innovations sourcées dans des laboratoires de recherche académique.

Une proposition pouvant présenter un grand nombre d'avantages structurels serait de concevoir un Venture Builder hybride combinant à la fois le développement de nouveaux actifs issus de la recherche et exploitation d'actifs matures issus de l'industrie pharmaceutique.

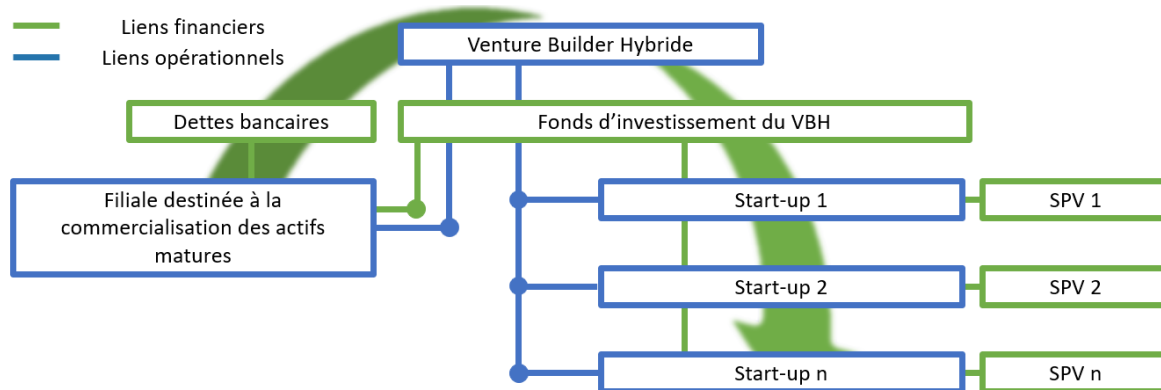


Figure 52 - Proposition d'organisation d'un Venture Builder hybride

Le modèle unique proposé dans le schéma ci-dessus repose sur l'acquisition et le repositionnement rapide de molécules déjà validées cliniquement, et pour la plupart déjà commercialisées et générant des revenus précoces et récurrents. Ces flux financiers sont pour partie réinvestis dans le développement de start-ups innovantes.

L'objectif sera, à terme, de mettre en œuvre un modèle de « **self-funded innovation engine** » : une structure profitant des revenus générés par les actifs matures pour porter le développement de médicaments innovants. Ainsi, après plusieurs années d'existence, il est possible d'envisager un modèle dans lequel le fonds d'investissement accolé au Venture Builder n'aurait plus qu'un seul LP : le Venture Builder lui-même.

Les avantages structurels d'un Venture Builder Hybride :

- **La diversification du risque** : en combinant des actifs innovants (présentant un risque élevé mais également des potentiels de rendement très élevés) et des actifs anciens, ayant déjà fait la preuve de leur efficacité et leur potentiel commercial (risque modéré, rendement rapide et prévisible), ce Venture Builder Hybride peut développer une proposition de valeur unique.
- **Une attractivité renforcée auprès des investisseurs** : En offrant la possibilité pour les LPs d'investir selon leur profil de risque, soit dans des actifs sécurisés, soit dans des start-up disruptives à forts potentiels de croissance, ou les deux, le Venture Builder hybride peut s'adresser à de nouvelles typologies d'acteurs.
- **Un effet de levier important** : La présence d'actifs déjà commercialisés, présentant des revenus récurrents et tangibles

