



HAL
open science

Utilisation de l'Intelligence Artificielle dans la gestion des maladies chroniques des animaux de compagnie : opportunités et freins à l'adoption

Nicolas Chauchart Du Mottay

► To cite this version:

Nicolas Chauchart Du Mottay. Utilisation de l'Intelligence Artificielle dans la gestion des maladies chroniques des animaux de compagnie : opportunités et freins à l'adoption. Médecine vétérinaire et santé animale. 2023. dumas-04303523

HAL Id: dumas-04303523

<https://dumas.ccsd.cnrs.fr/dumas-04303523v1>

Submitted on 23 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Année 2023

UTILISATION DE L'INTELLIGENCE ARTIFICIELLE DANS LA GESTION DES MALADIES CHRONIQUES DES ANIMAUX DE COMPAGNIE : OPPORTUNITÉS ET FREINS À L'ADOPTION

THÈSE

pour obtenir le diplôme d'État de
DOCTEUR VÉTÉRINAIRE
présentée et soutenue publiquement devant
la Faculté de Médecine de Créteil (UPEC)
le 24 octobre 2023

par

Nicolas, Charles, Marie CHAUCHART DU MOTTAY

sous la direction de
Chantal LEGRAND

JURY

Président du jury :	M. Henry CHATEAU	Professeur à l'EnvA
Directrice de thèse :	Mme Chantal LEGRAND	Maître de Conférences à l'EnvA
Examinatrice :	Mme Aline DE PAULA REIS	Maître de Conférences à l'EnvA

Liste des enseignants intervenant dans l'encadrement des thèses de Doctorat vétérinaire



Version juillet 2023

Liste des Professeurs et Maîtres de conférences titulaires de l'HDR

M	Adjou	Karim	Professeur	DPASP
M	Audigié	Fabrice	Professeur	DEPEC
M	Bellier	Sylvain	Professeur	DSBP
M	Blaga	Radu	Professeur	DSBP
M	Blot	Stéphane	Professeur	DEPEC
M	Boulouis	Henri-Jean	Professeur émérite	DSBP
Mme	Chahory	Sabine	Professeur	DEPEC
M	Chateau	Henry	Professeur	DSBP
Mme	Chetboul	Valerie	Professeur	DEPEC
Mme	Crevier-Denoix	Nathalie	Professeur	DSBP
M	Degueurce	Christophe	Professeur	DSBP
M	Denoix	Jean-Marie	Professeur	DEPEC
M	Desquilbet	Loic	Professeur	DSBP
Mme	Dufour	Barbara	Professeur émérite	DPASP
M	Eloit	Marc	Professeur	DSBP
M	Fayolle	Pascal	Professeur émérite	DEPEC
M	Federighi	Michel	Professeur	DPASP
M	Fontbonne	Alain	Professeur	DEPEC
Mme	Gilbert	Caroline	Professeur	DSBP
M	Grandjean	Dominique	Professeur	DEPEC
Mme	Grimard-Ballif	Bénédicte	Professeur	DPASP
Mme	Haddad-Hoang Xuan	Nadia	Professeur	DPASP
M	Jouvion	Gregory	Professeur	DSBP
M	Kohlhauer	Matthias	Professeur	DSBP
Mme	Le Poder	Sophie	Professeur	DSBP
Mme	Le Roux	Delphine	Maître de conférences HDR	DSBP
M	Manassero	Mathieu	Professeur	DEPEC
Mme	Maurey-Guénec	Christelle	Professeur	DEPEC
M	Millemann	Yves	Professeur	DPASP
Mme	Pilot-Storck	Fanny	Professeur	DSBP
M	Ponter	Andrew	Professeur	DPASP
Mme	Risco-Castillo	Véronica	Maître de conférences HDR	DSBP
Mme	Rivière	Julie	Maître de conférences HDR	DPASP
Mme	Robert	Céline	Professeur	DSBP
M	Tiret	Laurent	Professeur	DSBP
M	Tissier	Renaud	Professeur	DSBP
M	Verwaerde	Patrick	Professeur	DEPEC
Mme	Viateau	Véronique	Professeur	DEPEC

Liste des Maîtres de conférences et Ingénieurs de recherche DMV

M	Arné	Pascal	Maître de conférences	DPASP
Mme	Barassin	Isabelle	Maître de conférences	DPASP
Mme	Barbarino	Alix	Ingénieur de recherche (DMV)	DEPEC
Mme	Benckekroun	Ghita	Maître de conférences	DEPEC
Mme	Bertoni	Lelia	Maître de conférences	DEPEC
M	Bolnot	Francois	Maître de conférences	DPASP
Mme	Canonne-Guibert	Morgane	Maître de conférences	DEPEC
Mme	Chevallier	Lucie	Maître de conférences	DSBP
Mme	Cochet-Faivre	Noëlle	Maître de conférences	DEPEC
Mme	Cordonnier-Lefort	Nathalie	Maître de conférences	DSBP
Mme	Coudry	Virginie	Ingénieur de recherche (DMV)	DEPEC
Mme	Crepeaux	Guillemette	Maître de conférences	DSBP
M	Crozet	Guillaume	Maître de conférences	DPASP
Mme	De Paula Reis	Alline	Maître de conférences	DPASP
M	Delsart	Maxime	Maître de conférences	DPASP
Mme	Denis	Marine	Ingénieur de recherche (DMV)	DPASP
M	Deshuillers	Pierre	Maître de conférences	DSBP
M	Gauthier	Michel	Maître de conférences associé	DPASP
Mme	Guérin	Virginie	Maître de conférences	DSBP
Mme	Guélin-Poirier	Valentine	Maître de conférences	DPASP
Mme	Jacquet	Sandrine	Ingénieur de recherche (DMV)	DEPEC
M	Kurtz	Maxime	Ingénieur de recherche (DMV)	DEPEC
Mme	Lagrée	Anne-Claire	Maître de conférences	DSBP
Mme	Le Dudal	Marine	Ingénieur de recherche (DMV)	DSBP
Mme	Legrand	Chantal	Maître de conférences associé	DSBP
M	Mammeri	Mohamed	Maître de conférences	DSBP
Mme	Manguin	Estelle	Maître de conférences	DEPEC
Mme	Marignac	Genevieve	Maître de conférences	DSBP
Mme	Marotto	Stéphanie	Ingénieur de recherche (DMV)	DEPEC
Mme	Maurice	Emeline	Ingénieur de recherche (DMV)	DEPEC
Mme	Mespoules-Rivière	Céline	Ingénieur de recherche (DMV)	DEPEC
Mme	Mimmet	Narjès	Maître de conférences	DPASP
M	Mortier	Jérémy	Maître de conférences associé	DEPEC
M	Nudalman	Nicolas	Maître de conférences	DEPEC
M	Pignon	Charly	Ingénieur de recherche (DMV)	DEPEC
M	Polack	Bruno	Maître de conférences	DSBP
Mme	Quéré	Émilie	Ingénieur de recherche (DMV)	DEPEC
Mme	Ravary-Plumioën	Bérandère	Ingénieur de recherche (DMV)	DPASP
M	Reyes-Gomez	Edouard	Maître de conférences	DSBP
Mme	Rosa	Hélène	Maître de conférences associé	DSBP
M	Tanquerel	Ludovic	Maître de conférences	DEPEC

REMERCIEMENTS

Au Président du Jury de cette thèse, M Henry Chateau Professeur à l'EnvA,

Qui m'a fait l'honneur d'accepter la présidence de mon jury de thèse

Hommage respectueux

À Mme Legrand, Maitre de conférences à l'EnvA,

Qui m'a accompagné dans la rédaction de cette thèse avec patience et enthousiasme

Pour tous les échanges, les corrections et le temps dédié à cette thèse

Pour l'accompagnement dans la réalisation de mon projet professionnel

Sincères remerciements

À Mme De Paula Reis, Maitre de conférences à l'EnvA,

Qui a accepté d'être assesseur de cette thèse

Pour votre encadrement bienveillant et compréhensif durant ma scolarité

Pour votre écoute et votre réactivité dans le cadre de mes projets personnels

Sincères remerciements

TABLE DES MATIÈRES

REMERCIEMENTS.....	5
TABLE DES MATIÈRES.....	6
TABLE DES ILLUSTRATIONS.....	8
LISTE DES ABRÉVIATIONS.....	9
INTRODUCTION.....	10
Partie 1 : Définition de l'intelligence artificielle et de ses prérequis.....	11
1. Qu'est-ce que la data ?	12
A. Définition du concept de data et de data science	12
B. Différentes classifications de la donnée.....	12
C. Différentes sources de données	13
2. Le traitement des données	14
A. Caractéristiques des données Big Data.....	14
B. Collecte et stockage des données	15
C. Analyse des données.....	16
3. Qu'est-ce que l'Intelligence Artificielle (IA) ?.....	18
A. Définition du concept.....	18
B. Bref historique de l'Intelligence Artificielle.....	18
C. Qu'est-ce que le Machine Learning ?	20
D. Les problèmes posés par les algorithmes de Machine Learning.....	22
a. Qu'est-ce qu'un algorithme de Deep Learning ?.....	23
b. Le problème de l'opacité des modèles.....	24
4. Importance de la data dans la médecine de demain.....	25
A. Défis posés au système de santé actuel	25
B. Les solutions potentielles apportées par les technologies basées sur la data et l'IA aux défis posés aux systèmes de santé.	25
C. Médecine des « 4P » ou des « 6P »	27
Partie 2 : Utilisation des données et de l'intelligence artificielle dans la gestion des maladies chroniques en médecine humaine	29
1. Prérequis à l'utilisation des données en santé.....	30
A. Que sont les données de santé ?	30
B. Flots de données au sein des écosystèmes de santé	30
a. Qui produit les données de santé ?	30
b. Comment sont stockées les données de santé ?.....	31
c. Qui utilisent les données de santé ?	32
C. Contrôle de la qualité des données de santé	33
D. La formation des médecins à l'Intelligence Artificielle	34
2. Les applications de l'Intelligence Artificielle dans la gestion des maladies chroniques en médecine humaine	35
A. Contexte de l'étude	35
B. L'apport de l'intelligence artificielle dans le monitoring et la prévention des maladies chroniques..	36
a. L'apport de l'IA dans l'intégration des facteurs génétiques dans la prédiction des risques d'apparition des maladies chroniques.....	36
b. L'apport de l'IA dans la prédiction de l'apparition des maladies chroniques	37
c. L'apport de l'IA dans la création d'outils de sensibilisation.....	38
C. L'apport de l'IA dans le diagnostic des maladies chroniques	39
a. L'apport de l'IA pour le diagnostic précoce de maladies chroniques.....	39
b. L'IA permettant d'améliorer la fiabilité des outils dans le diagnostic de maladies chroniques	40

c. L'IA permettant de faciliter la mise en place d'un environnement favorisant le diagnostic des maladies chroniques.....	42
D. <i>L'apport de l'IA dans le traitement et le suivi long-terme des maladies chroniques</i>	43
a. La chirurgie assistée par l'IA dans la gestion des maladies chroniques.....	43
b. L'utilisation de l'Intelligence Artificielle dans la personnalisation des traitements des maladies chroniques	43
c. L'utilisation de l'IA pour évaluer l'évolution de la maladie et remonter les alertes.....	45
3. Réglementations et normes autour de l'Intelligence Artificielle en médecine humaine.....	46
A. <i>Réglementations en vigueur quant à l'utilisation de l'IA en médecine</i>	46
B. <i>Questionnements éthiques autour de l'utilisation des données de santé en médecine</i>	48
Partie 3 : Situation actuelle du développement de l'Intelligence Artificielle en santé animale et axes de développement	52
1. Etat de l'art de l'utilisation de l'IA en santé animale appliquée aux maladies chroniques chez les animaux de compagnie	53
A. <i>Utilisation de l'IA dans la prédiction de maladies chroniques chez les animaux de compagnie</i>	53
B. <i>Utilisation de l'IA dans le diagnostic de maladies chroniques chez les animaux de compagnie</i>	54
C. <i>Utilisation de l'IA pour le traitement de maladies chroniques chez les animaux de compagnie</i>	55
2. Les solutions disponibles embarquant de l'IA pour les vétérinaires dans la gestion des maladies chroniques des animaux de compagnie.....	56
A. <i>Outils disponibles dans le diagnostic des maladies chroniques des animaux de compagnie</i>	56
B. <i>Outils non spécifiques utilisant l'IA pour aider les praticiens dans leur pratique quotidienne</i>	57
3. Quelles applications semblent être les plus prometteuses et les plus applicables en médecine vétérinaire ?.....	58
4. Créer des modèles applicables pour les vétérinaires et diffusion au sein de la profession ...	61
A. <i>Développer des modèles d'IA : une histoire d'équipe et de données</i>	61
B. <i>Sensibiliser les vétérinaires à l'utilisation et au potentiel de l'IA</i>	63
C. <i>Proposer des solutions à valeur ajoutée pour les vétérinaires</i>	64
CONCLUSION	67
BIBLIOGRAPHIE	68

TABLE DES ILLUSTRATIONS

Figure 1 : Illustration du concept de base de données et d'entrepôt de données (WIKIVERSITE, 2017)

Figure 2 : Ecran issu du logiciel NVivo avec une carte des noeuds issus d'un entretien (Lebraty *et al.*, 2018)

Figure 3 : Illustration d'un réseau Bayésien (Soni, 2019)

Figure 4 : Développement des diverses disciplines de l'IA (Ertel, 2017)

Figure 5 : Illustration des différents domaines de l'Intelligence Artificielle (Berkeley Law, 2020)

Figure 6 : Illustration de la structure d'un réseau de neurones (Oppermann, 2021)

Figure 7 : Impact des différents axes de travail afin de repenser les systèmes de santé des pays développés (The Health Foundation, 2015)

Figure 8 : L'ensemble des données de santé présentes sur l'ENS (Fallery, 2020)

LISTE DES ABRÉVIATIONS

API : Application Programme Interface

AUROC : Area Under the Curve ROC

CAQDAS : Computer-Assisted Qualitative Data Analysis Software

CEPS : Comité Economique des Produits de Santé

CNEDiMETS : Commission Nationale d'Evaluation des Dispositifs Médicaux et Technologie de Santé

CNIL : Commission Nationale de l'Informatique et des Libertés

CPH : Cox Proportional Hazard

DBMS : DataBase Management System

DREES : Direction des Etudes, de l'Evaluation et des Statistiques

ENS : Espace Numérique de Santé

IA : Intelligence Artificielle

IMC : Indice de Masse Corporel

Inserm : Institut National de la Santé et de la Recherche Médicale

IRM : Image par Résonance Magnétique

JSON : JavaScript Object Notation

KOL : Key Opinion Leader

ML : Machine Learning

MRC : Maladie Rénale Chronique

NoSQL : Not only Structured Query Language

OCDE : Organisation de Coopération et de Développement Economique

OMS : Organisation Mondiale de la Santé

POC : Proof of Concept

SNVEL : Syndicat National des Vétérinaires d'Exercice Libéral

STAR : Smart Tissue Autonomous Robot

RGPD : Règlement Européen sur la Protection des Données Personnelles

SNDS : Système National des Données de Santé

SQL : Structured Query Language

XML : Extensible Markup Language

INTRODUCTION

Les données (aussi appelées data) sont considérées comme le nouvel or noir du 21^{ème} siècle. Elles constituent non seulement un enjeu majeur de développement pour les acteurs économiques, un enjeu géostratégique pour les acteurs géopolitiques, mais ont aussi une opportunité de révolutionner le mode de vie des particuliers. Le secteur de la santé n'est pas à l'écart de ces transformations. Que ce soit le partage des données entre deux professionnels lors des prises en charge ou leur exploitation pour créer des outils numériques utilisables par les structures de soins, les données deviennent un élément clé dans une nouvelle conception de la médecine. L'Intelligence Artificielle (IA) repose sur l'exploitation des données produites par l'ensemble des parties prenantes (liées ou non au secteur d'exploitation des données) et permet d'effectuer des tâches complexes, hors de portée des outils numériques traditionnels. L'IA peut modifier la façon dont les entreprises fonctionnent et il est plus que probable qu'elle impactera le mode de fonctionnement des cliniques vétérinaires.

Les évolutions numériques se font de plus en plus rapidement et sont présentes dans la majorité des secteurs d'activité. Pourtant, en médecine vétérinaire et plus particulièrement en pratique urbaine, les outils numériques à base d'IA sont encore peu nombreux. L'objectif de cette thèse n'est pas de dresser un état des lieux de la diffusion de l'IA au sein de la profession vétérinaire, celui-ci étant amené à changer profondément et rapidement. En revanche, il s'agit de constater le potentiel et surtout les limites à l'adoption de ces nouvelles technologies. En médecine humaine, le sujet est largement plus relayé et exploité. On retrouve des outils utilisant l'IA dans des domaines aussi variés que la chirurgie, la médecine sportive ou encore l'administration médicale. Comment expliquer ce retard pris par la médecine vétérinaire ? Et comment expliquer que même en pratique vétérinaire rurale, même si l'IA est plus présente, ce ne soient pas les vétérinaires qui se sont emparés du sujet ?

Le milieu vétérinaire est un milieu beaucoup plus petit que le milieu de la médecine humaine. C'est également un milieu en pleine mutation, particulièrement en France. Nous pouvons par exemple penser à l'émergence de groupes de cliniques (consolidation des structures de soins), qui ont connu une croissance exponentielle ces dernières années en médecine des animaux de compagnie. Nous pouvons aussi penser à la diversification des écoles de formation des vétérinaires, avec une traction forte des écoles situées à l'étranger. Nous observons également une croissance des assurances animales qui pourraient modifier la gestion de la santé des animaux sur le long terme. C'est également un milieu concentré. Nous pouvons prendre comme exemple, la concentration des logiciels de gestion de cliniques vétérinaires où 4 acteurs se partagent près de 70% du marché mais aussi le faible nombre d'hôpitaux (on en compte une dizaine en France). On comprend donc que certains acteurs ont un rôle prédominant dans la diffusion de pratiques et d'idées, à l'image du Conseil de l'Ordre des Vétérinaires ou encore du Syndicat National des Vétérinaires d'Exercice Libéral (SNVEL).

Pour ce travail, nous nous intéresserons à la pratique vétérinaire en milieu urbain avec une attention particulière pour les pratiques canines et félines. De plus, nous aurons une attention particulière pour la gestion des maladies chroniques. Nous faisons ce choix car la gestion des maladies chroniques représente l'activité de base des cliniques vétérinaires, que ce soit au travers de la prévention (vaccination par exemple), de la prédiction (visites annuelles chez le vétérinaire), du diagnostic ou du traitement et du suivi. Toutes les structures sont concernées. De plus, nous faisons le choix de nous intéresser à la pratique urbaine car c'est celle dont les problématiques et le fonctionnement sont les plus proches de la pratique humaine, comparaison que nous utiliserons au cours de cette thèse. Le but sera d'établir des comparaisons et des différences afin de mieux saisir les enjeux de l'utilisation de l'IA en santé animale.

Dans une première partie, nous explorerons les notions de données et d'Intelligence Artificielle afin d'en comprendre les spécificités et les prérequis. Nous verrons ensuite les applications des technologies basées sur l'IA en médecine humaine afin de comprendre comment elles sont utilisées dans un secteur plus mature sur le sujet que la médecine vétérinaire. Enfin, nous verrons les enjeux de la diffusion et les difficultés à l'adoption des outils à base d'IA au sein de la pratique vétérinaire.

Partie 1 : Définition de l'intelligence artificielle et de ses prérequis

1. Qu'est-ce que la data ?

A. Définition du concept de data et de data science

Selon le « Cambridge Dictionary », les données (également data) sont « de l'information, en particulier des faits ou des nombres, collectée afin d'être examinée et utilisée pour aider à la prise de décision, ou de l'information dans un format électronique qui peut être stockée et utilisée par un ordinateur ».

De manière plus large, tout ou presque peut être considéré comme de la data. Il peut aussi bien s'agir de l'avis d'un internaute sur le plat qu'il a mangé au restaurant que de l'historique des valeurs de glycémie d'un animal sur les 5 derniers jours. On comprend vite que la quantité de données produite par jour sur la planète est considérable. En 2020 par exemple, on estime qu'environ 1,7Mo (Mégaoctets) de données sont créées chaque seconde par chaque individu selon Oracle (ORACLE, 2022). Cela correspond à la taille d'une chanson au format MP3. De plus, la croissance de la production de données est exponentielle. On estime que 90% des données mondiales ont été produites au cours des deux dernières années.

Toujours selon Oracle, une entreprise spécialisée dans la gestion des données, la data science est une discipline visant à extraire de la valeur économique des données pour l'entité qui les exploite. C'est donc une discipline large qui s'appuie sur des outils mathématiques, statistiques, informatiques, de visualisation, ... L'émergence de la data science est directement issue du développement des bases de données et de l'Internet. En effet, les premiers écrits sur l'analyse de données remontent au début des années 60 et l'article « The Future of Data Analysis » de John Tukey. Il a fallu attendre le milieu des années 90 pour voir l'International Federation of Classification Societies (IFCS) organiser la première conférence avec une mention explicite du terme Data Science mais même alors, la data science n'était encore qu'un concept. Ce n'est qu'à partir des années 2000 et les avancées dans l'exploitation et le stockage de grandes quantités de données que la data science connaîtra sa démocratisation. Viendra ensuite le terme de Big Data (souvent traduit par « données massives ») qui correspond à de la donnée en grande quantité, créée rapidement et avec des formats variés. En effet, vers le milieu des années 2000, la création de données en quantité très importantes des utilisateurs de Facebook, Youtube et autres plateformes en ligne rendra ce terme populaire. Des outils permettant d'analyser ces larges volumes de données comme Hadoop ou Spark ont permis de travailler plus facilement avec des bases de données très importantes et surtout de les stocker à moindre coût.

Aujourd'hui, ce ne sont plus seulement les humains qui produisent ces données. Les objets connectés permettent de créer des données sur le comportement de leur utilisateur. Les algorithmes de Machine Learning produisent également de la donnée.

B. Différentes classifications de la donnée

Si nous produisons de plus en plus de données, la diversité d'origine et de structure de cette donnée s'accroît également.

Selon Charles Perez et Karina Sokolova dans leur livre « Un monde de Données : Initiation sans prérequis au domaine de la Donnée », nous pouvons comme suit catégoriser les données (Perez et Sokolova, 2020):

- Les données peuvent être structurées, non structurées ou semi-structurées :
 - o Les données structurées sont des données qui ont été formatées selon une structure précise, elles sont « organisées de sorte qu'il est facile d'identifier des couples clé/valeur ». C'est le type de données que l'on trouve traditionnellement dans des fichiers Excel. Par exemple, la clé serait la catégorie « Prénom » tandis que les valeurs seraient Chantal, Henry, Aline, ... On trouve ce type de données dans ce qu'on appelle des bases de données relationnelles. Ces données peuvent être représentées selon un format tabulaire avec des lignes et des colonnes ;

- Les données non structurées sont des données pour lesquelles on ne peut pas trouver de couple clé/valeur évident. Ce sont des données qui peuvent avoir des formats très divers. Les notes manuscrites, des posts sur des réseaux sociaux ou des données de capteurs IoT sont des exemples de données non structurées. Ces données non structurées ne peuvent pas être traitées par des outils « traditionnels » sans un traitement préalable spécifique. Ces données sont stockées dans des bases de données non-relationnelles (NoSQL : Not only SQL). Ce type de données est le type de données le plus courant et peuvent être porteuses d'informations précieuses pour les entreprises (avis sur Internet par exemple). A partir de données non-structurées, on peut obtenir des données structurées. Par exemple, on peut faire une analyse de l'occurrence de certains mots dans un message et extrapoler des données numériques ;
 - Les données semi-structurées sont des données possédant une partie structurée et une partie non-structurée. Par exemple, les mails sont des données semi-structurées dans la mesure où le contenu est non-structuré mais certaines données comme la date d'envoi, le destinataire, etc. sont des données structurées.
- Parmi les données structurées, on peut avoir un autre niveau de catégorisation qui correspond au type de variable :
- Les données nominales sont des données « textuelles » et ne prennent pas d'ordre précis. De plus, elles ne sont pas limitées à un ensemble précis. Les noms des produits commercialisés par une entreprise sont des données nominales par exemple ;
 - Les données ordinales sont également des données « textuelles » mais ne peuvent prendre qu'un certain nombre de valeur et possèdent un ordre. Par exemple, les tailles de vêtement sont des données ordinales (XS, S, M, L, XL) ;
 - Les données numériques sont des données non « textuelles » et peuvent avoir comme valeur tous les nombres réels.

C. Différentes sources de données

De la même façon que les données possèdent des structures variables, les fichiers contenant ces données peuvent être de nature variable également.

Un type de fichier courant sont les fichiers-texte. Dans ce type de fichier, les données sont stockées sous forme de texte et chaque valeur est séparée des autres par un délimiteur. Un délimiteur est une séquence d'un caractère ou plus qui sert à spécifier la frontière entre des catégories ou des valeurs. Ce délimiteur peut être une virgule (fichiers CSV), un espace (fichiers TSV), une barre verticale, ... Ce type de fichiers peuvent être traités par un grand nombre d'applications. Parmi d'autres types de fichiers courants, on peut retrouver les classeurs comme Microsoft Excel ou Google Sheet qui ont l'avantage de pouvoir être traités par un grand nombre d'applications et qui ne peuvent pas sauvegarder du code malicieux. D'autres formats sont tout aussi courants mais moins connus du grand public comme les fichiers JSON (JavaScript Object Notation) qui permettent de transférer des données structurées sur Internet ou encore les fichiers XML (eXtensible Markup Language) qui ressemblent aux fichiers texte.

Ces différents formats de fichiers proviennent de sources différentes. Les entreprises possèdent leurs propres applications internes pour maintenir les bases de données utilisées régulièrement. De tels systèmes nécessitent généralement de stocker les données sous des formats structurés et les bases de données

relationnelles sont utilisées dans ce cas. Les bases de données sont donc souvent des classeurs, des fichiers textes ou des fichiers XML.

Les sites Internet peuvent également être des sources de données précieuses. Ces sites fournissent des APIs (Applications Program Interfaces, c'est-à-dire des interfaces entre le serveur Web et une application qui définissent les règles selon lesquelles le serveur et les applications communiquent) qui permettent de transmettre de la donnée à analyser ou à traiter. Par exemple, il existe des APIs qui permettent d'extraire de la donnée depuis des Tweets ou des posts Facebook et ensuite de l'analyser (pour des sondages par exemple). Cela fait partie de ce que l'on appelle le « Web scraping » qui est le terme employé pour définir l'extraction de données depuis un site Internet.

Les flux de données sont aussi des sources de bases de données importantes. Ces flux de données peuvent provenir d'applications, d'objets connectés, de GPS, de programmes informatiques, ... et ils sont une des sources de données les plus prolifiques. Dans ces cas, les données sont fournies avec une temporalité ainsi qu'avec une géolocalisation. Pour exploiter ce type de sources, il faut des programmes spécialisés appelés lecteurs de flux qui permettent d'enregistrer et d'ordonner toutes les données depuis une source.

Il existe de nombreuses autres sources de données mais celles qui ont été présentées précédemment font partie des plus couramment exploitées dans le domaine de la santé.

2. Le traitement des données

A. Caractéristiques des données Big Data

Il n'y a pas de définition universelle du Big Data, cependant, la plupart des auteurs s'accordent pour attribuer aux données Big Data un certain nombre de caractéristiques qui différencient les données Big Data des données traditionnelles : les « 3V ».

- Volume : Il s'agit de la quantité de données qui est générée par les entreprises et les personnes. Le Big Data produit de gros volumes de données, non-structurées provenant de nombreuses sources différentes. Certaines entreprises doivent par exemple entretenir des bases de données pouvant s'élever jusqu'à des centaines de pétaoctets ($1 \text{ Po} = 10^{15}$ octets) ;
- Vitesse : Il s'agit de la fréquence à laquelle les données sont générées et stockées. Les entreprises ne peuvent compter sur ces données que si elles sont traitées aussi rapidement qu'elles sont générées. Avec les évolutions technologiques récentes, la vitesse de génération des données est de plus en plus élevée. La prise en compte de la vitesse de traitement des données est primordiale afin d'obtenir de la valeur des données ;
- Variété : Il s'agit de la diversité des sources de données mais aussi du type de données générées. Comme nous l'avons vu précédemment, il existe de nombreuses sources de données et de nombreux types de données qui ne seront ni stockées ni analysées de la même façon. Particulièrement dans le cas du Big Data, les données générées ne sont pas nécessairement structurées et nécessitent d'être traitées préalablement à leur analyse. (ORACLE FRANCE, 2023)

A ces 3 V peuvent s'ajouter d'autres caractéristiques (« V » également) :

- Véracité : Il s'agit de l'incertitude autour des données c'est-à-dire que la qualité et l'exactitude des données sont aléatoires. Dans le cas du Big Data, du fait de la multitude de sources notamment, la véracité des données devient difficile à évaluer et devient un enjeu majeur pour les utilisateurs des données ;
- Valeur : C'est l'analyse des données qui produit de la valeur. Il est parfois considéré que les données sont le pétrole du XXIème siècle. Cependant, les données n'ont pas toutes les mêmes valeurs et c'est

leur analyse qui leur donne leur valeur et selon Oracle, elles ne sont d'aucune utilité tant que leur valeur n'est pas découverte. Tous les V cités précédemment prennent leur sens dans le V de valeur. Si les données ne peuvent pas être traitées car elles sont trop diverses, pas à jour, en volume trop important pour être traitées ou qu'elles ne sont pas vraies, elles n'ont plus de valeur.

B. Collecte et stockage des données

Comme nous avons pu le voir, les données ont besoin d'être analysées afin d'avoir une valeur. Ainsi, avant de se lancer dans l'analyse de données, il faut être en mesure de savoir à quelles questions l'on cherche à répondre. Une fois que les objectifs de l'étude sont clairs, on peut appliquer une procédure de collecte des données :

- 1) Déterminer les informations que l'on veut collecter en fonction de l'objectif de l'étude : cette décision se fait en fonction des objectifs que l'on a mais aussi à partir des moyens et des sources dont nous disposons ;
- 2) Définir un plan de collecte des données : il est nécessaire d'établir la période sur laquelle on veut recueillir des données, quelle est la population de données cibles, ... C'est lors de cette étape qu'il est nécessaire de déterminer la quantité de données qui seront nécessaires afin de réaliser une analyse aux résultats crédibles ;
- 3) Identifier les méthodes de collecte des données qui seront mises en place : il est nécessaire de définir comment nous procéderons pour extraire des données des sources qui ont été identifiées. Ces méthodes dépendent du type de données voulues, de la période analysée souhaitée et du volume de données nécessaires.

Tout au long de ce processus, nous devons considérer trois caractéristiques relatives aux données que l'on souhaite collecter et aux méthodes utilisées :

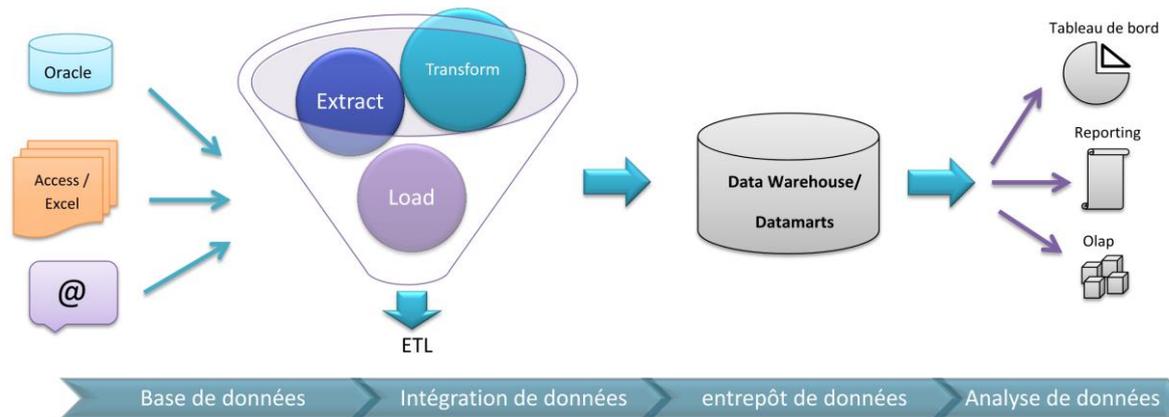
- Qualité des données : les données doivent être exemptes d'erreurs, elles doivent être complètes, précises, accessibles et pertinentes par rapport à l'objectif. Il peut être nécessaire de définir des tests de qualité ;
- Sécurité : les réglementations en termes de gouvernance des données et les procédures relatives à l'utilisation, l'intégrité et la disponibilité des données doivent être respectées. Le non-respect de ces réglementations peut avoir des conséquences graves pour l'organisation fautive ;
- Confidentialité : Les données collectées doivent pouvoir être accessibles en toute légalité sans enfreindre des réglementations de propriété privée.

Une fois que les données ont été collectées et organisées, celles-ci doivent être stockées dans des dépôts de données. On compte différents types de dépôts de données :

- Les bases de données : collection de données conçue pour la saisie, le stockage, la recherche, la récupération et la modification des données. Pour entretenir ces bases de données, on utilise un Système de Gestion des Bases de Données (DBMS : DataBase Management System). Il existe différents types de bases de données et leurs caractéristiques motivent l'utilisation de l'une ou l'autre des bases de données :
 - o Les bases de données relationnelles qui sont optimisées pour l'organisation de données structurées et qui permettent d'effectuer des opérations qui concernent de très grandes tables de données. Le langage utilisé pour utiliser ces bases de données est SQL (Structured Query Language) ;

- Les bases de données non-relationnelles (NoSQL) qui permettent de stocker de la donnée non-structurée. C'est le type de bases de données le plus utilisé pour l'analyse de données Big Data.
- Entrepôt de données (Data Warehouses) : il s'agit d'un dépôt central de données qui fusionnent des données provenant de sources variées et les rendent utilisables pour des analyses diverses. Il existe des dépôts de données qui dérivent des entrepôts de données comme les « Data Marts » ou les « Data Lakes ».

Figure 1 : Illustration du concept de base de données et d'entrepôt de données (WIKIVERSITE, 2017)



C. Analyse des données

Selon João Moreira, Andre Carvalho et Tomás Horvath dans leur livre "A General Introduction to Data Analytics", on peut définir l'analyse de données comme la science qui analyse les données brutes afin d'en extraire de la connaissance utile (Moreira *et al.*, 2018). Au sein de l'analyse de données, il faut différencier différentes sous-catégories :

- L'analyse statistique composée de l'analyse descriptive et de l'analyse inférentielle. L'analyse descriptive consiste à analyser les données et synthétiser les informations d'un échantillon de données. L'objectif est de présenter les données de façon à les rendre plus compréhensibles. Cette analyse utilise essentiellement des outils mathématiques de statistiques. Cependant, l'analyse descriptive est différente de l'analyse inférentielle dont le but est d'extrapoler les informations issues d'un échantillon et de le généraliser à une population plus large. Cette analyse inférentielle se base sur des tests d'hypothèse, des intervalles de confiance, des régressions, ... ;
- L'analyse diagnostique qui consiste à comprendre les causes de faits découverts grâce à l'analyse statistique. Ce type d'analyse permet d'identifier des comportements au sein de la donnée et a pour objectif de résoudre des problèmes similaires à l'aide de données fournies ;
- L'analyse prédictive qui consiste à étudier la probabilité de survenue des événements en se basant sur les informations du passé. Ce type d'analyse est particulièrement sensible à la qualité et à la quantité des données exploitées ;
- L'analyse prescriptive qui consiste à étudier les données afin de mettre en place une stratégie afin de résoudre un problème posé.

Dans l'industrie de la santé, ces quatre types d'analyse sont utilisées. Pour l'analyse statistique, on peut penser à l'étude de contagiosité d'un virus en examinant le taux de positivité à un test dans une population durant une période spécifique. En ce qui concerne l'analyse diagnostique, on peut chercher à déterminer les causes d'une pathologie héréditaire en étudiant l'implication de certains gènes. L'analyse prédictive peut trouver des applications en épidémiologie. Nous pouvons penser à la propagation d'une maladie en fonction des propagations passées de cette même maladie. Enfin, l'analyse prescriptive peut s'appliquer dans le domaine de la prévention de l'apparition de maladies en fonction des facteurs de risque que présentent un patient. (Cote, 2021)

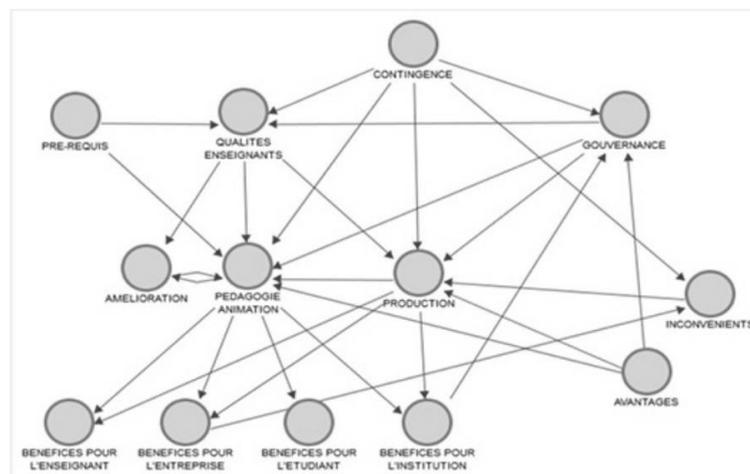
Ce dont nous avons parlé jusqu'à présent concerne l'étude de données quantitatives, mais il existe aussi l'analyse de données qualitatives. Les données qualitatives sont des données auxquelles on ne peut pas attribuer une valeur numérique. Elles ne peuvent donc pas être analysées en utilisant des outils statistiques. L'analyse de ces données demandent l'utilisation de logiciels spécialisés que l'on appelle des logiciels d'aide à l'analyse qualitative de données (Computer Assisted Qualitative Data Analysis Software : CAQDAS). Selon Jean-Fabrice Lebraty, Katia Lobre-Lebraty et Stéphane Trebucq dans leur livre « Les méthodes de recherche du DBA, 2018 », ces logiciels permettent de réaliser trois types de tâches :

- Organisation de données qualitatives ;
- Analyse ;
- Restitution.

Ces logiciels permettent de présenter les résultats sous 4 formes :

- Des cartes de concepts qui permettent de relier les concepts entre eux et qui peuvent être représentées sous forme de matrices ou de graphiques de matrices ;

Figure 2 : Ecran issu du logiciel NVivo avec une carte des noeuds issus d'un entretien (Lebraty et al., 2018)



- Des données liées aux documents sources qui permettent par exemple de montrer le poids de certains concepts au cours de l'analyse des données ;
- Des « arbres » qui permettent de lier des mots et des concepts entre eux ;
- Des nuages de mots.

Ce type d'analyse peut également trouver des applications dans le domaine de la santé. On peut par exemple penser à l'analyse documentaire sur des pathologies qui peut être réalisée à l'aide de logiciels d'intelligence artificiel. (Lebraty et al., 2018)

3. Qu'est-ce que l'Intelligence Artificielle (IA) ?

A. Définition du concept

Il n'y a pas de définition universelle du concept d'intelligence artificielle. La question de l'intelligence artificielle est une question complexe dans la mesure où définir l'intelligence n'est pas une chose aisée en soi. Le dictionnaire Larousse propose lui-même 5 définitions de l'intelligence. De plus, la mesure de l'intelligence n'est pas universelle non plus. Peut-on réduire l'intelligence d'une personne au niveau de son Coefficient Intellectuel ?

En 1955, John McCarthy fut le premier à donner une définition de l'intelligence artificielle. Selon lui, « le but de l'IA est de développer des machines qui se comportent comme si elles étaient intelligentes. ». Le psychologue Valentin Braitenberg pense que cette définition n'est pas complète parce que l'on pourrait faire reproduire à des objets inanimés à l'aide de circuits électriques des comportements qui auraient été décrits comme étant intelligents selon la définition de McCarthy.

Selon l'Encyclopedia Britannica, on peut définir l'IA de la manière suivante : « L'Intelligence Artificielle est la capacité d'ordinateurs numériques ou des robots contrôlés par ordinateur à résoudre des problèmes qui sont normalement associés aux capacités de traitement intellectuel supérieures des humains. ». On pourrait donc en déduire que résoudre des équations mathématiques complexes pourraient rentrer dans le cadre de ces définitions. Aujourd'hui, n'importe quel ordinateur possède la puissance de calcul pour effectuer des calculs qui prendraient beaucoup de temps à un être humain. Ce n'est pas pour autant qu'ils sont qualifiés d'intelligents.

Elaine Rich propose une troisième définition de l'Intelligence Artificielle : « L'intelligence Artificielle est l'étude de la façon dont on pourrait faire en sorte que les ordinateurs fassent des choses pour lesquelles, à l'heure actuelle, les humains sont meilleurs. ». Selon Rich, l'approche du concept d'Intelligence Artificielle est étroitement liée à la façon dont les humains raisonnent et donc à la neuroscience. L'intérêt de cette définition est qu'elle sera toujours valable en 2050 et qu'elle mime la façon dont l'IA s'est développée, à savoir progresser dans des domaines afin de pouvoir battre l'Intelligence Humaine. Cette comparaison avec l'intelligence humaine est reprise par la définition du Larousse qui définit l'IA comme : « Ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine ». Ainsi, l'IA semble se construire par rapport à l'intelligence humaine que ce soit pour l'imiter ou pour la dépasser.

B. Bref historique de l'Intelligence Artificielle

L'histoire de l'Intelligence Artificielle commence dans les années 1930s dans le sillage de précurseurs comme Kurt Gödel, Alonso Church ou encore Alan Turing. Le théorème de la complétude de Gödel tend à prouver que tout problème mathématique du premier ordre pouvant s'écrire avec des variables et dont la preuve s'établit par la valeur des variables peut être résolu à l'aide d'outils de calculs formels. Ainsi, des logiciels permettant de prouver des théorèmes peuvent être construits avec des outils de calculs formels. Le théorème d'incomplétude (de Gödel également) montre que pour des problèmes mathématiques d'ordre supérieur, il n'est pas toujours possible de prouver la véracité de certains énoncés vrais avec des outils formels. Turing a également identifié des limites pour les programmes intelligents. Peu après, dans les années 1940s, McCulloch, Pitts et Hebb ont imaginé le premier modèle mathématique de réseau de neurones mais les ordinateurs de l'époque manquaient de puissance pour simuler ce premier cerveau simplifié.

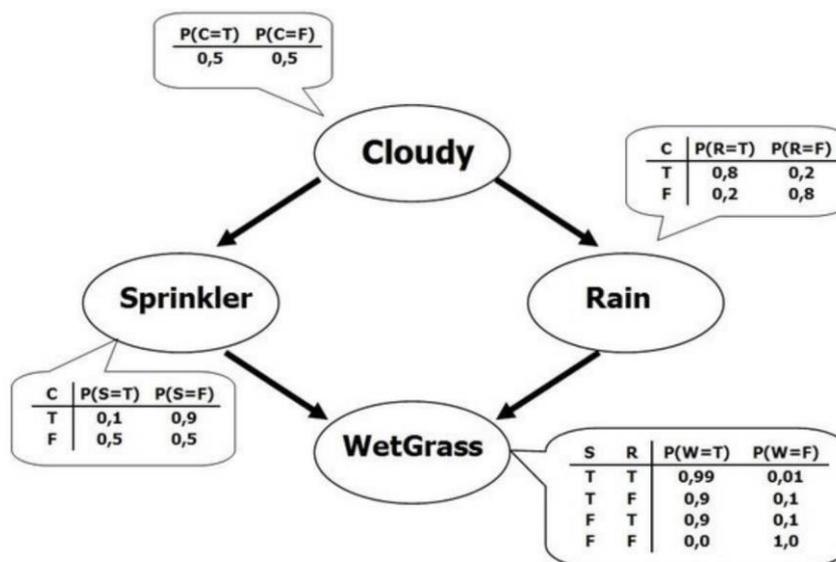
A partir des années 1950s, les ordinateurs sont devenus de plus en plus puissants et surtout sont devenus programmables ce qui a permis l'essor de l'IA. Newell et Simons ont par exemple développé Logic Theorist,

le premier programme de démonstration automatique. McCarthy dont nous avons parlé préalablement a également créé le langage LISP qui a grandement contribué à l'essor de l'IA. Dans les années 1970s, PROLOG a été créé et il correspond au pendant de LISP en Europe mais offrant plus de possibilités notamment en utilisant différents types de données. Jusque dans les années 1980s, la résolution de problèmes simples fonctionnait très bien mais les retombées économiques de son développement ont été jugées trop faibles et cela a conduit à une désillusion temporaire du grand public vis-à-vis de l'IA.

Cependant, c'est également au cours de ces années 1980s qu'est née une sous-discipline de l'IA, l'IA connexionniste. L'IA connexionniste est l'ancêtre du Machine Learning tel qu'il est connu aujourd'hui. L'IA connexionniste est basée sur des ordinateurs suffisamment puissants qui apprennent à partir d'exemples qui leur sont donnés. L'une de leurs forces est d'être capable de reconnaître des schémas et on trouve parmi ses applications la reconnaissance faciale à partir de photos et la reconnaissance d'écriture.

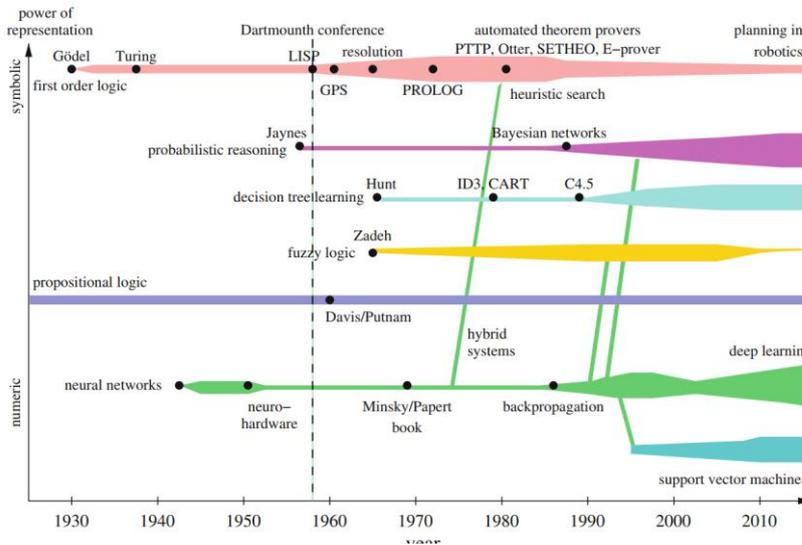
L'étape suivante du développement de l'IA correspond à l'apparition du Raisonnement Probabilistique qui fonctionne avec des probabilités conditionnelles pour résoudre des problèmes. Le Raisonnement Probabilistique s'appuie sur des réseaux Bayesiens. Les réseaux Bayesiens sont des modèles graphiques et probabilistes représentant un ensemble de variables aléatoires sous la forme d'un graphe orienté acyclique selon Wikipédia. Chaque pointe d'un tel graphe représente une variable aléatoire.

Figure 3 : Illustration d'un réseau Bayesien (Soni, 2019)



A partir des raisonnements dans l'incertain rendus possibles grâce aux réseaux Bayesiens, des systèmes comme CART ont pu construire de façon automatique des arbres décisionnels applicables à des données que l'on fournirait au système en entrée. Ces avancées ont été accompagnées de l'avènement du data mining qui correspond à l'extraction de connaissances à partir de base de données très larges. Les applications du data mining sont très nombreuses notamment en médecine. Des techniques de Machine Learning reposant sur des arbres décisionnels permettent de telles applications. Le succès le plus marquant fonctionnant de cette façon est probablement l'ordinateur Deep Blue développé par IBM qui battit le champion d'échec Garry Kasparov en 1997. Il s'agira de la rampe de lancement à la compétition entre intelligences humaines et artificielles dans de nombreux jeux. En 2016, une rencontre extrêmement médiatisée entre AlphaGo appartenant à Google et Lee Sedol, considéré alors comme l'un des meilleurs joueurs au jeu de Go, se termine sur le score de 4-1 en faveur de l'ordinateur. Cela fut considéré comme un événement marquant dans la mesure où le jeu de go était considéré comme un jeu où la puissance de calcul brute ne suffit pas à donner la victoire. (France Culture, 2021)

Figure 4 : Développement des diverses disciplines de l'IA (Ertel, 2017)



Bien évidemment, l'application de l'IA ne se limite pas aux jeux et de nombreuses applications existent dans divers secteurs. En voici quelques exemples (OECD, 2019):

- Dans le secteur des transports avec la transition vers les véhicules autonomes : l'OCDE calcule qu'avec un taux d'adoption de 10% de véhicules autonomes aux Etats-Unis, il y aurait 1100 vies sauvées chaque année et 38 milliards USD économisés par an ;
- Dans le secteur de l'agriculture avec des « robots cueilleurs » avec des systèmes d'IA qui pourraient permettre de répondre au défi posé de l'alimentation mondiale malgré la stagnation des terres cultivables ;
- Dans le secteur de la justice avec l'aide par des logiciels d'IA à mieux établir des connexions, détecter des schémas récurrents, prévenir et résoudre des crimes ;
- Dans le secteur de la santé qui sera l'objet des parties suivantes de cette thèse ;
- ...

Le point commun de ces exemples est l'utilisation de robots autonomes ayant pour but de fournir une réponse à une question grâce à des données fournies en entrée. L'objectif est également que ces robots deviennent de plus en plus performants avec le temps et qu'ils apprennent de leurs erreurs. C'est là qu'intervient le domaine du Machine Learning.

C. Qu'est-ce que le Machine Learning ?

Le Machine Learning est une sous-catégorie de l'IA dont l'objectif est de faire le lien entre des données observables « en entrée » et des variables « en sortie » qui dépendront des données observables. L'idée est de développer un système mathématique qui permettra de déterminer les variables en sortie en fonction des données fournies en entrée. Seulement, l'intérêt du Machine Learning est qu'il permet d'utiliser les capacités d'analyse des ordinateurs pour déterminer la valeur des variables malgré la complexité des relations entre les données observables et les variables. Le Machine Learning permet donc de construire des modèles

suffisamment performants pour résoudre ces problèmes complexes. On parle d'entraînement d'un modèle : le processus de construction du modèle et données d'entraînement les données qui permettent d'entraîner ce modèle. Une fois que le modèle est entraîné, il est capable de « prévoir » la valeur des variables de sortie en fonction des données fournies en entrée. Pour entraîner un modèle, il est nécessaire de posséder un grand nombre de données fiables (valeurs des variables en entrée et en sortie). La fiabilité du modèle dépendra de la quantité et de la qualité des données utilisées pour entraîner le modèle. Il est également nécessaire que ces données soient stockées dans un format susceptible d'être exploité par le logiciel. Les avancées récentes des capacités de stockage, de collecte et de traitement de la donnée sont donc un facteur prédominant à la démocratisation du Machine Learning.