- **La timeline de génération des revenus** : Les divestitures assets peuvent générer des flux financiers rapidement. Après le remboursement des dettes bancaires levés pour financer l'achat des droits de ces actifs, il est donc possible d'utiliser les revenus pour financer les start-ups lancées.
- **Une synergie importante des talents** : En effet, les talents du Venture Builder destinés à opérer la filiale et la commercialisation des actifs matures (réglementaire, affaires pharmaceutiques, accès au marché, marketing, etc) peuvent également être impliqués dans le succès des start-ups. Les revenus générés par ces actifs matures permettront d'embaucher des profils expérimentés. Les réseaux mis en place dans le cadre de la filiale (CRO, CMO, Distributeurs, consultants extérieurs, etc) peuvent également être sollicités pour assurer le succès des start-ups.
- **La capacité, à terme, à fonctionner sans LPs** : les revenus générés par les produits matures, une fois les emprunts bancaires remboursés, permettront au Venture Builder d'investir directement dans les nouvelles start-ups sans faire appel à des LPs. Ainsi, une part maximale de la valeur créée sera conservée au sein du Venture Builder. Ce modèle permet donc d'optimiser la capacité à prendre des décisions stratégiques à plus long terme, non influencées par les horizons d'investissement contraints des LPs. Il permet aussi une diminution progressive des coûts liés aux financements (multiples servis aux LPs). Enfin, le fait de pouvoir fonctionner sans LPs à terme permet au modèle de présenter une stabilité accrue face aux cycles économiques et une position concurrentielle renforcée.

Un modèle d'IA au service d'un Venture Builder hybride

Une des conditions de la réussite d'un modèle de ce type est de développer **une asymétrie d'information** par rapport aux concurrents (qu'ils souhaitent reprendre les droits d'actifs matures ou développer des start-up afin de développer des actifs issus de la recherche académiques) mais aussi par rapport aux sociétés pharmaceutiques avec lesquelles les deals seront passés.

L'objectif de cette asymétrie d'information sera de mieux sélectionner les deals et de prendre des décisions stratégiques plus éclairées.

Cela peut être mis en place au moyen d'un « système d'exploitation » du Venture Builder, une architecture d'agents IA :

- **Sourcing automatique des opportunités** :

L'objectif sera d'identifier de manière systématique et proactive les opportunités d'actifs pharmaceutiques existants disponibles à l'acquisition.

- **Agent de surveillance** :

Cet agent aura pour objectif de scraper automatiquement les bases publiques (Clinical Trials.gov, FDA, EMA, USPTO, Google Patents, etc) et de lancer des alertes automatisées en cas d'arrêt de développement clinique. Cet agent concernera surtout les shelved assets.

- **Agent de tracking de transactions** :

La mission de cet agent sera de suivre automatiquement les annonces de désinvestissement (à travers des communiqués de presse ou des résultats trimestriels par exemple). Cet agent sera également capable d'extraire automatiquement les deals historiques (transactions passées,

conditions financières, etc). Il sera également important que cet agent soit programmé pour détecter précocement des signaux faibles (discussions LinkedIn, forums spécialisés, etc)

- **Screening préliminaire des actifs identifiés :**

- **Agent de qualification IP :**

L'objectif de cet agent sera de vérifier automatiquement la qualité des brevets protégeant l'asset (quelle couverture géographique, durée de protection restante, etc). L'idée pourra être également d'analyser le statut réglementaire de l'asset.

- **Agent d'analyse du marché :**

Cet agent, en se basant sur des données épidémiologiques disponibles, sera en mesure de générer des premières hypothèses de marché (taille estimée, besoins non satisfaits, pricing, etc) et de générer des hypothèses de repositionnement en fonction des caractéristiques du produit.

- **Agent de validation scientifique :**

La mission de cet agent sera d'évaluer si le rationnel scientifique est encore pertinent (en répondant par exemple à la question de savoir si le mode d'action a été abandonné par d'autres molécules concurrentes, etc).

- **Agent de scoring médico-économique et de modélisation économique :**

Cet agent va estimer le potentiel de valeur ajoutée thérapeutique et économique : est-ce que le besoin médical est réellement non-satisfait ? Quel est le niveau de différenciation par rapport au standard-of-care ? Quel est l'impact envisageable sur la qualité de vie des patients ? Sur leur santé ? Sur le coût pour le système de santé ?

Une fois ces différentes informations recueillies, et en tenant compte des coûts de développement du projet, il sera possible de réaliser des projections économiques types : NPV, ROI, etc.

- **Agent de repositionnement thérapeutique :**

L'objectif de cet agent sera de proposer des indications alternatives possibles (sur l'exemple de ce qui a déjà été réalisé par FutureHouse, mentionné plus haut dans cette thèse). L'objectif sera, sur la base des données pharmacologiques générées, de repositionner le composé sur des indications orphelines ou de niche.