Cependant, pour certains problèmes donnés, il existe un grand nombre de variables qui peuvent être fournies en entrée. Toutes les proposer n'est pas toujours la meilleure solution. En effet, certaines variables peuvent n'exercer qu'une influence minime ou encore être des facteurs de confusion pour l'algorithme. Cela pourrait induire une diminution dans les performances du logiciel. Ainsi, il est laissé au spécialiste d'un sujet, la tâche de sélectionner les données qui devront être utilisées pour entraîner un modèle. Cette tâche est primordiale puisque la collecte et l'entretien de données peut être coûteuse en termes de temps et d'argent mais aussi parce qu'il faut faire attention à ne pas exclure de l'analyse des données pertinentes.

Du fait même du fonctionnement parfois opaque de l'apprentissage du logiciel (comme nous le verrons plus tard), il existe souvent plus d'un modèle entraîné pour résoudre un problème donné. Il est donc fréquent que plusieurs modèles soient entraînés et de décider *a posteriori* lesquels sont les plus adéquats. Il n'y a pas de règles pour définir la qualité d'un algorithme de Machine Learning et cela est souvent laissé à l'appréciation de spécialistes du sujet traité. De plus, la qualité d'un logiciel dépend de plusieurs facteurs comme ses données d'entraînement, les relations entre les variables d'entrée et de sortie ou encore la mémoire du système.

Jusqu'à présent nous avons parlé du Machine Learning comme s'il était uniforme mais ce n'est pas le cas. De la même façon qu'il existe des types de données différents, il existe également des modes d'apprentissage différents pour les logiciels : l'apprentissage supervisé, non-supervisé et par renforcement.

- Apprentissage supervisé : Le principe de l'apprentissage supervisé est celui que nous avons présenté jusqu'à maintenant. Il s'agit de s'appuyer sur des données d'apprentissage préalablement labélisées dans le but de construire un modèle prédictif. Pour ce faire, le modèle essaiera de construire une fonction dite de mapping. De façon mathématique, ce problème revient à déterminer la fonction F telle que $Y = F(X_1; X_2; \dots X_n)$ où X_t seraient les variables d'entrée et Y la variable de sortie. Au sein de l'apprentissage supervisé, la phase d'apprentissage est effectuée dans un contexte de classification ou de régression :
 - o Classification : L'output attendu peut prendre un nombre de valeurs finies et les variables impliquées sont des variables qualitatives. Par exemple, l'objectif d'un algorithme pourrait être de répondre si la photo qui lui est présentée représente un chat ou non. Dans ces cas, les valeurs possibles de la variable de sortie pourraient être Vrai ou Faux,
 - o Régression : Au contraire, l'output attendu peut prendre un nombre infini de valeurs et les variables impliquées sont des variables quantitatives. Il pourrait par exemple s'agir de déterminer le prix d'un appartement en fonction d'un certain nombre de critères comme la localisation, la superficie, le nombre de chambres, ...

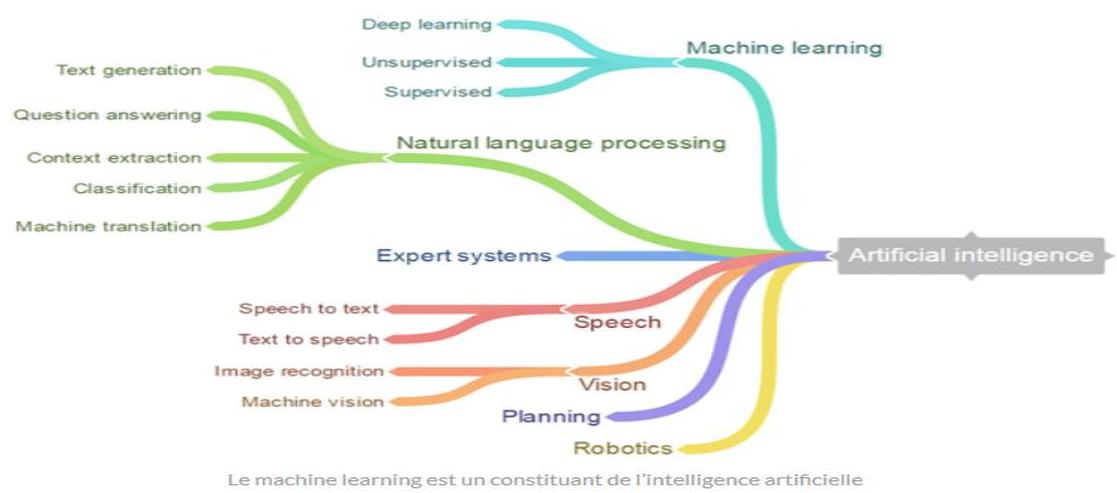
De cette façon, le logiciel apprend en fonction de ce qui lui a été présenté afin de réagir de façon adéquate lorsqu'on lui proposera une situation similaire. Le principe de ce système est que plus il est utilisé et plus le système gagne en performance ;

- Apprentissage non-supervisé : Le principe de l'apprentissage non-supervisé est moins intuitif. Dans ce cas, les données fournies à la machine ne sont pas annotées, contrairement à l'apprentissage supervisé. L'apprentissage par la machine se fera de façon indépendante. L'objectif de l'algorithme est de regrouper les données en fonction de leurs similitudes. Il ne s'agira pas de construire une

fonction de mapping mais d'être capable de regrouper les données en fonction de leurs ressemblances. L'ensemble des données sont traitées comme des variables aléatoires. Il est possible de catégoriser l'utilisation de l'apprentissage non-supervisé en deux types de problèmes : le clustering et l'association :

- Clustering : l'objectif du clustering est d'organiser en groupes les objets présentés sous forme de données parce qu'ils possèdent des caractéristiques homogènes. Cette technique est très utilisée en marketing pour regrouper des profils de consommateurs par exemple afin de leur proposer des publicités ciblées en fonction de leurs goûts, de leur pouvoir d'achat, ... Cela permet de segmenter les clients dans des applications Marketing par exemple.
 - Association : L'objectif de l'association est, comme son nom l'indique, de trier des objets présentés sous forme de données et de trouver des liens d'association entre eux en fonction de leurs caractéristiques. Les objets doivent être liés sans pour autant être identiques, ils n'ont pas forcément les mêmes caractéristiques. Par exemple, si l'on propose à un algorithme des photos de chats avec des photos d'objets dont des accessoires de chat, plutôt qu'associer tous les chats entre eux, l'algorithme associera les accessoires pour chats aux chats.
- Apprentissage par renforcement : Le mode de fonctionnement de l'apprentissage par renforcement est très différent de ce que nous avons vu précédemment. Il s'agit cette fois de placer l'algorithme dans un environnement et de le laisser prendre des décisions en fonction de l'état de l'environnement. En fonction de la transition opérée sur l'environnement par les actions de l'algorithme, celui-ci va recevoir des « récompenses » quantitatives qui peuvent être positives ou négatives. L'objectif de l'algorithme est de maximiser la valeur des récompenses obtenues. L'algorithme va donc apprendre en fonction de ses propres expériences et va chercher à développer un comportement décisionnel optimal (Zimmer, 2018). Par exemple, AlphaGo dont nous avons parlé précédemment a battu Lee Sedol grâce à l'apprentissage par renforcement.

Figure 5 : Illustration des différents domaines de l'Intelligence Artificielle (Berkeley Law, 2020)



D. Les problèmes posés par les algorithmes de Machine Learning

Les algorithmes de Machine Learning ont connu des résultats très encourageants dans de nombreux domaines ces dernières années. Cependant, ce mode d'apprentissage automatique pose des questions quant à l'interprétabilité des résultats. En effet, il est difficile de savoir comment apprend un algorithme car la génération du modèle par l'algorithme est opaque. Ainsi, comment interpréter le résultat si l'on ne sait pas

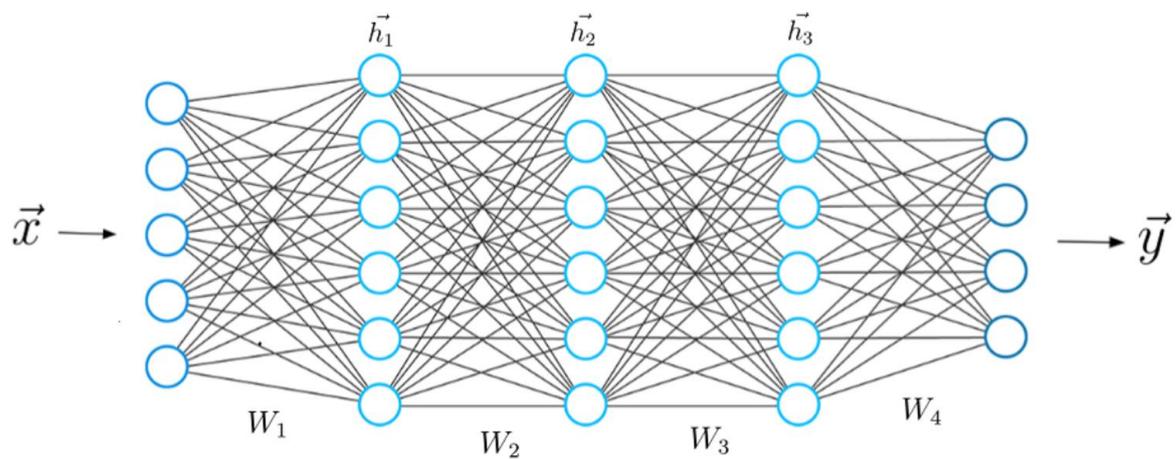
comment l'algorithme est arrivé à cette conclusion ? On parle de l'effet « boîte noire » des algorithmes de Machine Learning et plus particulièrement des algorithmes de Deep Learning.

a. Qu'est-ce qu'un algorithme de Deep Learning ?

Le Deep Learning (qui fait partie du Machine Learning) est un modèle populaire parmi les concepteurs d'algorithmes d'IA. De nombreuses applications et services basés sur le Deep Learning ont permis d'augmenter le niveau d'automatisation et de précision. Par exemple, les systèmes de détection de fraudes aux cartes bancaires reposent sur ce type de technologie. Les algorithmes de Deep Learning possèdent un avantage certain par rapport aux autres algorithmes de Machine learning, ils sont capables de travailler avec des données non-structurées avec peu de traitement préalable. Les algorithmes de Machine Learning classiques peuvent aussi utiliser des données non-structurées mais nécessitent de traiter ces données préalablement. Par exemple, si l'on fournit des photos d'animaux à un algorithme de Deep Learning, celui-ci trouvera automatiquement les éléments qui permettent de déterminer l'animal donné sur la photo. Un algorithme de Machine Learning classique aurait besoin d'un opérateur pour cela.

Le Deep Learning est une branche du Machine learning, c'est-à-dire qu'il s'agit toujours d'apprentissage automatique, mais dont l'objectif est de mimer le cerveau humain. C'est la raison pour laquelle on associe souvent le Deep Learning avec un réseau de neurones. En ajoutant plusieurs « niveaux » de neurones, on parle bien de réseau, on cherche à augmenter la précision et l'efficacité du modèle. Concrètement, les réseaux de neurones profonds consistent en de multiples couches de nœuds interconnectés. Chaque couche utilise les résultats donnés par la couche précédente pour affiner le résultat. La première couche (qui reçoit et ingère les données ou inputs) et la dernière couche (qui transmet le résultat final) sont appelées les couches « visibles ». On peut voir plus simplement ces couches comme étant des filtres qui traitent l'information de la granularité la plus grossière à la granularité la plus fine. Entre chacun des nœuds (ou neurones), il y a une connexion (assimilables aux axones du cerveau) qui possède un certain poids. Au fur et à mesure de l'apprentissage par l'algorithme, le poids entre les différents nœuds se modifient pour refléter l'importance de certains paramètres. Le problème est qu'il est impossible de savoir comment l'algorithme attribue le poids aux différents nœuds ni comment les interactions entre les nœuds donne le résultat final. C'est la raison pour laquelle on parle de « boîte noire ».

Figure 6 : Illustration de la structure d'un réseau de neurones (Oppermann, 2021)



Les logiciels de Deep Learning peuvent effectuer les mêmes tâches que les algorithmes de Machine Learning que nous avons vu précédemment (apprentissage supervisé, non-supervisé et par renforcement) mais l'inverse n'est pas vrai. La plupart des avancées récentes en intelligence sont dues à des algorithmes de Deep Learning.

b. Le problème de l'opacité des modèles

Comme nous l'avons vu précédemment, il existe un phénomène de « boîte noire » lié au fonctionnement des algorithmes de Deep Learning. D'après Jean-Marie John-Mathews de l'université de Paris-Saclay, l'impact de ce phénomène de boîte noire peut se décliner selon les contraintes suivantes :

- Prise en main par un opérateur humain : du fait même de leur utilisation comme outil d'aide à la prise de décision, les algorithmes de Deep Learning peuvent nécessiter l'adhésion d'un opérateur humain. En particulier dans des domaines d'application sensibles comme en médecine, la compréhension de la décision rendue par un algorithme peut être nécessaire avant de mettre en place la solution proposée ;
- Exigence de certification des modèles : cette exigence de certification vient notamment du nouveau cadre légal de la Commission Européenne qui exige que les algorithmes puissent être compréhensibles. Il s'agit notamment de prouver la cohérence d'un modèle par rapport au cahier des charges de son utilisation ;
- Emergence de nouveaux critères de validation des modèles : la performance prédictive d'un modèle n'est plus le seul critère pris en compte pour évaluer celui-ci. D'autres critères, à l'image du critère éthique où juridique, peuvent rentrer désormais en ligne de compte. Il est notamment nécessaire de vérifier l'absence de biais lors de l'apprentissage de l'algorithme qui peut être due à la sélection des données utilisées pour l'apprentissage.

Pour toutes ces raisons, des outils d'interprétation en Machine Learning ont vu le jour et permettent de cartographier les principales méthodes d'interprétation développées en Machine Learning. Il existe de nombreux outils à ce jour comme les méthodes locales, globales, intrinsèques, agnostiques, ... Cependant, aucune méthode ne semble constituer un Gold standard. Les auteurs travaillant sur le sujet manifestent leur volonté de vouloir mixer ces différentes méthodes (John-Mathews, 2019).

4. Importance de la data dans la médecine de demain

A. Défis posés au système de santé actuel

Nous avons vu jusqu'à présent les bases pour comprendre ce que sont les technologies de l'IA, les données sur lesquelles ces technologies se basent ainsi que la complexité technique à prendre en compte pour construire et entraîner ces modèles. Il faut maintenant mettre toutes ces notions dans le contexte médical.

Les systèmes de santé font aujourd'hui face à de nouvelles tendances et à de nouveaux défis qu'ils vont devoir relever. Parmi les tendances qui se dégagent, on retrouve le vieillissement de la population, particulièrement dans les pays développés, qui entraîne une augmentation des dépenses de santé ainsi que du nombre de patients. De plus, nous observons une modification dans l'origine des maladies avec l'émergence de maladies liées au mode de vie de plus en plus sédentaire. Tout cela conduit les pays développés à faire face à une augmentation de la demande en soins, des coûts ainsi que des contraintes (notamment en termes de capacités humaines) pour les systèmes de santé. A titre d'exemple, on estime qu'en 2030, les dépenses liées à la santé atteindront environ 17% du PIB en France. Au rythme du développement actuel, on estime ces dépenses à 27% du PIB en France en 2050 selon un étude McKinsey.

Ainsi, il y aura un défi en termes de quantité de soins à administrer mais également en termes de qualité. En effet, dans les pays de l'OCDE, il est estimé que 10% des patients seraient victimes d'une erreur médicale au cours de leur vie et qu'environ 50 000 personnes meurent chaque année à cause d'erreurs médicales en France d'après Sadek Beloucif, président du Syndicat national des médecins et chirurgiens des hôpitaux publics. Ainsi, avec l'augmentation du nombre de patients et la volonté de plafonner les dépenses de santé, la question de la qualité des soins administrés sera primordiale à l'avenir.

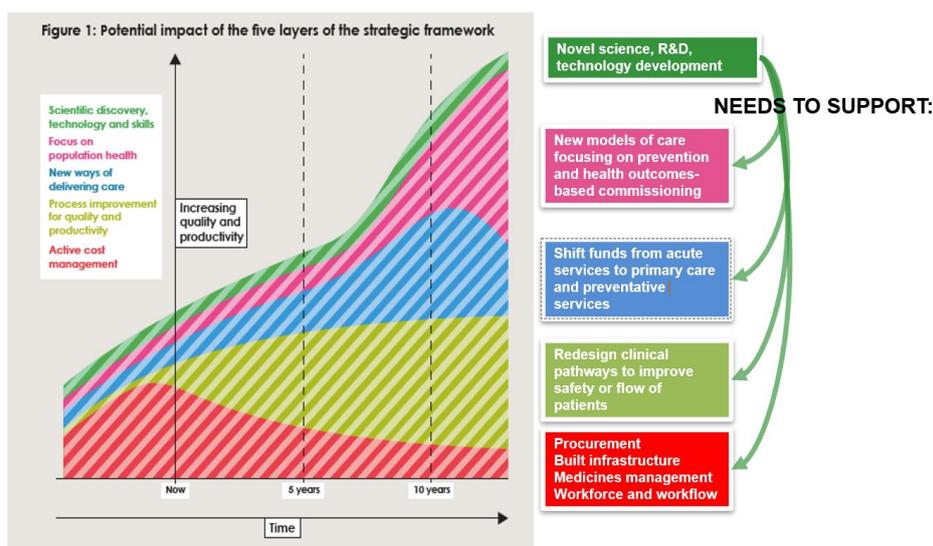
De plus, l'OMS a listé 13 défis liés à la santé pour la décennie à venir parmi lesquels on retrouve le développement des nouvelles technologies. Selon l'OMS, les nouvelles technologies ont la capacité de prévenir, diagnostiquer et de traiter de nombreuses maladies à l'aide de technologies comme l'édition de génome, la biologie synthétique ou encore l'Intelligence Artificielle. L'OMS convient néanmoins que l'utilisation de ces technologies soulèvent des questions quant au suivi et à la régulation de ces solutions (World Health Organization, 2020). Il est important que les implications éthiques et sociales de ces solutions ne fassent pas souffrir ceux qu'elles devraient aider.

B. Les solutions potentielles apportées par les technologies basées sur la data et l'IA aux défis posés aux systèmes de santé.

Selon James Barlow, professeur en Management de l'Innovation à l'Imperial College de Londres, ce sont les systèmes de santé des pays développés dans leur entièreté qui doivent être repensés. Pour cela, l'émergence de technologie comme l'Intelligence Artificielle permettront de supporter ce changement de paradigme profond du fonctionnement des systèmes de santé. Selon lui, les moteurs qui permettront aux systèmes de santé de gagner en efficacité demain ne sont pas les mêmes que durant les décennies précédentes. Pour repenser les systèmes de penser, il faudra développer de nouveaux axes :

- Nouveaux modèles de santé concentrés sur la prévention et sur la médecine de population au sens large ;
- Financement de la médecine préventive au détriment de structures concentrant tous les services comme les hôpitaux ;
- Construire différemment les parcours patients pour en améliorer l'efficacité.

Figure 7 : Impact des différents axes de travail afin de repenser les systèmes de santé des pays développés (The Health Foundation, 2015)



Comme on peut le remarquer sur cette illustration, selon James Barlow, la façon de repenser les systèmes de santé des pays développés passera par la construction de nouveaux parcours patients ainsi que par la création de nouveaux modèles de soins. Il apparaît que les nouvelles technologies telles que l'Intelligence Artificielle permettront de réinventer ces modèles de soins. Ainsi, l'innovation dans le domaine de la santé ne sera pas exclusivement technologique mais la technologie permettra de mettre en place de nouvelles façons d'exercer la médecine.

Selon le cabinet de conseil en stratégie McKinsey, parmi ces nouveaux modèles, nous trouverons que les soins seront guidés par l'utilisation des données. En effet, les patients seront monitorés de façon continue grâce à des objets connectés qui collecteront les données et les transmettront à des logiciels qui analyseront les données et qui pourront donner des indications personnelles au patient. Grâce à la quantité de données recueillies, les systèmes de prévention seront plus efficaces et permettront notamment de réduire les besoins en traitements individuels, ce qui réduira le coût et le temps passé par patient. De cette façon, les patients seront davantage acteurs de leur santé. Dans le domaine vétérinaire, l'utilisation de l'IA en médecine est encore très confidentielle même si certains outils commencent à apparaître. Nous sommes cependant encore loin d'un monitoring à distance des animaux. Nous développerons plus sur l'état actuel de l'avancée de l'IA en Santé Animale dans la troisième partie de cette thèse.

Les technologies de la data pourraient également permettre de répondre à la tension croissante de professionnels du domaine de la santé. Ce type de solution est notamment déjà mis en place en ce qui concerne le triage des patients à leur arrivée aux urgences. Certains sont déjà triés en fonction des objets connectés qu'ils portent déjà sur eux et qui permettent de passer beaucoup moins de temps sur les examens cliniques d'admission. D'autres sont triés grâce à la collecte d'informations pertinentes par des algorithmes de façon autonome sur les signes et les données cliniques.

C. Médecine des « 4P » ou des « 6P »

Nous avons vu que les systèmes de santé sont face à des défis structureaux mais que les technologies basées sur l'utilisation de la data sont susceptibles d'aider à relever ces défis. Cependant, la pratique de la médecine en tant que telle peut également bénéficier des technologies basées sur l'expérience des données.

Depuis quelques années, on assiste à l'émergence du concept de Médecine des 4P. C'est en 2013 que Leroy Hood définit la médecine personnalisée en tant que Médecine des 4P et en donne les 4 attributs (LEEM, 2019):

- Personnalisée : on s'intéresse au profil personnel de chaque individu, que ce soit le profil génétique, protéique, environnemental, ...L'objectif est de proposer des soins adaptés au profil particulier d'un individu à toutes les étapes de la prise en charge (prévention, diagnostic, traitement, suivi post-traitement) ;
- Préventive : on favorise le dépistage précoce des maladies et on améliore la qualité de vie des patients. La notion de « mieux-être » plutôt que de maladie est au centre de cette approche ;
- Prédicative : on essaye de sélectionner les traitements les plus appropriés en prenant en considération les risques de réactions secondaires à ces médicaments. On essaye également d'évaluer le risque pour chaque patient de développer une maladie ;
- Participative : L'objectif est d'amener les patients à être davantage acteurs de leur santé et de leur prise en charge. Etant plus informés de par l'accès à des ressources en ligne, ils sont également considérés comme « expert » dans le savoir subjectif de leurs troubles.

Grâce à l'apparition des nouvelles technologies s'appuyant sur les données comme le Big Data et leur analyse, chaque individu sera associé à un ensemble de données qui lui sont propres. L'objectif est donc de transformer ces données en de l'information utilisable afin d'améliorer l'expérience de soins du patient.

En ce qui concerne la médecine personnalisée, la collecte d'informations via des senseurs (comme des objets connectés) donnera aux praticiens d'avoir accès à un large panel d'informations (tel un tableau de bord) sur chacun des patients. Par exemple, en oncologie, en fonction de l'analyse du profil génétique de certaines tumeurs spécifiques à un patient, un praticien sera en mesure de prescrire un traitement personnalisé qui sera en mesure d'augmenter le taux de réponse au traitement. De façon plus accessible, en fonction des images d'une tumeur, le praticien pourra s'appuyer sur des algorithmes permettant de prédire l'évolution d'une tumeur et ainsi adapter la dosimétrie des rayons de radiothérapies à chaque patient afin de limiter les effets négatifs.

En ce qui concerne la médecine préventive, il y aura un volet de prévention populationnelle qui sera prédominant. Si l'on prend comme exemple le développement de maladies vectorielles dont le développement est affecté par le réchauffement climatique, la construction de modèles de déplacement des populations de vecteurs permettront de prévenir l'apparition de ces maladies dans les populations humaines. Cependant, les applications en médecine préventive se situent aussi à l'échelle individuelle. En prenant en compte les informations transmises par des senseurs (comme des objets connectés), le praticien sera à même de prévenir l'apparition de certaines maladies chez les patients en fonction de leurs conditions de vie. Par exemple, une équipe du MIT a mis au point un algorithme de Deep Learning capable de prédire à partir d'une mammographie si une patiente est susceptible de développer un cancer du sein à l'avenir (Yala *et al.*, 2019) et jusqu'à 5 années à l'avance.

En ce qui concerne la médecine prédictive, les technologies basées sur l'analyse des données permettent de nouvelles perspectives cliniques. L'analyse du génome d'un grand nombre d'individus a déjà permis d'identifier des mutations du génome qui favorisent l'apparition de tumeurs. Ainsi, la connaissance de ces facteurs de prédispositions couplés à des variables comportementales et environnementales permet

d'améliorer le dépistage de cancers chez des patients à risque mais également d'estimer les perspectives de survie des patients.