- **Agent d'évaluation du Développement Préclinique et Clinique :**

L'objectif de cet agent sera de déterminer les étapes nécessaires pour relancer l'actif vers une valorisation. L'agent devra tenir compte du package des données déjà générées et travailler à un plan de relance, proposer un design clinique alternatif en cas d'échecs cliniques précédents.

Bibliographie :

1. Rasouli M, Chiruvolu R, Rishch A. AI for Investment: A Platform Disruption [Internet]. arXiv; 2023 [cité 11 juill 2025]. Disponible sur: <http://arxiv.org/abs/2311.06251>
2. Ma S, Ni S, Si A, La R, Di M, Mo O, et al. Outcomes for children and adolescents with cancer: challenges for the twenty-first century. J Clin Oncol Off J Am Soc Clin Oncol [Internet]. 20 mai 2010 [cité 26 juill 2025];28(15). Disponible sur: <https://pubmed.ncbi.nlm.nih.gov/20404250/>
3. Ty O, Li H, Ji M, Li S, Hu MX, N THN, et al. Global mortality of chronic liver diseases attributable to Hepatitis B virus and Hepatitis C virus infections from 1990 to 2019 and projections to 2030. J Infect Public Health [Internet]. juill 2024 [cité 26 juill 2025];17(7). Disponible sur: <https://pubmed.ncbi.nlm.nih.gov/38838606/>
4. Poliomyelitis [Internet]. [cité 26 juill 2025]. Disponible sur: <https://www.who.int/news-room/fact-sheets/detail/poliomyelitis>
5. Castillo M. CNBC. 2016 [cité 11 juill 2025]. Chan Zuckerberg donate \$3 billion toward curing all disease by next 100 years. Disponible sur: <https://www.cnbc.com/2016/09/21/chan-zuckerberg-initiative-dedicates-3-billion-towards-disease-research.html>
6. Demis Hassabis: 'AI will affect the whole world ... It's going to change everything' [Internet]. [cité 11 juill 2025]. Disponible sur: <https://www.thetimes.com/life-style/celebrity/article/demis-hassabis-ai-could-cure-all-diseases-in-10-years-09pcqh7cb>
7. Anthropic's CEO thinks AI will lead to a utopia — he just needs a few billion dollars first | The Verge [Internet]. [cité 11 juill 2025]. Disponible sur: <https://www.theverge.com/2024/10/16/24268209/anthropic-ai-dario-amodei-agi-funding-blog>
8. Trump announces \$500B AI project — with biz titans saying it can cure cancer [Internet]. [cité 11 juill 2025]. Disponible sur: <https://nypost.com/2025/01/21/us-news/trump-announces-500b-ai-project-with-biz-titans-saying-it-can-cure-cancer/>
9. Accelerating the development of life-saving treatments - OpenAI x Moderna [Internet]. [cité 11 juill 2025]. Disponible sur: <https://openai.com/index/moderna/>
10. Armstrong A. New AI drug discovery powerhouse Xaira rises with \$1B in funding [Internet]. 2024 [cité 11 juill 2025]. Disponible sur: <https://www.fiercebiotech.com/biotech/new-ai-drug-discovery-powerhouse-xaira-rises-1b-funding>
11. Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov. mars 2012;11(3):191-200.
12. Unlocking R&D productivity - Deloitte [Internet]. [cité 11 juill 2025]. Disponible sur: <https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Life-Sciences-Health-Care/deloitte-uk-measuring-return-on-pharma-innovation-report-2018.pdf>
13. Novartis CEO Who Wanted To Bring Tech Into Pharma Now Explains Why It's So Hard [Internet]. [cité 11 juill 2025]. Disponible sur:

<https://www.forbes.com/sites/davidshaywitz/2019/01/16/novartis-ceo-who-wanted-to-bring-tech-into-pharma-now-explains-why-its-so-hard/>