Enfin, en ce qui concerne la médecine participative, les objets connectés ont un aspect ludique qui responsabilisent le patient dans le suivi de ces activités et de ces habitudes de vie. De plus, ils produisent eux-mêmes les données qui permettront aux praticiens d'être plus efficaces dans leur prise en charge. Les applications sont également capables d'identifier des déviations dans le comportement ou dans les variables physiologiques qu'elles monitorent et permettent au patient de prendre rendez-vous auprès d'un praticien plus précocement.

Ces exemples sont des aperçus de ce que peut proposer les sciences de la donnée que nous détaillerons davantage dans la partie suivante en faisant un focus sur le traitement des maladies chroniques.

Partie 2 : Utilisation des données et de l'intelligence artificielle dans la gestion des maladies chroniques en médecine humaine

1. Prérequis à l'utilisation des données en santé

A. Que sont les données de santé ?

Les données de santé sont des données particulières, considérées comme sensibles et sont définies à l'échelle européenne par ce qu'on appelle le règlement européen sur la protection des données personnelles (RGPD : Règlement Général sur la Protection des Données). Selon ce règlement, les données à caractère personnel concernant la santé sont les données relatives à la santé physique ou mentale, passée, présente ou future, d'une personne physique (y compris la prestation de services de soins de santé) qui révèlent des informations sur l'état de santé de cette personne.

Cette définition comprend par exemple :

- Des informations sur une personne physique collectées lors de son inscription afin de bénéficier de services de soins ou lors de la prestation de ces services. Il peut s'agir d'un numéro unique permettant d'identifier sans ambiguïté une personne ;
- Des informations obtenues lors d'un examen ou d'un test (d'une partie du corps, l'ensemble du corps ou substances corporelles) y compris des données génétiques ;
- Des informations concernant une maladie, un handicap, des antécédents médicaux, des traitements, ...

Ces données sont donc des données à partir desquelles il est possible de déduire l'état de santé d'une personne. C'est une définition très large qui doit être adaptée au cas par cas selon la CNIL (Commission Nationale de l'Informatique et des Libertés) en fonction de la nature des données collectées. Toujours selon la CNIL, cette notion de données de santé fait émerger trois catégories de données :

- Les données de santé par « nature » : antécédents médicaux, maladies, traitements, résultats d'examen, etc.
- Les données qui deviennent des données de santé du fait du croisement avec d'autres données et qui permettent de tirer une conclusion sur l'état de santé : croisement d'une mesure de la masse corporelle avec d'autres mesures (glycémie par exemple), ...
- Les données qui deviennent des données de santé du fait de l'utilisation qui en est faite sur le plan médical (mesure de l'activité par exemple).

Ces données, du fait de leur caractère individuel et sensible, sont gérées par un régime juridique particulier que nous verrons dans la suite de cette thèse. Ce régime juridique particulier fait en sorte que l'utilisation et l'accès à ces données est très contrôlé et qu'il s'agit donc d'un enjeu majeur pour leur utilisation pour les nouvelles technologies de l'information.

B. Flots de données au sein des écosystèmes de santé

a. Qui produit les données de santé ?

Les données de santé proviennent de multiples sources. En effet, nous avons vu qu'en fonction des croisements avec d'autres informations, de nombreuses données peuvent être considérées comme des données de santé. Ainsi, la typologie de ces données est plurielle. On peut penser immédiatement aux données cliniques ou aux données biologiques. Cependant, les données comportementales peuvent aussi être considérées comme des données de santé. En fonction de l'activité physique d'un individu, en corrélant cela avec des habitudes alimentaires, on peut faire des hypothèses sur l'exposition à certaines maladies cardiovasculaires par exemple. On peut également penser aux données environnementales. En fonction de l'environnement dans lequel vit un individu, cela peut le prédisposer à développer certaines maladies (cancers

pulmonaires chez les fumeurs passifs par exemple). Ainsi, du fait de la pluralité des données susceptibles de rentrer dans la définition de données de santé, de nombreux acteurs peuvent produire ces données.

Les patients produisent évidemment de nombreuses données de santé via leurs dossiers médicaux, les comptes-rendus d'hospitalisation, les traitements ou consultations renseignés à l'Assurance Maladie pour les remboursements, ... Ils produisent aussi des données-patient via les objets connectés et les applications (nombre de pas par jour, durée de sommeil, ...). Enfin, ils produisent également ce qu'on appelle des données d'opinion en s'exprimant sur les réseaux sociaux, les sites Internet, en effectuant des recherches sur des moteurs de recherche, ... Ces données font partie de ce que l'on appelle les données de vie réelle. Il s'agit de toutes les données qui ne sont pas obtenues dans un cadre expérimental. Parmi ces données de vie réelle, on retrouve également les bases de données médico-administratives, des registres, ... Les données peuvent également être produites par la recherche clinique via les essais cliniques (publics ou privés).

b. Comment sont stockées les données de santé ?

En ce qui concerne la situation française, la mise en place de l'Espace Numérique de Santé (ENS) à la suite du rapport Villani depuis juin 2018 va probablement rebattre les cartes en ce qui concerne le stockage et l'accessibilité aux données de Santé. Jusqu'à présent, existait en France le SNDS (Système National des Données de Santé) qui est un entrepôt de données médico-administratives pseudo-anonymisées (anonymisées mais reliées à des individus uniques) contenant l'ensemble des soins présentés au remboursement. Le SNDS contient :

- Les données de l'Assurance Maladie ;
- Les données des hôpitaux ;
- Les causes médicales de décès ;
- Les données relatives au handicap ;
- Certaines données d'organismes complémentaires.

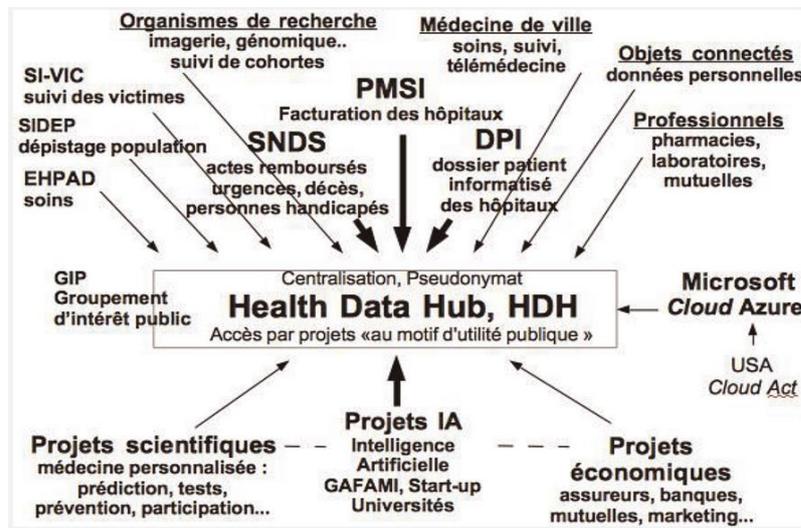
L'objectif de ce SNDS est de mettre à disposition des données dans le cadre d'études ou d'évaluations présentant un caractère d'intérêt public et s'inscrivant dans une des finalités définies. En revanche, de nombreuses données sont absentes du SNDS comme les résultats d'examen cliniques, les motifs de consultation, les facteurs de risques, les données sociales, les données produits par les objets connectés, les données issues d'études cliniques de laboratoires privés, ...

Le lancement de l'ENS répond à des priorités de la stratégie de la e-santé en France :

- Développer la médecine connectée à travers un plan de santé « Big Data » ;
- Encourager l'innovation en donnant accès à une base de données large ;
- Simplifier les procédures administratives pour les patients ;
- Renforcer la sécurité autour des données de santé.

L'ENS sera constitué de nombreuses parties comme d'une messagerie intégrée, un catalogue de services de référence, etc. Si l'on se concentre sur les données, l'ENS, contrairement au SNDS, contiendra davantage de données relatives aux individus comme les traitements, les résultats d'examen, ... L'ENS contiendra donc l'ensemble des données médicales des individus, les données de santé collectées par des senseurs, les données envoyées via une messagerie intégrée, des prescriptions digitales lors de téléconsultations, ... Cependant, comme on peut le voir dans la figure ci-dessous, l'objectif de l'ENS sera de rassembler la base de données de santé la plus riche possible et donc d'impliquer une grande variété d'acteurs depuis les laboratoires pharmaceutiques jusqu'aux acteurs économiques comme des assurances ou des banques.

Figure 8 : L'ensemble des données de santé présentes sur l'ENS (Fallery, 2020)



En 2023, l'ENS est encore un projet qui fait débat notamment à cause de l'hébergement des données mais aussi à cause de la réticence des patients à rendre accessibles leurs données (seuls 10 millions des 65,7 millions de détenteurs de compte ont activé leur carnet de santé numérique). C'est la société américaine Microsoft qui a été chargée par le gouvernement français d'accueillir les données *a priori* jusqu'en 2025 même si la décision n'est pas encore définitive. Cela pose un certain nombre de problèmes éthiques comme nous le verrons plus tard. Ainsi, à l'heure actuelle, il n'existe pas de base de données regroupant toutes les informations mais celles-ci sont plutôt segmentées en fonction des acteurs qui les produisent. Certaines sont présentes sur le SNDS, d'autres sont gérées par les acteurs privés comme des assurances ou des laboratoires pharmaceutiques, d'autres encore sont gérées par les entreprises qui n'ont à première vue pas de lien avec le secteur de la santé comme Apple ou Facebook. En effet, via les applications présentes sur leurs appareils ou les informations données par les utilisateurs sur les réseaux sociaux, ces géants des technologies possèdent une vaste base de données de santé.

c. Qui utilisent les données de santé ?

Si les données de santé peuvent être produites par une variété d'acteurs, elles sont également utilisées par une variété d'acteurs. Cependant, pour que celles-ci soient utilisées, il faut qu'elles soient accessibles et interopérables, c'est-à-dire qu'elles puissent être connectées entre elles. On comprend donc l'importance d'une plateforme unique afin d'accéder et d'exploiter les données. Si cette plateforme unique (qui est l'ambition de l'ENS) permet d'accéder aux données, elle permet aussi un contrôle par les autorités de la bonne utilisation de ces données. C'est une thématique importante du fait de la sensibilité des données. De plus, ce contrôle administratif permet de s'assurer de la bonne protection de ces données et notamment d'éviter les phénomènes de ransomware observé à l'encontre des établissements de santé depuis quelques années.

Une fois accessibles, les données de santé peuvent être utilisées :

- Par les patients via les services proposés sur la plateforme qui accueille les données. En effet, le projet de l'ENS est de centraliser les données mais également de permettre aux acteurs de la santé de proposer des services aux patients en utilisant les données disponibles sur la plateforme. En

fonction des données de santé qui sont présentes sur la plateforme et donc analysables par les professionnels de santé, ceux-ci peuvent conseiller aux patients des applications de prévention, de coordination, de qualité et de continuité des soins (Legifrance, 2022) ;

- Par les laboratoires pharmaceutiques en plus des données récoltées lors des essais cliniques. En effet, en collectant les données de vie réelle, les laboratoires peuvent améliorer le suivi de leurs médicaments sur le long-terme, observer des effets secondaires, découvrir des indications différentes des indications initiales, etc ;
- Par les professionnels de santé qui peuvent utiliser l'ensemble des données relatives aux patients afin de proposer une prise en charge personnalisée et ainsi améliorer l'efficacité de leurs interactions avec les patients ;
- Par les services d'accompagnement comme les assurances qui peuvent juger de l'efficacité des prises en charge mais aussi proposer des tarifs plus personnalisés (concept des néo-assurances) ;
- Par de nouveaux services proposant des solutions digitales qui utilisent les données afin de cibler les patients qui pourraient bénéficier des offres, construire leurs modèles de business et compléter des offres déjà existantes. Ces solutions digitales sont soumises à validation par la commission de référencement des services et outils numériques au catalogue d'applications de « Mon espace santé » et doivent être référencées dans un catalogue. Cette commission est composée de trois représentants d'usagers grand public des outils et services numériques ainsi que par quatre personnes qualifiées en raison de leur expertise dans les outils numériques de santé d'après l'Agence du Numérique en Santé (Agence du Numérique en Santé, 2022). Ce catalogue a vocation à évoluer mais certaines applications ont déjà été acceptées. Parmi ces applications, on peut en retrouver une créée par le laboratoire Abbott qui permettra aux patients diabétiques de directement partager leur taux de glucose moyen et des indicateurs de suivi de leur diabète avec « Mon espace santé ». Withings a également vu son application de suivi médical acceptée par la commission. La Poste, dont l'application « Mes médicaments chez moi » permet de faciliter la livraison des médicaments aux patients, pourra également intégrer l'ENS (Agence du Numérique en Santé, 2022).

C. Contrôle de la qualité des données de santé

Quand se pose la question de la qualité des données en santé, se posent en réalité deux questions. La première concerne la qualité des données qui permettent d'entraîner un algorithme. La deuxième question concerne la qualité des données proposées à l'algorithme afin qu'il retourne un résultat fiable.

Selon l'Inserm, la plupart des données médicales ne sont pas recueillies dans l'objectif d'entraîner un logiciel, elles ne sont donc pas parfaitement propres et annotées et donc exploitables directement. Toujours selon l'Inserm, la base de données du SNDS (en attendant celle de l'ENS) est difficile à exploiter pour les concepteurs de logiciels car elle a été créée pour l'analyse économique des prestations de santé plutôt que pour une analyse médicale. On pourrait relever jusqu'à 30% d'erreurs dans la description des pathologies associées aux malades. Corriger ces erreurs prend du temps et nécessite de croiser les données avec d'autres sources comme celles correspondant aux médicaments prescrits et en extrapolant ainsi les affections.

Selon Liaw et al, il n'existe pas de consensus afin de déterminer de la qualité des données. Cette diversité provient du fait que les besoins entre les chercheurs, les cliniciens, les informaticiens, ... peuvent différer (Liaw *et al.*, 2021). Selon eux, l'évaluation de la qualité des données doit être effectuée tout au long de la durée de vie de la donnée, c'est-à-dire de sa production à son archive. Cependant, les concepts permettant de définir la qualité d'une base de données varient d'une étude à l'autre et il est difficile de faire ressortir un consensus. Par exemple, au niveau des agences gouvernementales ou supra-gouvernementales, certains termes peuvent signifier des concepts différents. Ils définissent la qualité des données comme un mix de catégories intrinsèques à ces données et de catégories contextuelles à l'utilisation de ces données sans référence à un système établi de validation de la qualité.

En 2016, une collaboration de recherche internationale sur la qualité des données a développé un modèle afin d'évaluer les qualités d'une base de données. Ce modèle a connu des évolutions depuis sa création en 2016 afin de pouvoir continuer à être pertinent dans l'évaluation des bases de données utilisées en recherche ou en santé publique. Liaw a cependant suggéré qu'il était possible de compléter ce modèle en vérifiant un certain nombre de caractéristiques inhérentes aux bases de données :

- Des caractéristiques intrinsèques aux bases de données :
 - o Conformité : Homogénéité et conformité des données par rapport à leurs métadonnées (données sur les données),
 - o Complétude,
 - o Plausibilité.
- Des caractéristiques contextuelles quant à l'utilisation, la production des données :
 - o Accessibilité,
 - o Réutilisabilité,
 - o Gouvernance,
 - o Temporalité,
 - o Fiabilité.
- Des caractéristiques techniques des bases de données :
 - o Sécurité,
 - o Fragmentation,
 - o Traçabilité,
 - o Interopérabilité.

Ce que les auteurs ont surtout mis en lumière est l'absence de consensus sur le contrôle des données et ce tout au long de leur cycle de vie afin de s'assurer que celles-ci soient à même de répondre aux questions qui leur sont posées. Ils encouragent également les chercheurs à mener des enquêtes régulières sur la qualité de leurs données notamment dans un contexte où le Machine Learning et l'IA en général vont bouleverser les façons de faire de la recherche et de pratiquer la médecine.

D. La formation des médecins à l'Intelligence Artificielle

Dans le contexte d'une utilisation d'algorithmes d'IA en santé, la quantité et la qualité des données ainsi que les performances des logiciels en eux-mêmes ne suffisent pas à assurer une amélioration des services proposés par les professionnels de santé. En effet, ceux-ci doivent également être sensibilisés à ces thématiques afin de pouvoir utiliser correctement le potentiel de ces nouvelles technologies.

Selon une étude menée en 2018 par l'Observatoire ODOXA – UNESS (ODOXA, 2018), la formation des professionnels de santé vis-à-vis de l'IA est insuffisante. En effet, l'étude révèle que 73% des professionnels de santé s'estiment « démunis » et être mal formés en matière de numérique, d'intelligence artificielle et de robotisation alors même qu'une majorité (97% des professionnels et des étudiants) est convaincue qu'une révolution des systèmes de santé se prépare vis-à-vis de ces thématiques. Ainsi, 86% estiment qu'il est nécessaire de former davantage les professionnels de santé d'autant plus que 75% des sondés estiment que le développement des technologies digitales ainsi que de l'IA va permettre des améliorations de la formation des médecins et de la prise en charge des patients.

Toujours selon l'étude ODOXA – UNESS, il existe certes des formations proposées aux professionnels de santé mais seulement 9% les jugent efficaces, 33% les jugent inefficaces et 58% ne connaissent pas leur valeur. On comprend donc bien que si les technologies évoluent toujours plus vite et sont toujours plus performantes, elles s'avèreront inutiles sans le personnel apte à les utiliser.

De nombreuses formations existent désormais notamment dans le cadre de la formation continue obligatoire pour les professionnels de santé notamment sous la forme de Diplômes Universitaires mais ces formations sont éparpillées. Au contraire, le sondage de l'Observatoire ODOXA – UNESS a mis en évidence que les étudiants, les enseignants et les professionnels de santé étaient favorables à la création d'une plateforme numérique reconnue à l'échelle nationale afin d'établir des références communes.

Il est de la responsabilité des Universités de médecine de former leurs étudiants aux thématiques de l'IA et des nouvelles technologies en santé. Par exemple, la faculté de médecine d'Angers collabore étroitement avec le groupe ESEO (Ecole Supérieure d'Electronique de l'Ouest) afin de mieux former ingénieurs et médecins aux enjeux technologiques qui changeront la pratique de la médecine mais aussi afin de permettre aux ingénieurs de créer des dispositifs d'aide au diagnostic et au soin.

2. Les applications de l'Intelligence Artificielle dans la gestion des maladies chroniques en médecine humaine

A. Contexte de l'étude

Comme nous avons pu le voir dans les parties précédentes de cette thèse, les technologies liées à l'utilisation des données de manière générale et l'intelligence artificielle en particulier ont le potentiel d'impacter la façon dont nous pouvons pratiquer la médecine. Ces impacts peuvent se voir dans différents domaines comme la télé-régulation, l'admission dans les structures de soins ou encore en épidémiologie. Ce ne sont que des exemples et l'objectif est ici de montrer la diversité des impacts possibles. Pour la suite de notre étude, nous avons choisi de nous limiter à l'étude de l'impact de ces nouvelles technologies dans la gestion des maladies chroniques.

Selon l'OMS (ORGANISATION MONDIALE DE LA SANTE, 2018), on peut définir les maladies chroniques de la façon suivante : il s'agit « d'affections de longue durée qui en général, évoluent lentement ». Le Ministère de la Santé (Ministère de la Santé et de la Prévention, 2023) ajoute que ce sont des maladies ayant un impact sur le quotidien. On comprend que la définition de maladie chronique est large et Stéphanie Bernell ainsi que Stephen W. Howard remarquent qu'en fonction des organisations, les maladies comprises sous l'intitulé « maladies chroniques » diffèrent (Bernell et Howard, 2016). Dans la suite de cette thèse, nous adopterons la définition donnée par l'OMS ainsi que les maladies chroniques qui y sont citées : maladies cardiovasculaires, cancers, maladies respiratoires chroniques et diabète. Ce sont les causes majeures de mortalité dans le monde.

De plus, nous faisons le choix de nous intéresser à la façon dont les nouvelles technologies permettent d'aider les professionnels de santé à gérer ces maladies chroniques. Ainsi, nous nous intéressons à la gestion d'une manière générale, c'est-à-dire la prévention et le monitoring avant l'apparition d'une pathologie chronique chez un patient, le diagnostic de la pathologie ainsi que le traitement et la prise en charge sur le long terme des patients souffrant de maladies chroniques. De plus, nous ne nous intéresserons pas à l'apport des nouvelles technologies basées sur les données permettant d'améliorer les processus de recherche sur les maladies chroniques. La recherche privée connaît notamment des bouleversements importants depuis quelques années avec la création de plateformes par de grands laboratoires pharmaceutiques utilisant les données issues des essais cliniques et les données de vie réelle afin d'accélérer les processus de découverte et de

mise sur le marché. L'objectif y est d'utiliser des algorithmes de Machine Learning afin de scruter les données et de découvrir notamment de nouvelles indications à des molécules existantes.

B. L'apport de l'intelligence artificielle dans le monitoring et la prévention des maladies chroniques

a. L'apport de l'IA dans l'intégration des facteurs génétiques dans la prédiction des risques d'apparition des maladies chroniques

En ce qui concerne, la prévention de maladies chroniques, les apports de l'Intelligence Artificielle sont encore relativement limités et se cantonnent principalement à de la recherche à l'heure actuelle. Il existe des études faisant état du potentiel de ces technologies dans la prévention des maladies chroniques mais elles ne sont pas déployées à grande échelle.

Si l'on prend l'exemple de la génomique, selon Raquel Dias et Ali Torkamani, certaines études ont essayé de montrer les capacités d'algorithmes d'Intelligence Artificielle dans la prédiction de certains traits humains à partir d'une analyse génétique (Dias et Torkamani, 2019). Cependant, ces études n'ont pas été capables de montrer avec un degré suffisant de certitudes des corrélations entre patrimoine génétique et les phénotypes présentés.

Si certaines analyses statistiques relativement simples permettent une stratification du risque pour certaines maladies complexes communes, l'une des meilleures réussites en ce qui concernent le Machine Learning est une prédiction de la taille d'un individu en fonction de ces facteurs génétiques. Cette illustration montre le fait que des méthodes basées sur l'IA peuvent être développées. Raquel Dias et Ali Torkamani pensent cependant que l'utilité d'approches basées sur l'IA en ce qui concerne les prédictions de phénotypes en fonction du génotype viendra de l'intégration des données de santé et d'environnement à ces données génétiques. En effet, de nombreuses maladies chroniques sont le résultat d'interactions entre des facteurs de risques génétiques, environnementaux ainsi que des comportements individuels. Les facteurs de risques génétiques sont en effet des indicateurs de base pour savoir si une maladie chronique peut se développer mais l'intégration des autres facteurs est nécessaire afin d'affiner les stratifications de risques. Ainsi, sans accès à de telles bases de données (qui doivent être de taille suffisamment volumineuses et fiables), il est difficile d'envisager des algorithmes performants. De telles bases de données peuvent être construites. Une équipe de chercheurs menée par Andrew Lee a réussi à améliorer des modèles de stratification de risques de cancer du sein en associant des données génétiques avec des données sur d'autres facteurs de risques connus comme l'âge à la ménopause, l'âge du premier enfant ou encore la densité mammaire (Lee *et al.*, 2019). Des améliorations similaires ont été observées dans la stratification de risques pour des coronaropathies en intégrant des données génétiques et des données cliniques et environnementales.

L'apport des modèles basés sur l'Intelligence Artificielle pour déterminer le risque d'un individu d'être atteint d'une maladie chronique en fonction de facteurs génétiques, cliniques, comportementaux et environnementaux n'en est encore qu'à ses débuts mais ont un potentiel certain d'après Raquel Dias et Ali Torkamani. Ce potentiel pourrait être d'autant plus important quand le développement d'une maladie chronique est dû à des interactions nombreuses et faibles entre différents facteurs. Ces interactions sont difficilement captées par les modèles statistiques actuels mais pourraient l'être beaucoup mieux par des algorithmes basés sur l'IA. Outre-Manche, des bases de données sont en train d'être montées afin d'entraîner des modèles ayant cet objectif et les avancées sont d'autant plus prometteuses que des applications de l'IA sont déjà des standards de référence dans certains champs d'applications de la génomique.

b. L'apport de l'IA dans la prédiction de l'apparition des maladies chroniques

Pour certaines maladies chroniques, il peut être important de pouvoir déterminer des risques d'apparition de la maladie avant que celle-ci n'apparaisse effectivement afin de prévenir ou retarder son apparition. Nous illustrerons ce point en prenant l'exemple de trois maladies chroniques : les maladies apparaissant des suites de l'obésité, la maladie d'Alzheimer et l'asthme.

L'obésité est classée comme la cinquième cause de mort dans le monde. L'obésité n'est pas la cause de la mort en tant que tel mais contribue à l'apparition de nombreuses maladies chroniques comme des cancers, le diabète, des désordres métaboliques ou encore des maladies cardiovasculaires. Ainsi, la détection précoce et la lutte contre des facteurs de risque permettrait de limiter l'apparition de ces maladies. L'indice de Masse Corporel (IMC) est toujours considéré comme un outil de mesure efficace afin de classer le poids des individus adultes. Cette classification permet notamment de déterminer quels individus sont les plus susceptibles de présenter des complications médicales du fait de leur poids.