14. Mendoza C. LPs expect expanding role of AI in PE – Collier Capital Barometer [Internet]. Private Equity International. 2023 [cité 11 juill 2025]. Disponible sur: <https://www.privateequityinternational.com/lps-expect-expanding-role-of-ai-in-pe-coller-capital-barometer/>
15. Sofinnova Partners unveils Sofinnova.AI: A cutting-edge artificial intelligence platform set to transform its life sciences investment practice — Sofinnova Partners [Internet]. [cité 11 juill 2025]. Disponible sur: <https://sofinnovapartners.com/news/sofinnova-partners-unveils-sofinnova-ai-a-cutting-edge-artificial-intelligence-platform-set-to-transform-its-life-sciences-investment-practice>
16. OpenAI annonce l'arrivée de l'AGI pour 2025 [Internet]. [cité 11 juill 2025]. Disponible sur: <https://datascientest.com/openai-annonce-larrivee-de-lagi-pour-2025>
17. Hern A, editor AHU technology. Elon Musk predicts superhuman AI will be smarter than people next year. The Guardian [Internet]. 9 avr 2024 [cité 11 juill 2025]; Disponible sur: <https://www.theguardian.com/technology/2024/apr/09/elon-musk-predicts-superhuman-ai-will-be-smarter-than-people-next-year>
18. Experts L. Que devient l'IA symbolique ? [Internet]. FRENCHWEB.FR. 2018 [cité 11 juill 2025]. Disponible sur: <https://www.frenchweb.fr/que-devient-lia-symbolique/322592>
19. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need [Internet]. arXiv; 2023 [cité 11 juill 2025]. Disponible sur: <http://arxiv.org/abs/1706.03762>
20. Zoom sur Georges Doriot : Père du capital risque [Internet]. [cité 11 juill 2025]. Disponible sur: <https://www.alumneye.fr/zoom-sur-georges-doriot-pere-du-capital-risque/>
21. Ventures VB. A double-edged sword: VC timelines are getting longer in biotech [Internet]. BioVox. 2025 [cité 11 juill 2025]. Disponible sur: <https://biovox.eu/a-double-edged-sword-vc-timelines-are-getting-longer-in-biotech/>
22. Five charts showing how VC funding has shifted by sector [Internet]. [cité 11 juill 2025]. Disponible sur: <https://carta.com/data/vc-shifts-2023/>
23. Biopharma venture financing declines 20.2% YoY in Q1 2025 amid persistent investor caution, reveals GlobalData [Internet]. GlobalData. 2025 [cité 11 juill 2025]. Disponible sur: <https://www.globaldata.com/media/business-fundamentals/biopharma-venture-financing-declines-20-2-yoy-in-q1-2025-amid-persistent-investor-caution-reveals-globaldata/>
24. Data Driven VC Landscape 2025 [Internet]. [cité 16 août 2025]. Disponible sur: <https://datadrivenvc.io/data-driven-vc-landscape-2025>
25. EUVC. Embracing the VC data-driven transformation [Internet]. 2024 [cité 11 juill 2025]. Disponible sur: <https://www.eu.vc/p/embracing-the-vc-data-driven-transformation>
26. InReach Ventures, the « AI-powered » European VC, closes new €53M fund | TechCrunch [Internet]. [cité 26 juill 2025]. Disponible sur: https://techcrunch.com/2019/02/11/inreach-ventures-the-ai-powered-european-vc-closes-new-e53m-fund/?utm_source=chatgpt.com