De multiples études ont montré que l'obésité était un problème complexe issu d'une combinaison de facteurs comportementaux aussi bien que génétiques et qu'il peut se présenter sous de multiples formes et atteindre pratiquement tous les organes (Bischoff *et al.*, 2017). Comprendre les causes et les facteurs déterminants de l'obésité est donc une étape critique dans la mise en place de programmes de prévention afin de lutter contre l'apparition de maladies chroniques. A cet effet, il a été montré que l'utilisation d'algorithmes de Machine Learning avait montré des résultats encourageants dans la prédiction de l'obésité en fonction d'un certain nombre de facteurs de risques, aussi bien médicaux qu'environnementaux. Par exemple, Singh et Tawfik ont été capables de prédire l'apparition d'un IMC anormalement haut chez des adolescents avec une précision de 93,4% (Singh et Tawfik, 2020). Selon Safaei *et al.*, les techniques d'intelligence artificielle et particulièrement l'utilisation de réseaux de neurones permet d'améliorer grandement la prévision de l'obésité en fonction de facteurs cliniques, génétiques et environnementaux, notamment par rapport aux techniques statistiques traditionnelles (Safaei *et al.*, 2021). Cependant, selon eux, l'utilisation de tels algorithmes reste inexploitée et pourrait être plus développée en pratique régulière à la vue des bonnes performances de ces algorithmes.

Selon l'OMS, 30% des maladies chroniques à horizon 2030 seront des conséquences du mode de vie des individus et plus particulièrement de l'obésité engendrée par la sédentarité. Nous comprenons donc l'importance de la démocratisation de tels outils encore confidentiels aujourd'hui, qui permettrait d'accompagner en amont les individus présentant de fortes dispositions à l'obésité.

Si l'on s'intéresse à la maladie d'Alzheimer, la situation est différente puisqu'il s'agit d'une maladie pour laquelle les médicaments actuels ne sont pas jugés suffisamment efficaces pour être remboursés par la Haute Autorité de Santé. Cependant, il existe des habitudes qui permettent de ralentir le développement de la maladie si celle-ci se présente. Ainsi, l'enjeu de pouvoir détecter rapidement la maladie d'Alzheimer est très important. Les patients dans un stade précoce de la maladie d'Alzheimer sont dans une phase appelée Déficit Cognitif Modéré. Selon Petersen *et al.*, 30% à 40% des personnes présentant cet état développeront par la suite la maladie d'Alzheimer (Petersen *et al.*, 1999). En revanche, tous les malades atteints d'Alzheimer présentent cette phase. Ainsi, dans l'optique d'adopter des gestes empêchant ou retardant l'apparition de la maladie d'Alzheimer, il est important de pouvoir dépister le plus tôt possible ce déficit cognitif. Les méthodes de dépistage de cet état reposent sur des méthodes d'imagerie comme l'Image par Résonance Magnétique (IRM) ou la tomodensitométrie. Cependant, le diagnostic est difficile en raison de la résolution des zones du cerveau concernées (lobe moyen). Par ailleurs, le diagnostic clinique de la maladie d'Alzheimer est difficile et chronophage, il s'agit d'une combinaison de tests psychologiques, d'évaluations cliniques et d'imagerie. Gao et Lima ont montré que des algorithmes de Machine Learning étaient meilleurs que des opérateurs humains dans la détection de Déficit Cognitif Modéré ce qui permet donc *in fine* de détecter plus précocement les patients susceptibles de développer la maladie d'Alzheimer à l'avenir (Gao et Lima, 2022). Cependant, ils reconnaissent que la qualité des résultats dépend de la qualité des images fournies à l'algorithme du fait du changement de résolution du lobe moyen.

Enfin, l'asthme est la maladie chronique la plus fréquente chez les enfants, elle touche près de 340 millions de personnes dans le monde et la catégorie d'âge la plus touchée est la tranche d'âge des 10-14 ans. De nombreux enfants présentent une respiration sifflante sans pour autant développer de l'asthme plus tard. La question est donc d'être capable de déterminer les enfants qui ont des chances accrues de présenter de l'asthme plus tard afin de pouvoir mettre en place des mesures de prévention et ainsi réduire la prévalence de cette maladie (Patel *et al.*, 2022). A l'heure actuelle, les modèles de prédiction de l'asthme conventionnel reposent sur des méthodes statistiques utilisant des grilles de score en fonction des signes cliniques et renvoyant des intervalles de risque. Ces méthodes présentent des sensibilités comprises entre 36% et 95%, des spécificités entre 53% et 93% ainsi que des valeurs prédictives positives et négatives respectivement comprises entre 18% à 76% et 68% à 96%. Aucun modèle n'a cependant été capable de produire de bonnes valeurs dans toutes ces caractéristiques et aucun n'a été adopté à grande échelle. L'une des raisons pour lesquelles ces modèles n'ont pas présenté de bons résultats est le fait que l'asthme est une maladie associée à de nombreux facteurs de risque entretenant des relations complexes entre eux. Par ailleurs, peu d'études ont pris en compte les impacts de facteurs génétiques. Patel et al ont montré que l'utilisation d'algorithmes de Machine Learning dans la prédiction de l'apparition de l'asthme a permis d'obtenir des résultats plus précis que les modèles conventionnels utilisés avec, en fonction des modèles, une sensibilité allant de 73% à 100% et une spécificité allant de 71% à 100%. Cependant, dans ce cas précis, les études ont montré qu'il existait un risque de biais dû à la qualité des données utilisées pour entraîner les modèles.

Au travers de ces trois exemples, nous avons pu voir que les algorithmes de Machine Learning permettaient pour certaines maladies de détecter de façon précoce des probabilités de survenue des maladies chroniques. Dans le cas de la maladie d'Alzheimer, les algorithmes permettaient de détecter de façon précoce des déficits cognitifs qui peuvent être des signes précurseurs de la maladie. Dans le cas de l'asthme et de l'obésité, les algorithmes étaient capables de prendre en compte les différences de risques associés à ces états et de déterminer des probabilités d'apparition d'asthme et d'obésité (elle-même facteur déterminant dans l'apparition de maladies chroniques variées). La prise en compte de ces probabilités d'apparition permet donc aux médecins de prescrire des médicaments ou de mettre en place des procédures dans le but de prévenir l'apparition de la maladie ou au moins de la retarder. Cependant, il semble que ces processus soient encore loin d'être démocratisés et ne sont utilisés que par des équipes de recherche ou dans des services très en avance sur ces thématiques.

c. L'apport de l'IA dans la création d'outils de sensibilisation

Comme nous l'avons vu précédemment, certains comportements peuvent constituer des facteurs de risque liés au développement de maladies chroniques. Ainsi, sensibiliser les individus et particulièrement les adolescents est un enjeu majeur dans la prévention de certaines maladies. J.P Rowe et J.C Lester ont mis en avant l'apport d'algorithmes d'intelligence artificielle dans la conception d'outils interactifs permettant de sensibiliser les adolescents et de les filtrer en fonction de leur profil (Rowe et Lester, 2020). Par exemple, des applications digitales ou des jeux digitaux sont un véhicule naturel pour délivrer des messages de prévention qui viendront compléter les interactions des patients avec leur médecin ainsi que promouvoir de bonnes pratiques en termes de suivi médical. En exploitant les interactions des individus avec le logiciel, il est possible de déduire le risque d'exposition à certains facteurs et ainsi prédire l'apparition de maladies liées à ces facteurs. Les interactions des individus avec les logiciels peuvent être complétées à l'aide de senseurs disposés sur le support comme les expressions faciales, le langage ou encore des informations biométriques. Concrètement, Ozer et al ont développé Inspire, un jeu situationnel basé sur la consommation d'alcool (Ozer *et al.*, 2020). Au cours de ce jeu, des adolescents étaient mis dans différentes situations dans lesquelles ils devaient choisir un comportement à adopter. En fonction des choix qui étaient faits mais également en fonction du temps de réponse et d'autres paramètres, les situations évoluaient afin de cerner le mieux possible le profil de l'utilisateur.

Concrètement, 3 applications sont exploitées à l'heure actuelle pour construire le profil des utilisateurs :

- L'évaluation furtive : l'objectif est de mesurer de façon non-intrusive les connaissances d'un individu sur un sujet en particulier en utilisant un environnement situationnel. L'évaluation se fait de manière indirecte en collectant de plus en plus d'informations en fonction de l'avancement de l'utilisateur. Les données utilisées en entrée sont des informations de base sur l'utilisateur tandis que les données en sortie sont une prédiction du comportement de l'utilisateur face à certaines situations. Selon J.P Rowe et J.C Lester, évaluer les connaissances des adolescents sur les comportements bénéfiques à leur santé et les risques associés est un processus essentiel dans la présentation d'informations pertinentes aux adolescents afin de créer les fondations d'améliorations de leurs comportements ;
- La modélisation affective : l'objectif est de déterminer ce qui permet de répondre à des émotions négatives de la part des utilisateurs comme la frustration ou l'ennui, de mesurer les réactions positives à certains stimuli et de déterminer ce qui peut aider les utilisateurs qui sont confrontés à de l'anxiété ou de la confusion. La gestion de ces émotions a un impact sur la capacité des individus en général et des adolescents en particulier à ingérer les informations. L'objectif sur le long terme est de déterminer quels sont les formats et les outils permettant d'appréhender au mieux les émotions des individus et ainsi de pouvoir apporter du soutien dans le maintien des comportements bénéfiques pour la santé ;
- Reconnaissance des objectifs : l'objectif est de déterminer quels sont les véritables objectifs et préoccupations des utilisateurs sachant que ces objectifs sont souvent cachés au système et doivent être inférés des comportements des utilisateurs. Intuitivement, déterminer les objectifs des utilisateurs est essentiel afin d'anticiper leurs comportements et de proposer des approches cohérentes avec leurs objectifs et leurs comportements.

Selon Rowe et Lester, les technologies de santé adaptatives comme les algorithmes situationnels sont encore peu utilisées mais sont appelées à se développer. En effet, outre l'effet éducatif qu'ils proposent, ils sont un outil adapté à l'évolution de l'appétence des nouvelles générations pour les technologies. Ces méthodes présentent de meilleurs taux d'engagement notamment de la part des adolescents et permettent d'adapter les messages de prévention et d'éducation à chacun. Ils sont également compatibles avec le système de santé traditionnel dans la mesure où les informations en sortie fournies par ces logiciels peuvent aider les médecins à faire passer des messages pertinents s'ils ont accès aux résultats.

C. L'apport de l'IA dans le diagnostic des maladies chroniques

a. L'apport de l'IA pour le diagnostic précoce de maladies chroniques

Comme nous l'avons vu précédemment, il y a un réel intérêt à pouvoir détecter de façon précoce la survenue d'une maladie afin de pouvoir anticiper sa prise en charge. S'il est intéressant de pouvoir déterminer les facteurs de risque de survenue d'une maladie afin de pouvoir contrôler régulièrement l'état du patient, il est également primordial de pouvoir détecter précocement l'apparition d'une maladie. Pour illustrer ce point, nous prendrons l'exemple du diagnostic de certains cancers à commencer par les cancers oraux.

Les cancers oraux, sont la sixième cause de cancers dans le monde et présentent des taux de mortalité et de morbidité associés élevés (Ilhan *et al.*, 2021). En effet, il est estimé que le taux de survie après diagnostic est évalué à 45% pour deux ans. Ce taux de survie faible est expliqué par le fait que ce sont des cancers qui sont diagnostiqués à un stade loco-régional avancé. Il y a trois raisons qui peuvent expliquer le délai du diagnostic. La première raison vient de la durée avant une consultation. Selon Güneri et al, le délai entre l'apparition de signes et la première consultation oscille entre 1,6 et 5,4 mois (Güneri et Epstein, 2014). Cela peut être relié à la deuxième raison expliquant le délai qui est l'organisation des systèmes de santé. En effet, en fonction de

la réactivité des assurances, de la possibilité de prendre rendez-vous, des coûts, du fait d'être référé à un spécialiste, etc, le temps entre l'apparition des signes et le diagnostic peuvent être rallongé. Enfin, la troisième raison tient dans le fait que de nombreux praticiens généralistes ne sont pas à même de diagnostiquer ou d'orienter efficacement les patients vers le bon diagnostic. Cette assertion est également vraie pour les dentistes. Selon Seone et al, la performance diagnostique des dentistes pour des cancers oraux est de 57,8% de sensibilité et de 40% de spécificité environ (Seoane *et al.*, 2006). Par ailleurs, il a été montré par Seoane et al que l'accord de diagnostic entre dentistes et spécialistes n'était que de 40%. Ainsi, que ce soit à la suite d'un référé ou d'un mauvais diagnostic, le délai augmente. Pourtant, dans le cas des cancers oraux, le taux de rémission lorsque le diagnostic est rapide est supérieur à 90%. En revanche, en cas de métastases, le taux de survie à 5 ans tourne entre 16% et 23%. On comprend donc l'importance de pouvoir diagnostiquer rapidement cette maladie. Une bonne façon de pouvoir diagnostiquer rapidement ce type de maladie serait de faire en sorte que les médecins généralistes et les dentistes soient davantage capables de mettre en évidence ces pathologies. Un algorithme de Machine Learning disponible sur Internet et développé par une équipe de chercheurs chinois est capable, à partir de photos de lésions, de différencier les lésions en fonction de leur probabilité d'abriter des cellules cancéreuses avec une spécificité et une sensibilité importante. L'intérêt de cet algorithme est qu'il fonctionne avec des images qui peuvent être prises avec un téléphone portable à condition que la qualité de l'image soit suffisante. En plus de permettre à des non-spécialistes de pouvoir diagnostiquer ou au moins d'orienter le diagnostic, cette solution possède l'avantage d'être très économique et rapide et ainsi de réduire le délai de diagnostic des cancers oraux. Selon Ilhan et al, ces technologies sont encore en développement et ne sont pas encore utilisées à grande échelle mais possèdent des avantages certains qui pourraient en faire des outils disponibles dans de nombreux cabinets et non réservés à des équipes de spécialistes.

Pour certains cancers très prévalents à partir d'un certain âge comme le cancer de la prostate, des tests réguliers sont préconisés afin de détecter précocement cette maladie. Dans le cas du cancer de la prostate, un des tests fréquemment utilisé notamment aux Etats-Unis est un test sérologique. Cette technique présente une spécificité ainsi qu'une sensibilité relativement faible et conduit donc à un sur-diagnostic (et donc des examens complémentaires/traitements) dans des cas bénins et à des sous-diagnostic de cancers prostatiques effectifs. Ce test a donc été progressivement remplacé par des IRM qui présentent de meilleures sensibilité et spécificité. Cependant, du fait de la démocratisation des tests de prévention, les imageurs doivent faire face à une demande accrue de tests. Se pose donc le défi de pouvoir proposer suffisamment de tests (et d'interprétations) avec une qualité suffisante. Selon de Rooji et al, des algorithmes de Machine Learning peuvent être utilisés en premier lieu afin d'évaluer rapidement la qualité de l'image qui a été prise permettant ainsi d'assurer que tous les tests IRM effectués donnent des images exploitables afin d'obtenir un diagnostic (de Rooij *et al.*, 2021). Par ailleurs, d'autres algorithmes de Machine Learning permettent de classifier et donc d'établir un diagnostic quant à la présence ou l'absence de lésions de cancer de la prostate. Certains algorithmes commercialisés montrent même des améliorations dans la qualité du diagnostic par rapport à des spécialistes. Même s'il paraît prématuré que des algorithmes remplacent des imageurs humains pour des pathologies aussi graves, de Rooji et al notent que ces algorithmes servent aujourd'hui de triage préalable et de second avis afin d'affiner le diagnostic.

b. L'IA permettant d'améliorer la fiabilité des outils dans le diagnostic de maladies chroniques

L'amélioration des capacités diagnostiques est un apport très médiatisé quand il s'agit de l'utilisation de l'Intelligence Artificielle en santé. Par ailleurs, un rapport paru en 2020 du Grand View Research estime que le marché de l'Intelligence Artificielle pour le diagnostic médical passerait de 288 millions d'euros en 2019 à près de 3 milliards en 2027 (« Artificial Intelligence Market Size Report, 2022-2030 », 2022). Cependant, l'application concrète de telles technologies se limite essentiellement aujourd'hui à l'imagerie médicale. En effet, l'imagerie médicale (que ce soit l'interprétation ou l'acquisition d'images) est l'un des seuls secteurs qui permet aujourd'hui une intégration de ces outils dans une pratique courante de la médecine. Nous prendrons

donc cet exemple pour illustrer comment des outils basés sur l'Intelligence Artificielle peuvent présenter des compétences diagnostiques supérieures à celles de professionnels humains. Nous nous focaliserons sur les applications en imagerie thoracique en gardant à l'esprit que cela peut être généralisé à d'autres types de clichés.

Il est nécessaire de commencer par expliquer que les algorithmes sont codés soit avec pour objectif de répondre à une question unique (c'est-à-dire reconnaître une lésion en particulier) soit pour identifier un certain nombre de lésions. Dans la situation actuelle, il n'existe pas d'algorithme qui soit capable de détecter toutes les lésions détectables sur un cliché. Parmi les algorithmes développés pour une tâche en particulier, une tâche majeure consiste à diagnostiquer des nodules pulmonaires sur des radiographies thoraciques. L'objectif est de détecter la présence de nodules mais également de déterminer la probabilité de malignité de tels nodules. Par exemple, Athel et al ont développé un algorithme qui a été capable de diagnostiquer la présence de nodules avec une sensibilité de 92% et une spécificité de 86% (Ather *et al.*, 2020). De plus, Nam et al ont développé un algorithme de Deep Learning capable de détecter des nodules pulmonaires malins sur des radiographies et ont comparé ses performances avec celles d'un consortium de radiologues (Nam *et al.*, 2021). Ils ont montré que les performances de leur algorithme étaient supérieures à celles de 17 des 18 radiologues utilisés comme témoin. De plus, ils ont montré qu'avec l'aide de cet algorithme, tous les radiologues ont amélioré leur score. Des résultats similaires ont été démontrés sur d'autres affections comme la tuberculose ou des pneumonies. Rohmah et al ont par exemple développé un algorithme capable de détecter des signes de tuberculose sur des clichés radiographiques avec une précision de 95,7% (Rohmah *et al.*, 2013). Des solutions similaires ont montré être plus performantes que des radiologues professionnels. Il faut cependant rappeler que ces solutions ne sont entraînées que pour un type de lésion bien précis, ce qui rend leur potentiel d'utilisation limité.

Certains chercheurs se sont donc penchés sur la nécessité de construire des modèles capables de détecter des anomalies sur des clichés radiographiques. Rajpurkar et al ont cherché à comparer les performances de Chexnet, un algorithme développé afin de diagnostiquer 14 affections comme de la fibrose, des hernies, des cardiomégalies, ..., avec celles de radiologues. Ils ont montré que, excepté pour les cardiomégalies, de l'emphysème et des hernies hiatales, les performances de Chexnet étaient supérieures à celles des radiologues (Rajpurkar *et al.*, 2018). De façon encore plus générale, Hwang et al ont développé un modèle de Deep Learning capable de détecter si un cliché radiographique est normal ou anormal (Hwang *et al.*, 2019). L'algorithme s'est avéré plus performant que des médecins généralistes, des radiologues et même que des radiologues spécialistes de la région thoracique. Par ailleurs, pour tous les médecins participant à l'expérience, leur précision a été améliorée à la suite de l'utilisation de l'algorithme.

Nous avons ici pris l'exemple de radiographies mais des études similaires ont été menées sur des images de scanner. Par exemple, Lung CAD VB10A développé par Siemens et disponible depuis 2006 a montré avoir une sensibilité de 96,7% dans la détection de nodules pulmonaires. Cependant, la sensibilité tombe à 50% lorsque l'on prend en compte les nodules pulmonaires subsolides. Cela illustre la difficulté à mettre en place des algorithmes pouvant détecter plusieurs lésions. Un autre exemple est le développement par Walsh et al d'un outil permettant de détecter des maladies pulmonaires fibrotiques. Même si l'algorithme a présenté une meilleure sensibilité que les radiologues pris en comparaison, cette sensibilité n'était que de 73,3% (Walsh *et al.*, 2018).

Comme nous avons pu le voir, même si l'usage d'outils basés sur l'IA n'est pas encore démocratisé, son utilisation peut permettre d'obtenir des résultats supérieurs à ceux obtenus par des opérateurs humains seuls. Dans le domaine de l'imagerie, où ces outils sont davantage utilisés que dans d'autres domaines, l'application se retrouve comme support pour les imageurs puisque nous avons vu que de tels outils permettaient aux professionnels d'améliorer leur résultat lorsque l'algorithme fait office de premier ou de second lecteur. Cependant, si les algorithmes basés sur l'IA peuvent trouver une application directement dans le diagnostic, ceux-ci sont également utilisés pour fluidifier la mise en place de paramètres facilitant le diagnostic.

Nous avons choisi d'illustrer les capacités de l'IA en matière de diagnostic en prenant l'exemple de l'imagerie médicale. Cependant, il faut garder en mémoire que si l'imagerie est un domaine dans lequel l'utilisation d'outils à base d'IA est répandue, ce n'est pas le seul. Dans le domaine de l'oncologie par exemple, le « gold standard » en matière de diagnostic est souvent l'histopathologie. Dans ce domaine également, des outils basés sur l'IA sont capables de détecter un certain nombre d'anomalies et d'établir des diagnostics.

c. L'IA permettant de faciliter la mise en place d'un environnement favorisant le diagnostic des maladies chroniques

Nous avons vu jusqu'à présent que les technologies basées sur l'IA peuvent être utilisées à des fins diagnostiques soit pour permettre un diagnostic plus précoce qu'avec des outils traditionnels mais aussi qu'ils pouvaient avoir des performances supérieures à celles de professionnels humains dans les domaines où ce type de technologie est répandu. Cependant, l'IA peut également permettre d'améliorer le diagnostic en facilitant le travail du praticien et en le mettant dans de meilleures conditions pour qu'il puisse poser un diagnostic.

Par exemple, si nous poursuivons avec l'exemple de l'imagerie, un radiologue ne passe « que » 53,8% de son temps à analyser des images (Schemmel *et al.*, 2016). Le reste du temps est passé sur des tâches non-interprétatives comme la construction de protocoles ou des entretiens avec des cliniciens. En gagnant du temps sur d'autres tâches non-interprétatives, les radiologues seraient à même de passer plus de temps sur l'interprétation d'image et donc d'être en mesure de poser de meilleurs diagnostics. A titre d'exemple, les radiologues passent plus de 6% de leur temps à construire des protocoles. Lee a développé un algorithme afin d'automatiser cette procédure. Il a été rapporté que cet algorithme présentait de très bons résultats lorsqu'il s'agit de déterminer le bon protocole d'imagerie en fonction des patients et de ses besoins (Lee, 2018). Implémenter un tel algorithme dans des services d'imagerie ayant une forte activité permettrait de fluidifier le travail des radiologues et de passer plus de temps sur l'interprétation d'images.

Toujours en imagerie, les technologies basées sur l'IA permettent également d'améliorer le processus d'acquisition des scans. En effet, l'une des faiblesses majeures de l'utilisation des IRM est le long temps d'acquisition pouvant être désagréable pour le patient. Ainsi, des artefacts de mouvements peuvent apparaître. De plus, plus le temps d'acquisition est long et plus le débit de patients pouvant utiliser le l'IRM est faible. Ainsi, l'objectif serait d'augmenter le débit de l'IRM tout en permettant une meilleure résolution d'image. Hyun *et al* ont montré que leur algorithme était capable de recréer une image de qualité suffisante seulement à partir d'un échantillon des données (Hyun *et al.*, 2018). En utilisant cette technique, on réduit le temps d'acquisition de l'IRM, réduisant ainsi le temps d'attente pour les patients et donc la probabilité d'apparition d'artefacts de mouvement. En implémentant ces techniques dans un service, nous pourrions donc augmenter la qualité des images ainsi que leur nombre et permettre ainsi aux cliniciens de travailler sur davantage de patients tout en bénéficiant d'images de meilleures qualités.

D'autres tâches sont actuellement réalisées par des algorithmes et permettent aux cliniciens de gagner du temps sur leurs tâches non-cliniques. Par exemple, des algorithmes peuvent être utilisés pour prédire la probabilité qu'un patient ne se présente pas à son rendez-vous. L'algorithme utilise un certain nombre de paramètres comme les informations liées au patient (âge, sexe), les caractéristiques du rendez-vous (durée, jour de la semaine, nombre de rendez-vous en une journée) ou encore les conditions météorologiques afin de prédire les patients qui ne se présenteront pas au rendez-vous. Ainsi, ces informations peuvent être utilisées pour améliorer l'expérience patient et de permettre aux cliniciens de ne pas expérimenter de temps mort. Enfin, l'IA peut également améliorer l'efficacité quotidienne de services d'imagerie en générant les comptes-rendus automatiquement et permettant ainsi aux cliniciens de gagner du temps. À la suite d'un examen, des

algorithmes utilisent les informations de l'examen afin de générer rapidement un compte-rendu exhaustif. Ces comptes-rendus peuvent être aussi bien utiles aux jeunes cliniciens afin de leur montrer des modèles-types qu'aux cliniciens plus expérimentés en leur faisant gagner du temps. De plus, ces algorithmes sont très utiles en termes de comparaison de l'évolution d'une maladie puisqu'ils sont objectifs. Ils permettent d'éviter de mauvaises interprétations sur l'évolution d'une maladie qui pourrait être dues à des différences d'interprétation d'un même examen entre les différents praticiens.

D. L'apport de l'IA dans le traitement et le suivi long-terme des maladies chroniques

a. *La chirurgie assistée par l'IA dans la gestion des maladies chroniques*

Des robots sont régulièrement utilisés dans le cadre de chirurgie que ce soit dans un but de filtration des tremblements, de stabilisation des instruments ou de visions en 3D (Shin *et al.*, 2022). Il a été montré que l'utilisation de telles technologies permettait de réduire le temps de récupération post-opératoire, de réduire les pertes sanguines ou encore de réduire le temps de chirurgie. L'utilisation de robots est importante notamment dans le domaine de la chirurgie cardiaque à l'image de chirurgies de la valve mitrale ou d'opérations sur la fibrillation atriale. Le nombre de centres et d'hôpitaux utilisant ces technologies augmente du fait du nombre d'études montrant l'intérêt de tels robots pour la sécurité des patients ou encore des économies en termes de coûts associés. Ces robots reposent (partialement ou non en fonction des robots) sur des mécanismes de Machine Learning. Par exemple, STAR (Smart Tissue Autonomous Robot) est un robot autonome qui a été utilisé dès 2016 sur des opérations de l'appareil intestinal. Les résultats ont montré que STAR était plus agile et plus précis que des médecins humains. STAR est autonome mais peut également être contrôlé à distance par des opérateurs humains. STAR a été entraîné par des algorithmes d'Intelligence Artificielle. De nombreux autres robots comme ROBOCAST (spécialisé en neurochirurgie) ont été développés par la suite. Un intérêt de ces robots est leur potentiel d'utilisation pour réaliser des opérations aussi peu invasives que possible. C'est notamment le cas pour les interventions où le risque est élevé pour les patients.

Si l'on prend un exemple de chirurgie qui permet le traitement des maladies chroniques, des robots basés sur l'Intelligence Artificielle sont utilisés pour des transplantations rénales chez les patients souffrant d'insuffisance rénale chronique. Defortescu et Tillou ont montré que les robots présentaient de très bons résultats aussi bien chez les donneurs que chez les receveurs, permettaient de réduire les douleurs post-opératoires et présentaient un intérêt particulier chez les patients obèses et/ou sous anticoagulants (Defortescu *et al.*, 2016). Les algorithmes d'IA sont utiles pendant l'intervention en elle-même mais trouvent également une utilité dans le planning préopératoire. Par exemple, ils peuvent incorporer des images obtenues par électrocardiographie ou par tomodensitométrie afin de déterminer la taille de valve appropriée à chaque patient dans le cadre de chirurgies cardiaques.

b. *L'utilisation de l'Intelligence Artificielle dans la personnalisation des traitements des maladies chroniques*

Dans le cadre du traitement des maladies chroniques, afin de déterminer la dose d'un médicament à administrer, les médecins suivent un protocole clinique. Ils débutent le traitement par une dose standard, désignée par le protocole en fonction des médicaments puis évaluent à intervalles réguliers l'effet du médicament sur le patient afin d'ajuster les doses en fonction des effets sur la maladie ainsi que des effets secondaires. Afin de gagner du temps et de limiter les effets secondaires, le développement d'algorithmes permettant de suggérer des doses en fonction du profil du patient pourraient s'avérer être une aide précieuse pour les médecins. Par ailleurs, il est estimé que près de 280 000 hospitalisations par an sont dues aux effets indésirables des médicaments aux Etats-Unis (130 000 en France) (Coulet et Jay, 2019).

Pour répondre à cette problématique, Nicolas Jay et Adrien Coulet ont mis au point avec leur équipe un algorithme capable de prédire quels patients nécessiteraient une dose de médicament plus faible que la dose standard. Ils se sont basés sur les données médicales électroniques de l'hôpital de Stanford. Ils sont parvenus à construire des profils de patients pour 34 médicaments différents pour des affections variées (insuffisance pulmonaire, symptômes cutanés, greffe de moëlle osseuse, ...) et à croiser ces informations avec le devenir de la prescription initiale. Concrètement, ils ont regardé si les doses avaient été diminuées ou augmentées par rapport à la prescription initiale. Sur les 34 médicaments, leur algorithme a été capable d'anticiper des réductions de dose pour 23 médicaments. En revanche, les résultats n'ont pas été satisfaisants en ce qui concerne les augmentations de dose.

Si cet algorithme français n'a pas eu des résultats aussi brillants qu'espérés, l'algorithme DeepSurv développé par Jared L. Katzman et al apparaît comme une référence parmi les algorithmes de prédiction de traitement. Cet algorithme permet non seulement de donner des recommandations de traitement en fonction des maladies sur lesquelles il est utilisé mais permet également de prévoir les pronostics des patients en fonction de leurs affections et de leurs caractéristiques (âge, sexe, antécédents médicaux, ...) (Katzman *et al.*, 2018). Le modèle de survie standard utilisé pour établir les modèles de survie est appelé le CPH (Cox Proportional Hazards model). Cependant, ce modèle est considéré comme trop simpliste pour certaines applications comme des recommandations de traitements personnalisés par exemple. DeepSurv est une amélioration de ce modèle et il a été démontré qu'il était meilleur dans la prédiction de survie des patients que les modèles standards mais aussi qu'il était capable de comprendre les relations entre les caractéristiques de patients et leur réponse aux traitements. Il est ainsi en mesure de proposer des traitements personnalisés en fonction des caractéristiques des patients. En appliquant les recommandations de DeepSurv, les médecins sont capables d'augmenter l'espérance de vie des patients.

L'intérêt de DeepSurv est triple. Il peut être appliqué à plusieurs maladies. En effet, en fonction des données utilisées pour l'entraîner, DeepSurv s'est montré plus performant que les modèles traditionnels pour des applications pour différents types de cancer par exemple. Enfin, en plus de pouvoir proposer des protocoles de traitement plus efficaces que les modèles traditionnels, sa capacité à prédire l'espérance de vie des patients en fonction de leurs caractéristiques permet aux médecins d'affiner leur proposition de traitement aux patients.

Un autre domaine dans lequel l'utilisation de l'Intelligence Artificielle permet de donner des recommandations en termes de traitement est la diabétologie. En effet, dans les suivis standards des cas de diabète, les doses d'insuline injectées sont calculées par les médecins. Cependant, une équipe de l'Université de Cambridge a développé un logiciel qui prédit les besoins en insuline des enfants atteints du diabète de type 1 et qui est basé sur l'Intelligence Artificielle (Gutierrez C, 2022). Ce système, CamAPS FX, mesure la glycémie des enfants tout au long de la journée et détermine la quantité d'insuline à injecter. CamAPS FX analyse les profils d'alimentation et d'activités des enfants et prédit leur taux glycémique. Selon un essai clinique publié le 20 janvier 2022 dans le *New England Journal of Medicine*, ce système permet de maintenir la glycémie entre les repas dans une fourchette normale pendant plus de temps que les dispositifs traditionnels. Ce système permet d'allonger la charge qui incombe aux patients et à leurs parents sur le long terme d'autant plus que ce dispositif possède un système de secours en cas de panne et évite ainsi que l'enfant ne se retrouve sans insuline en attendant que le problème soit résolu. Pour aller encore plus loin, depuis le 29 Septembre 2021, le système DBLG1 commercialisé par Diabeloop est inscrit sur la Liste des Produits et des Prestations Remboursables (LPPR) de l'Assurance Maladie (Paitraud, 2021). Ce système dédié à la gestion automatisée du diabète de type 1, communément appelé « pancréas artificiel », est composé de trois systèmes connectés par Bluetooth. Il est doté d'un système de mesure en continu du glucose interstitiel qui s'insère au niveau de l'abdomen dans le tissu sous-cutané. La valeur de glucose interstitiel est calculée toutes les 5 minutes. Il est relié à un système d'administration de l'insuline, c'est-à-dire une pompe avec tubulure externe. Grâce aux informations du capteur, le système d'administration de l'insuline calcule la dose à administrer. Le système gère tout comme « un vrai pancréas » (CHU LYON, 2022). Enfin, le système possède un système d'interface avec le patient et permet de visualiser les informations de glucose interstitiel et l'état du système, de contrôler la pompe à insuline, de centraliser les alarmes et de proposer des repas ou des activités physiques. Le système embarque de l'IA notamment pour calculer la dose d'insuline à administrer ainsi que pour proposer des heures de repas

ou d'activité. L'intérêt de cette solution est qu'elle fonctionne à l'aide de technologies embarquant du Machine Learning. L'idée est donc que l'outil s'adapte de mieux en mieux aux variations de glycémie du patient qui l'utilise. Medtronic a également développé un système qui fonctionne comme un pancréas artificiel. Il s'agit du MiniMed 780G. Ce système fonctionne sur le principe d'une boucle semi-fermée, le système ajuste automatiquement les niveaux d'insuline en fonction de la glycémie mesurée. On parle de boucle semi-fermée car le patient doit toujours réaliser certaines actions : validation de bolus, calibration de l'appareil, indications de resucrage ou d'activité physique.

c. L'utilisation de l'IA pour évaluer l'évolution de la maladie et remonter les alertes

La gestion sur le long terme des maladies chroniques est devenue de plus en plus maîtrisée par les pays Européens et Nord-Américains. La population vieillissant, les séniors sont de plus en plus touchés par de nombreuses maladies chroniques comme l'hypertension, le diabète, les maladies cardiovasculaires, etc. Le développement continu des technologies Big Data, IA, blockchain ainsi que les objets connectés font rentrer la médecine dans une nouvelle ère digitale. Cependant, selon le Dr Nicolas Postel-Vinay (médecin dans l'unité d'hypertension artérielle de l'hôpital Georges-Pompidou), les objets dont la fiabilité et l'intérêt pour la santé ont été prouvés cliniquement demeurent rares. De plus, il n'existe pas de registre ou de certification officielle permettant de les identifier. Afin d'être commercialisé au sein de l'UE, un dispositif médical doit disposer d'un marquage « CE », gage de sa conformité vis-à-vis des exigences définies dans les textes européens. Cependant, le marquage CE n'est en rien une certification de l'intérêt médical d'un dispositif. Chaque pays est donc libre de réguler la commercialisation des dispositifs médicaux et il n'existe pas de critères d'évaluation médicale de la valeur apportée par un dispositif à l'échelle européenne.

En France, c'est la Commission Nationale d'Évaluation des Dispositifs Médicaux et des Technologies de Santé (CNEDiMETS) qui examine la demande de prise en charge des dispositifs médicaux par l'Assurance Maladie. Le Comité Economique des Produits de Santé (CEPS) est chargé de fixer avec les fabricants un tarif de remboursement ou un prix limite de vente (HAUTE AUTORITE DE SANTE, 2011). Pour chaque dispositif médical, la CNEDiMETS évalue le service attendu en fonction de :

- l'intérêt du produit (son effet thérapeutique, diagnostique ainsi que les effets secondaires ou les risques liés à son utilisation). La place dans la stratégie thérapeutique est aussi analysée ;
- l'intérêt de santé publique (impact sur la santé de la population, sur le système de santé ou sur les politiques de santé publique) (HAUTE AUTORITE DE SANTE, 2012).

Ces outils servent essentiellement pour le diabète, l'hypertension artérielle ou les insuffisances cardiaques.

L'Intelligence Artificielle permet, dans le domaine des maladies cardiovasculaires par exemple, de suivre en temps réel l'évolution des signes cliniques, que ce soit à la maison ou dans une structure de soins. Parmi les outils que l'on peut utiliser chez soi et qui peuvent intégrer un système d'alerte, on retrouve le BPM Core de Withings. Cet outil est certifié dispositif médical et offre diverses fonctionnalités :

- Tensiomètre dont la précision a été validée cliniquement à +/- 3 mmHg ou 2% de la mesure de la fréquence cardiaque : il permet de mesurer la tension artérielle systolique et diastolique ;
- Electrocardiogramme qui recueille les données ECG et détecte les fibrillations auriculaires ;
- Stéthoscope électronique qui détecte les valvulopathies.

Cet outil peut permettre au médecin un meilleur suivi de l'hypertension chez les patients et permet de s'assurer de l'efficacité du traitement prescrit.

Il est cependant nécessaire de différencier ce type de dispositif des outils dont l'unique objectif est de collecter les données et de les envoyer ensuite à un professionnel de santé. Notamment dans le domaine de la cardiologie, il existe des outils de télésurveillance couplés à des défibrillateurs automatiques implantables qui

collectent les données et les envoient à intervalles réguliers à un professionnel de santé. Cependant, ces dispositifs ne possèdent pas de fonctionnalités d'analyse et ne remontent pas d'alertes, « juste » des données brutes.

3. Réglementations et normes autour de l'Intelligence Artificielle en médecine humaine

A. Réglementations en vigueur quant à l'utilisation de l'IA en médecine

Les logiciels et algorithmes d'Intelligence Artificielle fonctionnent grâce à l'exploitation des données. La réglementation majeure qui encadre l'accès et l'utilisation des données est le règlement général sur la protection des données (RGPD). Le RGPD est un texte réglementaire européen qui encadre le traitement des données de manière homogène sur tout le territoire de l'Union Européenne. Ce règlement est entré en vigueur en 2018.

D'après le Ministère de l'Economie, des Finances et de la Souveraineté Industrielle et Numérique, le RGPD s'inscrit dans la continuité de la loi française « Informatique et Libertés » de 1978 qui a posé les bases légales de l'utilisation des données sur le territoire français (Ministère de l'économie, des finances et de la souveraineté industrielle et numérique, 2019). Du fait qu'ils utilisent des données, les dispositifs médicaux utilisant l'Intelligence Artificielle se doivent d'être en conformité avec le RGPD. La Commission Nationale de l'Informatique et des Libertés (CNIL) donne des indications afin de s'assurer que de tels dispositifs suivent bien le RGPD (CNIL, 2022).

- Définition d'une finalité : Selon le RGPD, un système d'IA doit être développé, entraîné et déployé avec une finalité bien définie et précise s'il repose sur l'exploitation de données personnalisées. Cette finalité doit être définie en amont du projet et doit respecter certaines clauses : elle doit être compatible avec les missions de l'organisme qui le développe ainsi que connue et compréhensible ;
- Etablissement d'une base légale : Un système d'IA qui exploite des données personnelles ne peut être mis en œuvre que s'il correspond à une justification légale : le consentement, le respect d'une obligation légale, l'exécution d'un contrat, l'exécution d'une mission d'intérêt public, la sauvegarde des intérêts vitaux et la poursuite d'un intérêt légitime. Cette base légale doit être choisie en amont du développement du projet et aura une implication sur les obligations de l'organisme qui développe le système. Il est intéressant de noter que la recherche scientifique ne peut pas constituer en soi une base légale du traitement ;
- Constituer une base de données : La constitution des bases de données est une étape importante pour le développement des systèmes d'IA seulement cette constitution ne peut se faire au détriment du droit des personnes concernées par ces données. Les personnes ont un droit à l'information qui peut avoir lieu préalablement à la collecte des données ou dans un délai d'un mois après la réception des bases par les tiers. La constitution des bases de données peut se faire via la collecte spécifique de données personnelles à cette fin, soit via la réutilisation de données collectées pour une autre finalité, mais se pose alors la question de la compatibilité des finalités et des conditions dans lesquelles la base initiale a été constituée ;
- Minimisation des données : Le principe de minimisation des données correspond au fait que les données collectées doivent être adéquates, pertinentes et limitées à ce qui est nécessaire à l'accomplissement de l'objectif défini. Cette attention particulière doit être renforcée dans le cadre des

données sensibles. La quantité des données nécessaires doit aussi être calculée au regard de l'objectif du traitement des données (principe de proportionnalité). Dans le cas des systèmes d'IA, les contraintes sont plus fortes en phase de production qu'en phase d'apprentissage. Par exemple, il peut être nécessaire de resserrer la typologie des données personnelles aux seules qui se sont avérées utiles pendant la phase d'apprentissage ;

- Définition d'une durée de conservation : Le RGPD demande qu'une durée soit définie au bout de laquelle les données devront être supprimées ou archivées. La durée de conservation est définie en fonction de l'objectif de l'analyse. Les systèmes d'IA peuvent nécessiter la disponibilité des données sur des périodes plus longues que des analyses de données classiques. Le RGPD tient compte de cette spécificité dans le cadre du principe de proportionnalité (performance du système ou dans des cas de recherche scientifique par exemple) ;
- Encadrement de l'amélioration en continu : Les systèmes d'IA peuvent être amenés à passer par des phases de réapprentissage afin d'améliorer leurs performances. Dans ce cas, des données utilisées lors de la phase de production sont réutilisées pour cette phase de réapprentissage. Il faut dans ce cas se poser la question de la finalité des deux utilisations (fournir une information et servir au réapprentissage) et de la protection des données. Le RGPD est flou par rapport à cette notion mais il est nécessaire de se poser cette question afin d'éviter tout litige juridique ;
- Prémunition quant aux risques liés aux modèles d'IA : Les systèmes d'IA sont susceptibles d'être attaqués par des processus spécifiques (inférence d'appartenance, attaque par exfiltration de modèle, ...). Par ailleurs, des études ont montré que les systèmes d'IA pouvaient « mémoriser » des informations traitées lors de la phase d'apprentissage. Les concepteurs de systèmes doivent donc mettre en œuvre des mesures afin de minimiser les risques de fuites d'informations. Ainsi, un modèle d'IA peut être considéré comme une donnée personnelle. En cas d'attaque réussie, il faut procéder au retrait du modèle et notifier de cette fuite auprès de l'autorité de protection des données compétente ;
- Assurance de l'information et de l'explicabilité : Le RGPD exige que toute information ou communication relative au traitement des données personnelles soit concise, transparente, compréhensible et aisément accessible. La complexité et l'opacité des systèmes d'IA peut rendre difficile l'explication de l'utilisation des données aux personnes concernées. Il peut être possible de déroger à ce droit à l'information s'il est démontré que l'information des personnes concernées exige des efforts disproportionnés par rapport aux résultats attendus ;
- Possibilité de mettre en œuvre l'exercice des droits : Toute personne concernée par un traitement de ses données personnelles dispose de droits dans le but de pouvoir garder la maîtrise de leurs données. L'organisme qui collecte les données doit expliquer aux personnes concernées comment exercer leurs droits inhérents au traitement de leurs données personnelles. Ces droits s'appliquent aussi bien dans le cadre de la phase d'apprentissage que dans le cadre de la phase de production ;
- Encadrement de la prise de décision automatisée : Parmi les droits dont disposent les individus dont les données sont exploitées, on trouve le droit de ne pas faire l'objet d'une décision entièrement automatisée si cela a un effet juridique ou l'affecte sensiblement. Ainsi, le propriétaire du système doit prévoir une possible intervention humaine pour permettre à la personne concernée par l'analyse de

bénéficier d'un réexamen de sa situation ou d'obtenir une explication sur la décision prise par le système ;

- Evaluation du système : L'évaluation du système est nécessaire afin de valider l'approche testée lors de la phase d'apprentissage, de minimiser les risques de dérive du système au cours du temps et de s'assurer de la satisfaction des besoins opérationnels pour lesquels un système a été conçu. Le propriétaire du système doit donc mettre en place des mesures d'évaluation du système et les adresser à la CNIL contenant des mesures objectives de performance, une analyse des erreurs du système, une analyse des risques de biais entraînés par l'utilisation du système... ;
- Evitement des discriminations algorithmiques : Les systèmes d'IA peuvent entraîner des risques de discrimination à cause des données utilisées pour l'apprentissage ou à cause de faille de conception. Afin de respecter le RGPD, les propriétaires des systèmes doivent présenter à la CNIL les mesures mises en place afin d'éviter que ces risques ne se produisent.

Le RGPD est la réglementation élémentaire en termes de contrôle de l'utilisation des données mais elle n'est pas spécifique aux systèmes d'IA. En 2021, une réglementation Européenne, le AI Act a été proposé à la Commission Européenne avec les 4 objectifs suivants (COMMISSION EUROPEENNE, 2021) :

- S'assurer que les systèmes d'IA mis sur le marché commun respectent les lois sur les droits fondamentaux des citoyens de l'UE (Union Européenne) ;
- S'assurer de créer une base légale solide afin de favoriser les investissements dans l'IA au sein de l'UE ;
- Promouvoir la bonne application des mesures prévues par les lois relatives aux droits fondamentaux des citoyens de l'UE ;
- Faciliter le développement du marché de l'IA au sein de l'UE et éviter la fragmentation des marchés.

L'AI Act n'est pas une proposition centrée sur les applications médicales mais tout système d'IA, quel que soit son utilisation finale (et donc les systèmes d'IA avec des finalités relatives au monde de la santé) devront respecter cette loi si elle venait à être votée.

B. Questionnements éthiques autour de l'utilisation des données de santé en médecine

Avec l'émergence des outils utilisant l'Intelligence Artificielle, de nombreuses initiatives ont vu le jour afin d'émettre des recommandations pour promouvoir un usage raisonné et fiable de ces outils. Parmi ces initiatives, on peut citer AI4People (Floridi *et al.*, 2018), Asilomar AI principles (FUTURE OF LIFE INSTITUTE, 2017), ou encore la Déclaration de Montréal pour un Développement Responsable de l'Intelligence Artificielle (Université de Montréal, 2019). La Commission Européenne a également mandaté un groupe d'experts en Intelligence Artificielle afin d'étudier les questions liées à l'éthique. Ce rapport ne se focalise pas sur le domaine de la santé mais se veut généraliste afin d'englober l'ensemble des secteurs et utilisations possibles de l'Intelligence Artificielle.

L'objectif de ce groupe d'experts (High-Level Expert Group on Artificial Intelligence, 2019) est de fournir un cadre pour permettre le développement de l'IA de façon « robuste » et éthique. Selon ce groupe, les principes éthiques auxquels le développement de l'IA doit adhérer sont : respect de l'autonomie humaine, la prévention de tort, l'équité et l'explicabilité. Cependant, ces principes peuvent devenir contradictoires et que dans ce cas, il est nécessaire que les décisions se prennent de façon démocratique et que l'avis de la société civile soit

écouté et pris en compte. Par ailleurs, le groupe insiste sur le fait qu'une attention particulière doit être portée aux situations qui incluent des groupes vulnérables (enfants, personnes présentant un ou des handicaps, groupes ethniques et sociaux historiquement défavorisés, ...) ou dans lesquelles il existe une asymétrie de pouvoir ou d'information. Enfin, si l'utilisation de l'IA présente de nombreux intérêts, elle présente également de nombreux risques qui peuvent être difficiles à anticiper, identifier et mesurer. Il est donc nécessaire d'adopter des mesures afin de gérer ces risques.

A partir des 4 considérations éthiques identifiées et citées plus haut, le groupe d'experts mandaté par la Commission Européenne fournit des recommandations que les systèmes d'IA devraient suivre afin d'être dignes de confiance.

Parmi ces recommandations, on trouve : le contrôle par des opérateurs humains, la sécurité technique, la gestion réfléchie des données et l'intimité, la transparence, la diversité, l'équité, le bien-être sociétal et environnemental et la responsabilité. Afin d'implémenter ces recommandations, les moyens peuvent être techniques ou non.

Le groupe insiste sur la nécessité de mettre à jour régulièrement cette liste en faisant appel à de nouvelles générations d'expert afin de rester proche des besoins de la société. De même, la formation des parties prenantes ou affectées par des systèmes d'IA est primordiale afin que tous soient au fait des bénéfices mais également des risques que ces systèmes impliquent.

L'intérêt de ce rapport est qu'il se veut global et international (les rédacteurs appellent d'ailleurs à une réflexion à l'échelle mondiale sur les potentiels risques de l'IA). Les considérations soulevées par ce rapport sont similaires à celles que l'on peut retrouver au sein d'autres institutions internationales comme l'Organisation des Nations Unies. Cependant, les limites de ce rapport se situent justement au niveau de la généralité des considérations soulevées. En restant général, le rapport ne peut prétendre proposer des solutions pour des situations concrètes et spécifiques. Certains auteurs se sont donc penchés sur la façon d'appliquer les conclusions du rapport de groupe d'experts au domaine de la santé.

Kristine Bærøe et son équipe ont mis en avant cinq autres considérations éthiques à prendre en compte dans le cadre de l'IA en santé (Bærøe *et al.*, 2020). Ces cinq considérations sont les suivantes : des objectifs contradictoires, des contextes d'application inégaux, des risques et des incertitudes particulières, des coûts d'opportunité et des déficits démocratiques.