27. Alumni jessica. Correlation Ventures: Taking the « Gut Feel » out of VC [Internet]. Digital Innovation and Transformation. [cité 26 juill 2025]. Disponible sur: <https://d3.harvard.edu/platform-digit/submission/correlation-ventures-taking-the-gut-feel-out-of-vc/>
28. Patent thicket. In: Wikipedia [Internet]. 2025 [cité 11 juill 2025]. Disponible sur: https://en.wikipedia.org/w/index.php?title=Patent_thicket&oldid=1283467368
29. Home [Internet]. [cité 11 juill 2025]. Disponible sur: <https://www.dennemeyer.com/octimine/>
30. PatentPal [Internet]. [cité 11 juill 2025]. PatentPal. Disponible sur: <https://patentpal.com/>
31. Spohr C. Build an Instant Patent Landscape With Gen AI [Internet]. LexisNexis Intellectual Property Solutions. 2025 [cité 11 juill 2025]. Disponible sur: <https://www.lexisnexisip.com/resources/instant-patent-landscape/>
32. gurjinder.rana@xlpat.com. Empowering Patent Analysis with AI: A New Era in Intellectual Property Management - XLSCOUT [Internet]. 2024 [cité 11 juill 2025]. Disponible sur: <https://xlscout.ai/empowering-patent-analysis-with-ai-a-new-era-in-intellectual-property-management/>
33. Write Patents With AI | Solve Intelligence [Internet]. [cité 11 juill 2025]. Disponible sur: <https://www.solveintelligence.com>
34. Patlytics • Premier AI-Powered Patent Intelligence [Internet]. [cité 11 juill 2025]. Disponible sur: <https://www.patlytics.ai>
35. IPRally | AI Patent Search, Review & Classification [Internet]. [cité 11 juill 2025]. Disponible sur: <https://www.iprally.com/>
36. Patsnap | AI-powered IP and R&D Intelligence [Internet]. [cité 11 juill 2025]. Disponible sur: <https://www.patsnap.com/>
37. Raleigh A. Deconstructing the Diligence Process: An Approach to Vetting New Product Theses [Internet]. LifeSciVC. 2024 [cité 11 juill 2025]. Disponible sur: <https://lifescivc.com/2024/04/deconstructing-the-diligence-process-an-approach-to-vetting-new-product-theses/>
38. Building Effective AI Agents [Internet]. [cité 11 juill 2025]. Disponible sur: <https://www.anthropic.com/engineering/building-effective-agents>
39. Gottweis J, Weng WH, Daryin A, Tu T, Palepu A, Sirkovic P, et al. Towards an AI co-scientist [Internet]. arXiv; 2025 [cité 11 juill 2025]. Disponible sur: <http://arxiv.org/abs/2502.18864>
40. FutureHouse Platform: Superintelligent AI Agents for Scientific Discovery | FutureHouse [Internet]. [cité 11 juill 2025]. Disponible sur: <https://www.futurehouse.org/research-announcements/launching-futurehouse-platform-ai-agents>
41. Demonstrating end-to-end scientific discovery with Robin: a multi-agent system | FutureHouse [Internet]. [cité 16 août 2025]. Disponible sur:

https://www.futurehouse.org/research-announcements/demonstrating-end-to-end-scientific-discovery-with-robin-a-multi-agent-system?utm_source=chatgpt.com

42. Ghafarollahi A, Buehler MJ. SciAgents: Automating scientific discovery through multi-agent intelligent graph reasoning [Internet]. arXiv; 2024 [cité 11 juill 2025]. Disponible sur: <http://arxiv.org/abs/2409.05556>
43. Su H, Chen R, Tang S, Yin Z, Zheng X, Li J, et al. Many Heads Are Better Than One: Improved Scientific Idea Generation by A LLM-Based Multi-Agent System [Internet]. arXiv; 2025 [cité 11 juill 2025]. Disponible sur: <http://arxiv.org/abs/2410.09403>
44. Lo AW, Siah KW, Wong CH. Machine Learning With Statistical Imputation for Predicting Drug Approvals. Harv Data Sci Rev [Internet]. 3 juill 2019 [cité 11 juill 2025];1(1). Disponible sur: <https://hdsr.mitpress.mit.edu/pub/ct67j043/release/10>
45. Using machine learning to better predict clinical trial outcomes | MIT Sloan [Internet]. 2019 [cité 18 août 2025]. Disponible sur: <https://mitsloan.mit.edu/ideas-made-to-matter/using-machine-learning-to-better-predict-clinical-trial-outcomes>
46. Fu T, Huang K, Xiao C, Glass LM, Sun J. HINT: Hierarchical Interaction Network for Trial Outcome Prediction Leveraging Web Data [Internet]. arXiv; 2022 [cité 18 août 2025]. Disponible sur: <http://arxiv.org/abs/2102.04252>
47. Cho C, Lee S, Bang D, Piao Y, Kim S. ChemAP: predicting drug approval with chemical structures before clinical trial phase by leveraging multi-modal embedding space and knowledge distillation. Sci Rep. 3 oct 2024;14(1):23010.
48. Leading AI Drug Development Company | Intelligencia AI [Internet]. [cité 18 août 2025]. Disponible sur: <https://www.intelligencia.ai/>
49. Streamline Clinical Trials with AI and Digital Twins of Patients [Internet]. [cité 18 août 2025]. Disponible sur: <https://www.unlearn.ai/>
50. Yue L, Xing S, Chen J, Fu T. ClinicalAgent: Clinical Trial Multi-Agent System with Large Language Model-based Reasoning [Internet]. arXiv; 2024 [cité 11 juill 2025]. Disponible sur: <http://arxiv.org/abs/2404.14777>
51. Crumling M. Breaking Network Barriers in the Era of Data-Driven Venture Capitalists [Internet]. Rochester, NY: Social Science Research Network; 2024 [cité 11 juill 2025]. Disponible sur: <https://papers.ssrn.com/abstract=4941953>
52. Sawers P. AI agents could birth the first one-person unicorn — but at what societal cost? [Internet]. TechCrunch. 2025 [cité 11 juill 2025]. Disponible sur: <https://techcrunch.com/2025/02/01/ai-agents-could-birth-the-first-one-person-unicorn-but-at-what-societal-cost/>
53. Why Venture Studio Model Have Higher Long-Term Success Rates [Internet]. [cité 11 juill 2025]. Disponible sur: <https://www.bundl.com/articles/why-venture-studio-startups-have-higher-long-term-success-rates>
54. How To Mine Biopharma's IP Storeroom for Rare Disease Drugs, Just Like SpringWorks - BioSpace [Internet]. [cité 11 juill 2025]. Disponible sur:

<https://www.biospace.com/business/how-to-mine-biopharmas-ip-storeroom-for-rare-disease-drugs-just-like-springworks>

55. Krishnamurthy N, Grimshaw AA, Axson SA, Choe SH, Miller JE. Drug repurposing: a systematic review on root causes, barriers and facilitators. BMC Health Serv Res. 29 juill 2022;22(1):970.
56. Bakker A. Don't let « shelved assets » gather dust. Make them into new lifesaving drugs [Internet]. STAT. 2021 [cité 11 juill 2025]. Disponible sur: <https://www.statnews.com/2021/12/13/dont-let-shelved-assets-gather-dust-turn-them-into-lifesaving-drugs/>

Université de Lille

UFR3S-Pharmacie

DIPLOME D'ETAT DE DOCTEUR EN PHARMACIE

Année Universitaire 2025/2026

Nom : DEMAILLY

Prénom : ALEXANDRE

Titre de la thèse :

Prise de décision d'investissement dans les biotechs : vers une disruption par l'Intelligence Artificielle ? - L' « evidence-based investment » à la portée des fonds de Venture Capital

Mots-clés : Investissement, Capital-Risque, Biotech, Intelligence Artificielle, Agents autonomes

Résumé :

Cette thèse a pour objectif d'étudier la manière dont l'intelligence artificielle pourrait aider les fonds de capital-risque spécialisés en santé dans leurs prises de décision d'investissement dans des biotechs. L'allocation optimale du capital dans des phases précoces de développement médicamenteux étant essentielle à la mise à disposition des patients des meilleurs traitements.

Un panorama de l'utilisation de l'IA dans les fonds d'investissement a été dressé à travers une étude internationale des pratiques.

Les avancées récentes de l'Intelligence Artificielle permettent de concevoir des outils en mesure d'analyser la propriété intellectuelle, les données pré-cliniques ou cliniques générées. Ces différents cas d'usage sont explorés dans cette thèse.

Les bases d'autres outils – un copilote permettant de suivre un portefeuille de sociétés, un outil de company building ou encore un outil de sourcing d'actifs abandonnés par l'industrie pharmaceutique – sont également imaginées et discutées.

Membres du jury :

Président & Directeur de thèse : M. MORGENROTH THOMAS, Maître de Conférence des Universités Droit et d'Economie Pharmaceutique – UFR3S Pharmacie

Assesseur(s) : Mme. PINÇON CLAIRE, Maître de Conférence des Universités en Biomathématiques – UFR3S Pharmacie

Membre extérieur : Mme. FOUTEL VERONIQUE, CEO – InBrain Pharma

Membre extérieur : M. MARCHAND LAURENT, Responsable du programme scientifique – Paris Santé Campus