- Des objectifs contradictoires : Le secteur de la santé peut être affecté par des forces contradictoires liées à son importance politique, économique et sociale. Par exemple, la question de l'accès au soin universelle permet à l'intelligence artificielle de se démocratiser mais les profits issus de son exploitation ne sont pas redistribués équitablement mais seront tournés vers l'industrie de l'Intelligence Artificielle (majoritairement dans des pays développés). Cela mène à toujours davantage d'accumulation de richesse dans les pays riches, au maintien des inégalités de niveau de vie et par conséquent des inégalités en termes de santé. La décision de fin de vie peut être un autre exemple d'objectif contradictoire (les objectifs du patient et de sa famille peuvent être contradictoires avec les objectifs de la structure de soins). On comprend donc la nécessité d'établir des recommandations adaptées au secteur de la santé ;
- Des contextes d'application inégaux : En fonction du contexte économique des pays, les standards acceptables pour l'Intelligence Artificielle peuvent être inégaux. Dans le cadre des pays riches par exemple, il est possible que la mise en pratique de certains outils puisse dégrader le soin aux patients (en termes de relations aux patients par exemple) du fait que la main d'œuvre médicale soit bien formée et qu'elle fasse face à des problèmes de santé relativement connus maintenant. Ainsi, les standards d'acceptation d'un algorithme seraient très élevés pour être certain d'apporter un bénéfice au patient. Dans les pays plus pauvres en revanche, le personnel médical est souvent moins bien formé et les critères d'acceptation d'un algorithme pourraient être beaucoup plus bas. On comprend

donc l'importance d'établir des normes qui soient adaptables aux réalités du terrain dans les différents pays ;

- Des risques et des incertitudes particulières : Du fait du caractère personnel de la santé, la mise en place et l'accessibilité de données sensibles peuvent constituer un risque à l'avenir. Si l'on prend comme exemple la surveillance des personnes atteintes de démence, on peut imaginer que la surveillance de leurs mouvements puisse être un pas en avant vers la surveillance à plus grande échelle de la société à l'avenir par l'amélioration des performances des algorithmes. De manière générale, lorsque les personnes sont atteintes de pathologies graves, elles sont plus susceptibles d'accepter des outils qui pourraient présenter des risques à l'avenir, surtout si ces risques sont potentiels et incertains. Ainsi, dans le monde de la santé, le caractère personnel des données et l'importance des enjeux doit mener à une réflexion encore plus approfondie des potentiels risques que l'implémentation d'outils d'Intelligence Artificielle pourrait entraîner ;
- Des coûts d'opportunité : Il y a des coûts d'opportunité à considérer en cas de choix d'implémentation d'algorithmes aux dépens d'opérateurs humains. En effet, en implémentant des algorithmes, le jugement humain se relocalise en amont de la chaîne de commandement puisque des opérateurs humains doivent être en mesure de superviser les décisions prises. Cependant, plus le niveau technique pour comprendre les décisions prises augmente et plus les concepteurs d'algorithmes deviendront l'autorité compétente pour trancher en cas de décision ambiguë. La croissance du rôle des concepteurs se fera au détriment du rôle du personnel médical ;
- Des déficits démocratiques : Le groupe d'experts de l'UE a insisté sur le fait que les démocraties développées devaient administrer le développement de l'Intelligence Artificielle. Cependant, d'après Kristine Bærøe et son équipe, la façon d'organiser le système de santé de manière générale n'était pas claire dans tous ces pays. Comment organiser correctement le développement de l'IA dans le domaine de la santé dans ce cas ? On voit donc une limite au rapport d'expert de l'UE. Par ailleurs, on sait que l'Etat français a demandé à Microsoft de gérer le stockage des données destinées à être intégrées à l'ENS. La gestion par une entreprise étrangère des données de santé des patients français peut soulever des questions notamment celle de la responsabilité de Microsoft et de l'applicabilité des lois françaises sur celle-ci.

On peut donc observer que les risques liés à l'utilisation des données ont été intégrés dans les textes réglementaires au niveau européen et national. En effet, le RGPD garantit une bonne utilisation des données personnelles des patients afin de limiter les risques de mauvaises utilisations. Cependant, il existe une nécessité de renforcer le contrôle de l'utilisation de l'IA et particulièrement dans le domaine de la santé. Si l'IA Act ouvre la voie à une réglementation spécifique aux logiciels d'Intelligence Artificielle, il apparaît que d'autres considérations éthiques doivent être étudiées afin de garantir la sécurité des patients mais aussi de la société de manière générale. Ces considérations sont encore trop peu considérées et doivent faire l'objet d'études afin de permettre un développement raisonné et cohérent de l'IA en santé.

En conclusion de cette partie, nous pouvons remarquer, que même en médecine humaine, l'utilisation de technologies basées sur l'Intelligence Artificielle sont loin d'être complètement exploitées. Dans un premier temps, on remarque qu'il n'existe pas de normes internationales ou reconnues permettant de s'assurer de la qualité des données. Qu'il s'agisse des données utilisées pour entraîner les modèles ou des données dont on attend une analyse, il est nécessaire de pouvoir s'assurer de la qualité des données particulièrement avec des

enjeux aussi importants que la santé. Par ailleurs, les bases de données elles-mêmes sont éparpillées du fait de la pluralité des acteurs qui les produisent mais aussi parce qu'il est difficile de tout agréger au niveau national voire mondial. S'il n'existe pas de normes consensuelles mondialement pour la qualité des données, il n'en existe pas non plus de régulation globale quant à l'utilisation de ces données. Certaines organisations comme l'UE ont adopté des textes de lois mais ces textes ne sont respectés que par ses signataires. D'ailleurs, il n'existe pas encore de textes réglementaires s'appliquant spécifiquement à l'IA. Ainsi, l'utilisation de l'IA n'est pas encore suffisamment encadrée pour pouvoir être utilisée à grande échelle. Enfin, d'un point de vue médical, il existe de nombreuses initiatives notamment dans les centres de recherches mais peu d'outils sont utilisés en routine. La démocratisation des systèmes d'IA n'est pas encore une réalité sauf dans le domaine de l'imagerie où son utilisation s'étend. Cependant, les potentiels de l'IA sont très importants et il est probable que celle-ci devienne de plus en plus connue et maîtrisée par le personnel de santé. Des initiatives comme AI for Health portée par Startup Inside tendent à faire connaître les utilisations possibles de l'IA et poussent à un développement raisonné des outils basés sur l'IA.

Maintenant que l'on a vu l'écosystème IA autour de la santé humaine, intéressons-nous à l'écosystème IA en santé vétérinaire afin de voir le potentiel de cette technologie en santé animale et les difficultés de sa diffusion au sein de la pratique vétérinaire.

Partie 3 : Situation actuelle du développement de l'Intelligence Artificielle en santé animale et axes de développement

1. Etat de l'art de l'utilisation de l'IA en santé animale appliquée aux maladies chroniques chez les animaux de compagnie

L'objectif de cette première thématique n'est pas de dresser une liste exhaustive de tous les algorithmes ou solutions développées par des équipes de chercheurs. Mais il s'agit de montrer que des projets existent dans différentes spécialités (cardiologie, endocrinologie, oncologie, ...) mais dans des proportions bien moindres qu'en médecine humaine.

A. Utilisation de l'IA dans la prédiction de maladies chroniques chez les animaux de compagnie

La Maladie Rénale Chronique (MRC) est une maladie progressive qui touche essentiellement les chats âgés. Elle se caractérise par une azotémie persistante ainsi que par une diminution de la capacité de concentration des urines de l'animal. Elle résulte d'une perte de néphrons fonctionnels menant à une baisse du taux de filtration glomérulaire. Le diagnostic précoce de cette maladie est difficile car la relation entre le taux de filtration glomérulaire et les marqueurs du fonctionnement rénal ne sont pas linéaires. Au début de la maladie, lorsque les reins sont encore peu endommagés, l'adaptation des néphrons restants fait que les marqueurs de filtrations glomérulaires évoluent peu. Vincent Biourge et son équipe ont essayé de développer un algorithme basé sur l'utilisation de l'IA. Ils ont mis en place différents modèles afin de maximiser soit la spécificité et la sensibilité soit la spécificité et la valeur prédictive positive dans la prédiction de développement d'une MRC chez les chats. Ils ont réussi à mettre au point un système qui, sur la seule base d'une analyse du taux de créatinine plasmatique, du taux de concentration d'azote uréique et de la densité urinaire au cours d'une seule visite prédit la probabilité d'apparition d'une MRC chez les chats de plus de 7 ans dans les 12 mois (Biourge *et al.*, 2020). Les résultats de l'algorithme donnent une spécificité de 87%, une sensibilité de 70%, une Valeur Prédictive Positive de 53% et une Valeur Prédictive Négative de 92% à 12 mois sur des chats dont les résultats étaient dans la norme. Ces résultats montrent que l'algorithme est loin d'être infaillible mais peut donner une première idée au vétérinaire praticien de la probabilité de l'animal à développer une MRC à l'avenir. Il peut notamment l'encourager à chercher une cause sous-jacente au développement d'une MRC alors même que les valeurs de l'animal sont dans les normes. Greene et son équipe ont montré qu'une perte de poids pouvait être un facteur indicateur du développement d'une MRC même s'il ne s'agit pas d'un marqueur systématique (Greene *et al.*, 2014). En croisant des mesures régulières de poids et l'utilisation de cet algorithme, les vétérinaires seraient en mesure de dépister plus précocement une MRC chez un patient. Plus intéressant encore, cet algorithme se base sur des valeurs obtenues sur un seul rendez-vous, et donc indépendamment de données historiques (on peut imaginer de multiples raisons pour lesquelles le vétérinaire n'aurait pas les mesures de poids précédentes de l'animal). Cela se prête donc bien à une pratique quotidienne au cours de laquelle les vétérinaires n'ont pas toujours accès à l'historique de l'animal ou s'il s'agit d'une première visite.

Richard Bradley et son équipe ont développé un logiciel similaire dont le but est également de prédire l'apparition d'une MRC. Il diffère légèrement dans le sens où il prend les mêmes paramètres en entrée mais en y ajoutant l'âge. Ils obtiennent une sensibilité de 63% pour une prédiction à 12 mois et de 44,2% pour une prédiction à 24 mois. En revanche, dans les deux cas, la spécificité reste aux alentours de 99% (Bradley *et al.*, 2019).

Ces deux algorithmes ont été développés dans le cadre d'une étude mais ils montrent que s'ils étaient disponibles, ils apporteraient de la valeur ajoutée au vétérinaire en orientant vers des examens complémentaires, en mettant en place un traitement préventif. L'étude menée par Richard Bradley a permis de développer un outil, le RenalTech™ développé par WALTHAM Petcare Science Industry, appartenant à Mars Animal Health. L'institut déclare que la précision de son outil est supérieure à 95% pour une prédiction à 24 mois (WALTHAM, 2019).

En termes de relation client, cela permettrait aussi d'avertir les propriétaires de l'éventualité de cette maladie qui peut être coûteuse pour les propriétaires afin qu'ils puissent s'organiser et se préparer à ces

dépenses à venir. La relation entre le praticien et le propriétaire en serait améliorée également. Ceci confère un avantage compétitif important aux cliniques du groupe Mars équipées de cette technologie.

Il existe d'autres exemples chez le chien. On peut par exemple citer l'équipe d'Imogen Schofield qui a mis au point un algorithme de prédiction de survenue de la maladie de Cushing (Schofield *et al.*, 2021). Dans cette étude, la sensibilité est de 71% et la spécificité de 82%, ce qui donne une performance globale de l'algorithme mesurée par l'AUROC de 85%.

B. Utilisation de l'IA dans le diagnostic de maladies chroniques chez les animaux de compagnie

Même s'il existe des dizaines d'études sur le potentiel des algorithmes d'IA pour le diagnostic chez les animaux de compagnie, nous avons choisi d'en présenter deux, relevant de deux spécialités différentes. La première porte sur le diagnostic de la myocardiopathie hypertrophique chez le chat tandis que la seconde porte sur l'hypoadrénocorticisme chez le chien.

La cardiomyopathie hypertrophique est un syndrome commun chez le chat. On estime qu'il concerne 10% à 15% des chats adultes (Payne *et al.*, 2015). Les symptômes associés sont des difficultés respiratoires, un souffle cardiaque ainsi que de l'apathie. Cependant, cette affection peut également provoquer une mort subite de l'animal et fait même partie des 10 causes les plus courantes de mort subite chez le chat. Anatomiquement, cette affection provoque un épaississement du myocarde, les muscles cardiaques. La forme la plus courante de la cardiomyopathie hypertrophique provoque une augmentation de la taille de la paroi du ventricule gauche conduisant à une diminution de la compliance du myocarde. Cette diminution de compliance du myocarde empêche le cœur de se relâcher et de se remplir correctement durant la diastole. Progressivement, cela entraîne une dilatation de l'atrium gauche. Il existe plusieurs examens permettant d'orienter le diagnostic mais un examen radiographique associée à une échographie cardiaque constitue la méthode de diagnostic de référence. Jinhyung Rho et son équipe ont construit 5 modèles différents d'IA à l'aide d'une base de données de 231 images radiographiques selon une orientation ventro-dorsale, c'est-à-dire que l'animal est positionné sur le dos (Rho *et al.*, 2023). Leur modèle le plus abouti montre une sensibilité de 100% et une spécificité de 91% environ.

L'intérêt de cette approche est que le diagnostic de certitude de la myocardiopathie hypertrophique incombe généralement à un vétérinaire spécialiste. En effet, cela demande une expertise en imagerie ainsi que d'un équipement coûteux. En routine en revanche, l'utilisation de la seule radiographie couplée à l'auscultation permettrait au vétérinaire généraliste d'orienter vers ce diagnostic et donc de diriger l'animal vers un spécialiste pour le diagnostic de certitude et la prise en charge.

L'hypoadrénocorticisme chez le chien est le plus souvent causé par une destruction immune du cortex surrénalien. Cela déclenche donc une diminution de la production d'hormones surrénaliennes : les glucocorticoïdes et les minéralocorticoïdes. Les présentations de cette affection sont diverses, les symptômes étant souvent vagues et non spécifiques. Ils peuvent être confondus avec des maladies rénales ou gastrointestinales, des insuffisances hépatiques, ... Certains marqueurs peuvent également alerter comme une hyponatrémie ou une azotémie associée à de l'hypoalbuminémie, un faible taux d'aldostérone, une hypercholestérolémie, une hypoglycémie, ... Cependant, même pour un chien souffrant effectivement d'hypoadrénocorticisme, ces marqueurs peuvent être dans les normes. Reagan et son équipe ont développé un algorithme permettant de diagnostiquer cette affection (Reagan *et al.*, 2020). Ce modèle a montré une sensibilité de 96,3% et une spécificité de 97,2%. Il a été démontré que cet algorithme utilisant seulement une Numération Formule Sanguine et des analyses biochimiques complètes accessibles par des analyseurs classiques permettait de trier efficacement les chiens atteints de cette affection. Dans cet exemple, les facteurs limitants ne sont ni techniques ni financiers mais proviennent plutôt de la présentation de la maladie, souvent

non spécifique. En revanche, les analyses requises par cet algorithme sont simples et accessibles à tous les vétérinaires, d'autant plus que des données historiques ne sont pas nécessaires.

Une autre implication de l'IA dans le diagnostic dont nous avons peu parlé jusqu'à présent est l'apport de cette technologie dans la construction des outils diagnostiques. Pour cela, nous pouvons citer l'analyseur VETSCAN IMAGYST développé par Zoetis. Le système VETSCAN IMAGYST est doté de trois composants : un outil de préparation d'échantillon, un scanner d'analyse et un logiciel d'analyse. Le logiciel d'analyse embarque des algorithmes d'IA et permet de faire trois types d'analyses : des analyses de frottis sanguin, des analyses coproscopiques et des analyses cytologiques. Yoko Nagamori et son équipe ont cherché à étudier les performances du VETSCAN IMAGYST System ainsi que de comparer sa technique de préparation d'échantillons avec les méthodes de flottaison traditionnelles. Dans le cadre de cette thèse, seule l'analyse des performances du logiciel d'analyse nous intéressera.

Les analyses coproscopiques font partie des examens de routine pour les vétérinaires. On y cherche certains éléments parasitiques tels que les œufs, les oocystes, les larves, ... Cependant, le résultat dépend beaucoup de la qualité de la préparation et des compétences de l'opérateur.

Yoko Nagamori et son équipe ont comparé les résultats de l'analyse donnée par le système VETSCAN IMAGYST avec un consensus de trois parasitologues. Selon eux, les résultats du système sont proches du résultat des parasitologues (Nagamori *et al.*, 2020). Le système ne reconnaît encore aujourd'hui qu'un nombre limité de parasites mais ses fonctionnalités sont appelées à se développer. De plus, ce système enregistre les images utilisées pour faire l'analyse, ce qui permet de les envoyer à des parasitologues confirmés en cas de besoin. Cela permettrait un gain de temps et d'efficacité pour les opérateurs et les vétérinaires pour une utilisation en routine étant donné que cet analyseur est à destination des cliniques vétérinaires.

C. Utilisation de l'IA pour le traitement de maladies chroniques chez les animaux de compagnie

Un dernier aspect que nous pouvons aborder est le traitement des maladies chroniques. Pour illustrer ce point, nous pouvons prendre l'exemple de la chimiothérapie. Zach Bohannon et son équipe ont essayé de prédire la réponse au traitement de chiens atteints d'un lymphome à l'aide de l'analyse du profil d'expression des antigènes des cellules tumorales ainsi que de la réponse *ex-vivo* de cellules cancéreuses vis-à-vis d'agents de chimiothérapie. Cette prédiction a été faite grâce à l'utilisation d'un algorithme d'IA.

Le lymphome fait partie des cancers les plus courants chez le chien et sa forme la plus courante est le lymphome multicentrique (environ 75% des lymphomes). Pour traiter ce type de cancer, le traitement de choix est la chimiothérapie en utilisant plusieurs agents anti-cancéreux. On note l'utilisation de cyclophosphamide, de doxorubicine, de vincristine, de prednisone, ... L'utilisation conjointe de ces molécules résulte en une rémission clinique comprise entre 73% et 96% et des durées de survie comprises entre 275 et 344 jours en fonction du grade du cancer (Curran et Thamm, 2016). Cependant, afin d'assurer une meilleure prise en charge, l'utilisation de traitements personnalisés au patient apparaît comme essentielle.

Bohannon et son équipe ont montré que leur algorithme présentait de bons résultats en termes de prédiction de réponse au traitement chez les chiens atteints d'un lymphome pour des traitements mono-agents (Bohannon *et al.*, 2021). En identifiant les molécules les plus à même d'induire une réponse positive au traitement, ils considèrent que cela permettrait aux vétérinaires de faire de façon plus routinière des recommandations de traitement personnalisées et donc *in fine* de réduire le poids du traitement pour les propriétaires et d'améliorer ainsi les durées de survie des animaux atteints.

Ce que l'on peut noter de manière générale au travers de ces exemples est qu'il existe des études démontrant l'efficacité d'outils à base d'IA dans la gestion des maladies chroniques chez les animaux de compagnie. Cependant, ces solutions restent pour la grande majorité à l'échelle de l'expérimentation et ne sont pas utilisables par les vétérinaires. Il s'agit davantage d'algorithmes qui démontrent le potentiel des technologies basées sur l'IA dans la gestion des animaux.

Ainsi, le potentiel de l'IA en santé animale pour la gestion des maladies chroniques des animaux de compagnie est certain. Quelles sont les solutions effectivement présentes sur le marché et quel est leur niveau de pénétration sur le marché des cliniques vétérinaires en France ?

2. Les solutions disponibles embarquant de l'IA pour les vétérinaires dans la gestion des maladies chroniques des animaux de compagnie

A. Outils disponibles dans le diagnostic des maladies chroniques des animaux de compagnie

L'outil le plus abouti en ce qui concerne le diagnostic en médecine vétérinaire est PicoxIA. Il s'agit d'un outil de lecture des radiographies vétérinaires développé par Emilie Boissady et Aloïs de La Comble. Il dispose de deux composantes : PicoxIA Acquisition® et PicoxIA Intelligence®. PicoxIA Acquisition® permet d'aider le vétérinaire à réaliser des radiographies de qualité optimale pour faciliter le diagnostic. Il permet aussi d'embarquer PicoxIA Intelligence® : outil qui analyse les images produites et identifie les lésions détectées. C'est cette partie de la solution qui comporte des algorithmes d'IA (PICOXIA, 2023). Les performances de cette solution ont été publiées. Un coefficient de corrélation intraclasse a été utilisé afin de comparer le degré d'accord entre les résultats donnés par l'algorithme et un consensus de deux radiologues à la lecture d'un même cliché radiographique. Les études ont été menées sur des clichés radiographiques dans le but de quantifier l'élargissement du cœur chez le chat et le chien. Les coefficients de corrélation sont tous les deux de 0,998 (Boissady *et al.*, 2021). Selon les auteurs, l'utilisation d'un tel algorithme en routine permettrait de limiter la variabilité de lecture entre les professionnels de santé et permettrait à terme d'augmenter la base de données servant à contrôler les performances de l'algorithme. Cela permettrait également à des vétérinaires de diagnostiquer plus facilement des atteintes telles que l'élargissement cardiaque.

Les régions anatomiques analysées sont l'abdomen (avec 13 lésions détectées), le thorax (avec 16 lésions détectées) ou les hanches (avec 5 lésions détectées). Cet outil est effectivement utilisé en clinique avec près de 1700 utilisateurs réguliers de l'outil et plus de 300 000 radiographies analysées au 13/04/2023 (PICOXIA, 2023). L'intérêt de cet outil est qu'il peut s'interfacer avec des logiciels de gestion de clientèle vétérinaires (AssistoVet et VetoPartner) et donc de diminuer la charge de travail supplémentaire du vétérinaire en complétant directement les informations dans l'espace dédié à l'animal. PicoxIA n'est pas le seul outil embarquant de l'IA permettant d'analyser des radiographies. SignalPet, une entreprise américaine propose également d'analyser les clichés des animaux de compagnie mais dispose d'un panel plus large que celui de PicoxIA (SignalPet, 2023).

Comme nous avons pu le voir précédemment avec le VETSCAN IMAGYST, certains analyseurs embarquent des algorithmes d'IA afin d'améliorer le processus diagnostique à l'image du SediVue Dx développé par IDEXX. Concrètement, l'analyseur utilise un algorithme pour analyser les échantillons. Par exemple, le SediVue DX alerte sur la présence ou l'absence de bâtonnets ou de coques de bactéries présentes dans les urines mais également les cellules présentes (cellules épithéliales, hématies, ...), les cristaux présents (struvite, biurate d'ammonium, ...) et les cylindres urinaires. Contrairement à l'exemple précédent, l'objectif ici n'est pas d'assister le vétérinaire dans l'interprétation des résultats mais plutôt de lui fournir des informations plus rapidement pour minimiser le temps passé sur l'analyse des échantillons. Cela permet de se concentrer davantage sur les patients (IDEXX CANADA, 2023).

Ces analyseurs permettent un gain de temps mais peuvent aussi se montrer plus efficaces que d'autres outils en embarquant des algorithmes d'IA. Toujours chez IDEXX, le ProCyte OneMC utilise des algorithmes d'IA afin de donner une Numération Formule Sanguine plus précise afin que des fragments n'interfèrent pas dans le décompte par exemple. Selon l'entreprise, cela permet d'obtenir des résultats plus fiables et d'alerter le praticien en cas d'anomalies. Il n'existe cependant que peu d'études attestant de l'efficacité de ces analyseurs de nouvelle génération.

B. Outils non spécifiques utilisant l'IA pour aider les praticiens dans leur pratique quotidienne

En pratique, d'autres outils non spécifiques permettent d'aider les vétérinaires dans leur pratique quotidienne. En France, on peut citer l'exemple de ZAG, un outil numérique d'aide à la décision en médecine vétérinaire. ZAG propose 11 fonctionnalités :

- Hypothèses diagnostiques ;
- Fiches informatives sur les affections ;
- Axes thérapeutiques ;
- Proposition d'examens complémentaires ;
- Accès à des consensus et des publications scientifiques de référence ;
- Proposition de tests génétiques pour les races présentant des prédispositions ;
- Informations sur le statut sanitaire des affections (zoonose, déclarations, ...) ;
- Guide pratique de l'utilisation des fluides (calcul de besoin hydrique, supplémentation, ...) ;
- Informations importantes sur les médicaments ;
- Guide sur le degré d'urgence des affections ;
- Accès à des sources bibliographiques en lien avec les affections.

Certaines de ces fonctionnalités comme les hypothèses diagnostiques utilisent une technologie appelée « système expert » pour émettre des recommandations ou donner des résultats (PRONOZIA, 2023). Si cette technologie n'est pas de l'IA à proprement parler, elle s'en approche. Le système peut répondre à des questions et construire un raisonnement à partir de faits et de règles connues. S'il y a tout de même un apprentissage des règles et des faits (reposant bien sur des bases de données), cet apprentissage est différent de ce qu'on peut voir en Deep Learning par exemple. Pour répondre à un problème, l'outil prend en compte l'anamnèse, les commémoratifs fournis par le vétérinaire, les symptômes, ... C'est la pertinence des informations fournies qui permettra au logiciel de donner une réponse pertinente et ces logiciels ne sont donc pas voués à remplacer les vétérinaires (Driot, 2023). Selon le fondateur de ZAG, il aurait été beaucoup plus facile de développer cette solution dans des pays où l'accès aux données est plus aisé, comme aux Etats-Unis.

Cette solution était disponible sous la forme d'un abonnement mensuel pour les vétérinaires jusqu'à fin 2021 puis a été rachetée par l'entreprise Dômes Pharma, holding qui détient le laboratoire pharmaceutique TVM par exemple. La diffusion de cette technologie a été très lente initialement, témoin des difficultés de diffusion au sein de la profession. Depuis le rachat par Dômes Pharma, la diffusion s'accélère grâce au travail des équipes commerciales.

L'application Sofie de LifeLearn est en revanche toujours disponible sur le marché et utilise le logiciel d'IA Watson développé par IBM. Concrètement, Sofie effectue des recherches en langage naturel afin de fournir la réponse la plus pertinente possible à une recherche du vétérinaire. Sofie s'appuie sur des manuels reconnus par la communauté scientifique et non disponible gratuitement à l'image du *Textbook of Veterinary Internal Medicine* d'Ettinger. L'apport de cette technologie est que les recherches ne se fondent pas uniquement sur des mots-clés mais aussi sur le contexte qui est fourni par le vétérinaire lors de sa recherche. Sofie peut renvoyer des informations aussi diverses que des diagnostics différentiels, des protocoles de traitement, des procédures d'urgence, ... Sofie ne donne pas une réponse fermée à une question posée mais permet

d'accéder à la bonne information en fonction du contexte. Selon ses concepteurs, Sofie permet au vétérinaire de gagner du temps, d'optimiser son flux d'activité, d'améliorer la prise en charge des patients et d'encourager la mise à niveau constante des connaissances des vétérinaires sur le long terme (LIFELEARN ANIMAL HEALTH, 2019).

Enfin, depuis quelques mois, nous voyons l'émergence d'outil que l'on appelle LLM (*Large Language Models*). Le plus populaire étant probablement ChatGPT, développé par la société OpenAI. Un LLM est un modèle statistique qui a intégré une quantité très importante de données sous forme de texte. Pour faire simple, le modèle apprend quelle suite de mots doit « répondre » à une question posée par un utilisateur. Cependant, ce n'est pas un simple copié-collé, il s'agit d'une production originale. ChatGPT n'a pas été conçu pour accompagner la pratique des médecins et des chercheurs selon l'INSERM (INSERM, 2023). Cependant, tout professionnel de santé comme les vétérinaires mais aussi les patients peuvent recourir à ces modèles pour se renseigner sur des affections, réaliser des diagnostics différentiels, vérifier des posologies, ... Selon Marc Cavazza, diplômé de la Faculté de Médecine Xavier Bichat et Docteur en Biomathématiques de l'Université Paris Cité, ce type de modèle n'est pas capable d'avoir des raisonnements hypothético-déductifs, pourtant essentiels au diagnostic différentiel (Cavazza, 2023). Cependant, d'autres études, comme celle menée par Arya Rao et son équipe, ont montré que ChatGPT démontrait de bonnes performances en termes de diagnostic, de diagnostic différentiel et de connaissance générale en médecine (Rao *et al.*, 2023). S'il paraît difficile aujourd'hui d'envisager une utilisation routinière et fiable des LLM pour les vétérinaires, il est indéniable que ceux-ci peuvent trouver des applications concrètes et quotidiennes dans des problèmes peu complexes, qui demandent d'aller trouver une information précise et parfois difficilement accessible. Par ailleurs, ces algorithmes peuvent être utilisés dans le cadre de ce pour quoi ils ont été développés, l'intelligence conversationnelle. Ce type d'outil pourrait permettre de mettre en place des chatbots par exemple, accessibles via le site des cliniques. En fonction de la pertinence de leur réponse, ils pourraient permettre de filtrer un certain nombre d'appels de propriétaires et ainsi dégager du temps pour les vétérinaires ou les auxiliaires vétérinaires.

3. Quelles applications semblent être les plus prometteuses et les plus applicables en médecine vétérinaire ?

Même si les outils que nous avons listés précédemment en médecine vétérinaire ne constituent pas une liste exhaustive, il existe assez peu d'outils disponibles pour les vétérinaires. Au contraire, en médecine humaine, de très nombreuses entreprises se sont lancées dans le développement d'outils. Selon l'entreprise AI for Health, qui crée une base de données contenant les start-ups françaises dédiées à l'IA, il en existe au moins 165 travaillant sur le sujet (AI FOR HEALTH, 2023). Ces start-ups sont présentes dans de nombreux domaines du secteur de la santé comme la médecine préventive, la logistique, le suivi des patients, le diagnostic, ... D'ailleurs, une étude menée par le cabinet de conseil Ernst et Youngs en partenariat avec le CHRU de Nancy a montré que près de 86% des directeurs d'hôpitaux et des Présidents de la Commission Médicale d'Etablissement considèrent que le principal défi qui se pose à l'IA aujourd'hui est l'identification et la priorisation des domaines où l'IA apporte le plus de bénéfices (Ernst & Youngs et CHRU Nancy, 2019). Il n'est donc plus temps de se demander dans quels domaines l'IA pourrait avoir une application mais plutôt de savoir quels domaines prioriser. Selon cette même étude, 67% des répondants affirment que des dispositifs d'IA sont d'ores et déjà en place dans l'hôpital. Ils le sont cependant en nombre limité. Pour 86% du panel, il n'y a qu'entre 1 et 5 dispositifs actuellement mis en œuvre. La recherche reste le domaine prioritaire, suivi de l'appui aux pratiques médicales et soignantes et le pilotage des activités. Au contraire, les processus administratifs et logistiques ou la gestion des ressources ne font pas encore l'objet de projets à base d'IA. Il semblerait que cette tendance soit cependant en train de s'inverser. Selon l'étude *Health et AI : Now et Next* publiée en novembre 2022 par le cabinet Capgemini Invent et l'entreprise AI for Health, près de 78% des

hôpitaux de leur panel ont des projets en cours sur l'organisation des soins. Soit davantage que dans les autres domaines comme le diagnostic ou la chirurgie (Capgemini Invent et AI for Health, 2022).

Dans une étude sur le potentiel de l'application de l'IA en santé, Davenport et son équipe distinguent plusieurs domaines d'application de l'IA (Davenport et Kalakota, 2019) :

- Applications liées à la prédiction, le diagnostic et le traitement : Dans ce domaine, de nombreuses études font état des performances de l'IA dans le diagnostic et la prédiction d'atteintes pathologiques. Cependant, en pratique, peu d'hôpitaux ou de médecins les utilisent. Cela est dû au fait que si les algorithmes sont souvent meilleurs que les professionnels, ils sont difficiles à intégrer dans un parcours de soin. Par ailleurs, des questions d'éthique se posent notamment en termes de responsabilité du praticien ainsi que les façons de repenser la relation entre professionnels de santé et patients.

En médecine vétérinaire, ces systèmes sont peut-être plus faciles à envisager. Dans un premier temps, l'organisation de l'offre de soins est très différente de la médecine humaine. La proportion de vétérinaires « généralistes » amenés à détecter et à prendre en charge un nombre plus large d'affections est bien supérieure à la proportion de médecins généralistes en médecine humaine. A titre d'exemple, seulement 330 vétérinaires sont reconnus comme étant spécialistes par le Conseil National de la Spécialisation Vétérinaire (soit 1,7% des vétérinaires inscrits à l'Ordre National des Vétérinaires) alors que près de 55% des médecins revendiquent une activité intégrant au moins une spécialité. Les vétérinaires peuvent évidemment être amenés à référer des cas à un spécialiste mais pour un certain nombre d'affections, le vétérinaire prendra en charge le patient durant tout le processus. On peut par exemple penser à des affections rénales chroniques, très courantes et bien prises en charge par les vétérinaires. De plus, les structures vétérinaires sont généralement moins grosses que les structures hospitalières humaines. Lorsque l'on compte plus de 3000 hôpitaux en France selon la DREES (Direction des Etudes, de l'Evaluation et des Statistiques), on comptait 11 structures vétérinaires hospitalières fin 2020 selon l'Ordre des Vétérinaires (Conseil National de l'Ordre des Vétérinaires, 2021). Ainsi, par rapport à la médecine humaine, les vétérinaires travaillent majoritairement dans des petites structures et suivent des patients de bout en bout ou presque. Cela signifie donc que la gestion des parcours de soins est plus légère qu'en médecine humaine et qu'il est donc plus facile d'y intégrer des algorithmes d'IA pour la prédiction et le diagnostic d'affections. D'ailleurs, PicoxIA, système de diagnostic radiographique, fait partie des plus répandus en médecine vétérinaire et témoigne de la faisabilité de diffuser ce type de solution sur le marché.

En ce qui concerne le traitement en revanche, l'intérêt de l'utilisation de l'IA est moins évident. En médecine humaine, Watson d'IBM a été remis en cause pour sa capacité à proposer des protocoles adéquats pour les patients atteints de cancer, considérés comme des maladies très complexes à traiter. En revanche, de nombreux observateurs pensent qu'ils auraient dû se concentrer sur des applications plus accessibles. En médecine vétérinaire, il y a beaucoup moins d'engorgement des soins au niveau des spécialistes qu'en médecine humaine. Ainsi, lorsqu'une pathologie dépasse le niveau de compétence du clinicien qui la diagnostique, le fait de référer est moins problématique qu'en médecine humaine en termes de délai de prise en charge. Il est probable que le spécialiste sera en mesure de prendre en charge la pathologie sans avoir besoin d'un algorithme pour l'aider à définir un traitement. On comprend donc que dans le cas de maladies chroniques (et non d'urgences où la question pourrait être plus pertinente), l'utilisation de l'IA pour aider les vétérinaires à mettre en place des traitements n'est sans doute pas une priorité. En revanche, la question de l'observance du traitement et du suivi peut s'avérer pertinente ;

- Applications liées à l'engagement patient et à l'adhérence au traitement : Il s'agit d'un domaine critique dans la gestion des maladies chroniques. Plus les patients sont appliqués dans le suivi de leur traitement, meilleurs sont les résultats (médicaux et financiers). De la même façon, pour certaines maladies chroniques, si le patient ne met pas en place lui-même les ajustements comportementaux nécessaires, le traitement thérapeutique aura moins ou peu d'effets. Pourtant, selon une étude menée par Davenport, les cliniciens considèrent que moins de 50% des patients sont impliqués dans leurs traitements sur de longues durées (Davenport *et al.*, 2018). On peut estimer que cette proportion peut

être appliquée à l'implication des propriétaires dans le traitement de leurs animaux. Pourtant, il y a une attention croissante sur les possibilités apportées par l'IA d'utiliser les données récoltées par des objets connectés (montres, smartphones, ...) afin de faire remonter des alertes ou de mettre en avant des recommandations ou des rappels afin d'aider les patients et les professionnels de santé à bien gérer le suivi long-terme des patients.

En médecine vétérinaire, l'intérêt de telles technologies paraît pertinent. Le premier niveau de difficulté quand on parle d'adhérence au traitement intervient au niveau de l'engagement des propriétaires d'animaux qui peuvent être réticents à appliquer un traitement régulier sur de longues périodes. Le second niveau correspond à la coopération de l'animal. Il peut être difficile de forcer un animal à prendre son traitement (faire avaler une pilule à un chat réticent par exemple) ou même à suivre le comportement de l'animal (calcul de la quantité d'eau bue par un chat présentant des problèmes rénaux).

Peu d'initiatives sont sur le marché à ce jour. En revanche, certains outils utilisent des technologies qui pourraient être intégrées à des solutions pensées pour faire le suivi des animaux. Le distributeur d'eau et de nourriture Catspad propose par exemple un suivi de l'alimentation des chats et permet de contrôler la prise alimentaire de l'animal. En cas de baisse de prise d'alimentation, le système couplé à une application remonte une alerte à l'utilisateur. Nous pourrions imaginer un système similaire avec calcul de la prise de boisson couplé à d'autres systèmes de monitoring. Cela nécessiterait par exemple un couplage avec le logiciel de gestion de la clinique ;

- Applications administratives : Ces applications sont particulièrement importantes en médecine humaine. En effet, le personnel infirmier en hôpital passe près de 25% à s'occuper des tâches administratives plutôt que du soin des patients (HEALTHLEADERS, 2010). Cela peut correspondre au management des données cliniques, la gestion de plaintes, de la documentation clinique, ... L'IA peut rendre des services dans ce domaine. On peut penser aux multiples chatbots qui servent de premiers intermédiaires entre patients et structures de soins.

En médecine vétérinaire, ces applications semblent plus anecdotiques. Comme nous l'avons vu précédemment, les structures vétérinaires sont en moyenne plus petites que les structures en santé humaines. Ce qui entraîne une moindre lourdeur administrative. L'utilisation des chatbots est encore peu développée du fait que peu de cliniques effectuent de rendez-vous en ligne (Vétolib est beaucoup moins développé que Doctolib par exemple). Cependant, avec le développement de ce type d'outils (prise de RDV en ligne, boutiques en ligne, ...), nous pouvons envisager que les applications se développent à l'avenir pour aider administrativement les vétérinaires.

En conclusion, les applications d'IA en médecine vétérinaire sont aujourd'hui peu nombreuses. Si l'efficacité de l'IA en médecine n'est plus à démontrer, son applicabilité pose davantage de questions dans le domaine de la santé animale. En médecine humaine, de nombreuses initiatives sont mis en place par des instituts de recherche mais peinent à passer à l'échelle réelle. En médecine vétérinaire, même au niveau des équipes de recherche, le sujet est nettement moins développé. Cependant, il existe des applications pour lesquelles l'utilisation de technologies à base d'IA pourrait rendre des services dans le cadre de la gestion des maladies chroniques pour les animaux de compagnie. En diagnostic tout d'abord, PicoxIA démontre que des solutions pertinentes peuvent être utilisées. L'IA pourrait permettre de diagnostiquer plus précocement et avec plus de précision certaines affections dont les marqueurs seraient peu visibles (maladies rénales chroniques débutantes par exemple). Dans le suivi des pathologies, l'IA pourrait permettre de construire une nouvelle relation avec les propriétaires. Le vétérinaire serait prévenu plus vite des difficultés d'observance du traitement ou des changements de comportement. Cela pourrait permettre de planifier des rendez-vous plus intelligemment, lorsque la nécessité est réelle. Ces initiatives sont aujourd'hui peu nombreuses. Nous nous proposons d'explorer les raisons et les initiatives qui pourraient être menées afin de favoriser la diffusion de cette technologie au sein de la pratique vétérinaire.

4. Créer des modèles applicables pour les vétérinaires et diffusion au sein de la profession

A. Développer des modèles d'IA : une histoire d'équipe et de données

Nous avons vu précédemment que les modèles d'IA ont besoin de grandes quantités de données afin d'être développés. Sans rentrer dans trop de détails, pour construire et s'assurer de l'efficacité d'un modèle, nous utilisons trois jeux de données différents :

- Un jeu de données d'apprentissage sur lequel le modèle va construire ses paramètres afin d'accomplir la tâche qu'on lui demande au fil des itérations ;
- Un jeu de données de validation utilisé également pendant la phase d'apprentissage afin d'ajuster les paramètres construits par le modèle. Ce sont des données sur lesquelles le modèle n'apprend pas mais qui permettent de s'assurer qu'il ne fait pas de surentraînement (le modèle construit des paramètres non reliés à la tâche qu'on lui demande de faire) ;
- Un jeu de données de test qui ne sert pas à ajuster le modèle mais de tester ses performances comme s'il était mis en situation d'utilisation réelle. Ce jeu doit être composé exclusivement de données nouvelles pour le modèle afin de ne pas biaiser le résultat avec des échantillons qu'il aurait déjà vus.

Pour construire des algorithmes, la quantité et la fiabilité des données est donc primordiale. Se procurer les données nécessaires est un enjeu de taille pour les concepteurs d'algorithmes. Existe-t-il des bases de données de qualité suffisantes et comment peuvent faire les entrepreneurs pour y avoir accès ?

Si l'on prend en exemple PicoxIA, les concepteurs annoncent dans leurs publications avoir eu accès à des radiographies provenant d'un centre de référent. Ces radiographies ont été annotées par des spécialistes diplômés. Ces bases de données ne sont pas accessibles librement. Il est difficile d'anticiper les moyens d'accès à ces bases de données ainsi que d'évaluer leur pertinence.

De façon générale, nous pouvons considérer que les données diagnostiques sur la base d'examen sont fiables dans les centres de référent. En effet, les spécialistes, sur la base d'examen sont entraînés à poser les bons diagnostics. En revanche, nous pouvons imaginer que d'autres données comme l'anamnèse ou l'historique médical peuvent être erronées (problèmes de communication, oubli de la part du propriétaire ou du clinicien, ...). Lorsqu'une équipe veut entraîner un modèle sur la base de données issues de centres de référent, elle doit donc se poser la question de la qualité de toutes les données qu'elle récolte et utilise.

Par ailleurs, on peut envisager que les centres de référent ne sont pas enclins à fournir leurs données (pour des questions de confiance par exemple). Il faut donc se tourner vers d'autres sources de données.

Les bases de données vétérinaires sont peu recensées sur le web, ce qui constitue un premier obstacle à la construction d'un algorithme basé sur des technologies d'IA. Nous pouvons tout de même citer le Veterinarian Companion Animal Surveillance System (VetCompass™). C'est une initiative qui part du principe que de nombreuses informations sur les patients sont stockées de façon délocalisée dans les systèmes de management des cliniques individuelles. L'objectif de cette initiative lancée par le Royal Veterinary College à Londres est de collecter ces données en les anonymisant et en les regroupant dans une même base de données. Cette initiative lancée à la fin des années 2000 s'est étendue et compte plus de 20 millions d'animaux provenant de plus de 1800 cliniques anglaises (Royal Veterinary College, 2023). Cette initiative s'est propagée

à la pratique équine en 2016 et à l'Australie en 2013. L'accès à cette base de données permet à des chercheurs de créer des études afin de faire avancer la recherche en médecine vétérinaire. Bien que cette base de données ne soit pas non plus d'accès public, elle a le mérite d'avoir été réalisée à une échelle bien supérieure à celle d'une seule clinique, même de référé. Elle permet donc de fournir des données sur des cas plus nombreux et plus diversifiés. Concrètement, les vétérinaires n'ont rien à faire de particulier selon les auteurs de l'initiative. VetCompass™ va chercher directement les informations dans les bases de données des cliniques. Les informations recherchées sont les diagnostics, les notes cliniques, les traitements mis en place ainsi que des informations relatives aux patients (âge, sexe, identification, ...). Ces données sont évidemment anonymisées et protégées selon la réglementation en vigueur.

Il est difficile de dresser une liste exhaustive de toutes les bases de données existantes. On peut cependant également mentionner la Golden Retriever Lifetime Study financée par la Morris Animal Foundation. Il s'agit également d'une base de données destinée à des fins de recherche (Morris Animal MORRIS ANIMAL FOUNDATION, 2023). Initialement, les données étaient principalement à destination de recherche contre différents cancers (lymphomes, hémangiosarcomes, ostéosarcomes, ...). Selon la fondation, des acteurs de l'industrie, des instituts de recherche et des organisations à but lucratif peuvent candidater dans le cadre de partenariats de recherche. Le processus de sélection est à la discrétion de la fondation. Du fait que près de 500 études ont été réalisées et publiées grâce à l'utilisation de cette base de données, on peut supposer que celle-ci est de qualité suffisante mais il est difficile de se prononcer sur sa capacité à entraîner des modèles d'IA.

Ainsi, on peut noter qu'il semble exister des bases de données suffisamment fiables pour être utilisées à des fins de recherche ou d'entraînement d'algorithmes (cf. PicoxIA). Cependant, celles-ci sont peu accessibles, leur fiabilité est difficile à évaluer ainsi que leur taille. S'il est donc possible de créer des solutions à l'aide de partenariats par exemple, il apparaît que ce manque d'accessibilité est un obstacle certain à la diffusion à large échelle de technologies à base d'IA au sein de la population vétérinaire. Il faut aussi se rappeler que plus la qualité de la base de données est faible, plus la quantité de données nécessaires à l'entraînement d'un algorithme est importante. Ainsi, vu qu'il n'existe pas de bases de données à grande échelle en France facilement accessible, la qualité des données est primordiale alors même qu'elle paraît difficile à évaluer.

La solution à ce problème pourrait venir de grands groupes à l'image de compagnies d'assurance dans les pays où celles-ci sont très implantées, des chaînes de cliniques ou de groupes pharmaceutiques possédant des chaînes de clinique. En effet, nous pouvons imaginer que ces chaînes disposent de données issues de toutes les cliniques qu'elles possèdent. Elles peuvent donc créer des bases de données vastes, sur plusieurs années, avec des animaux provenant d'endroits différents. De plus, elles peuvent mettre en place des procédures au sein des cliniques afin d'harmoniser dès la collecte le type de données à récolter. On comprend donc que ces grandes structures, si elles ne possèdent pas encore des bases de données prêtes à permettre l'entraînement d'algorithmes, elles pourraient les mettre en place dès à présent. Par ailleurs, il est probable que ces structures ne soient pas enclines à permettre l'exploitation de leur base. Ce qui peut poser à terme des problématiques de non-concurrence ou de concentration du marché.

Si les données sont essentielles pour développer des modèles à base d'IA, disposer d'une équipe pluridisciplinaire apparaît aussi primordial. En effet, à quelques exceptions près, les vétérinaires ont peu de connaissances en science des données. En effet, cela ne fait pas partie de leur cursus initial mais en plus, ces connaissances sont peu utiles dans une pratique courante. De même, il n'existe pas (ou peu) de passerelles permettant de disposer d'une formation en science des données et en médecine vétérinaire. Si un vétérinaire décide de se former en science des données, il doit le faire par ses propres moyens via des plateformes d'apprentissage en ligne (Coursera par exemple), soit par des formations diplômantes comme des M2 science des données. Ces profils restent rares aujourd'hui, ce qui explique la nécessité de construire des équipes avec des profils aux parcours différents. Une nouvelle fois, des grosses structures sont peut-être mieux positionnées pour mettre sur pied ces équipes pluridisciplinaires. En effet, elles ont les capacités

financières pour absorber des équipes de science des données et elles possèdent l'attractivité pour les recruter. Ces équipes sont probablement déjà existantes ne serait-ce que pour maintenir à jour les bases de données centralisées et les systèmes informatiques (des structures ou de l'ensemble de l'organisation). Il ne s'agit donc pas de créer *de novo* une nouvelles équipe mais plutôt de l'agrandir.

B. Sensibiliser les vétérinaires à l'utilisation et au potentiel de l'IA

Comme nous l'avons vu dans les parties précédentes, l'IA peut avoir de multiples applications : prédiction, prévention, diagnostic, traitement, observance, ... Cependant, ces applications ne sont utiles que si les vétérinaires les connaissent et les comprennent suffisamment bien pour pouvoir les utiliser et les expliquer aux propriétaires d'animaux si besoin. Qu'en est-il du niveau de connaissances des vétérinaires sur le sujet en 2023 ?

Il y a deux étapes dans la formation des vétérinaires : la formation initiale dispensée pendant les années d'études puis la formation continue tout au long de la carrière des vétérinaires.

La formation continue est prise en charge par les ENV (Ecoles Nationales Vétérinaires). A l'EnvA, la notion d'IA est introduite auprès des étudiants dès la deuxième année dans le cadre plus large de la e-santé vétérinaire. Les étudiants ont un cours sur le sujet ainsi qu'une conférence. Ils ont par ailleurs accès à des livres blancs ou des supports de présentation sur la e-santé (incluant souvent une partie sur l'IA). Les étudiants voient notamment des exemples d'applications qui existent aujourd'hui. Cependant, le digital étant un secteur en évolution rapide, ce que les étudiants auront vu au début de leur cursus ne sera plus nécessairement valable à leur sortie d'école. Une étude réalisée par VetoNetwork en 2022 intitulée « Jeunes vétérinaires et Numérique » sur les étudiants des ENV françaises montre que plus de 85% des étudiants n'avaient aucune formation sur l'Intelligence Artificielle durant leur formation. Ainsi, la formation des étudiants sur le sujet évolue vers davantage de sensibilité au sujet même si aujourd'hui encore, ces notions ne sont que brièvement présentées en début de scolarité.

Concernant la formation continue une fois sortie d'école, les vétérinaires doivent se former eux-mêmes sur le sujet s'ils veulent se l'approprier. Un certain nombre de supports et de moyens de s'informer existent mais le principe fonctionne sur la base du volontariat. Les vétérinaires qui souhaitent se former sur le sujet peuvent le faire mais ceux qui ne le souhaitent pas n'aborderont pas le sujet, au risque de laisser passer des opportunités.

Parmi les moyens d'information, on peut citer les conférences de Vet-in-Tech, qui sont accessibles en rediffusion, des salons tel que Numanima dans lesquels il y a des interventions concernant l'IA. Enfin, on peut citer les blogs ou les sites internet personnels à l'image de Vetiverse ou E-santé animale qui sont en libre accès. La presse spécialisée produit également des articles sur le sujet comme le Point Vétérinaire. De manière générale, ces outils parlent de l'avancée du digital dans son ensemble mais on peut retrouver du contenu sur l'IA. Il faut noter l'importance des Key Opinion Leaders (KOLs), c'est-à-dire des leaders d'opinion, dans la diffusion de la technologie au sein de la profession. Les solutions d'IA n'ont vocation ni à remplacer le vétérinaire généraliste ni à surpasser les compétences des vétérinaires spécialistes. Il faut donc éduquer les KOLs, notamment les spécialistes dont le message est largement relayé, sur l'utilisation réelle de l'IA. Ainsi, ils pourraient délivrer un message éclairé sur le sujet largement diffusé au sein de la profession.

L'acculturation des vétérinaires peut aussi se faire au travers du positionnement d'instances publiques ou politiques du milieu. Nous pouvons notamment penser à des institutions comme l'Ordre des Vétérinaires ou le SNVEL (Syndicat National des Vétérinaires d'Exercice Libéral). En ce qui concerne le SNVEL, le syndicat prend peu partie quant à l'IA. En effet, il n'y a aucune publication en leur nom sur le sujet sur leur site Internet. Même s'il ne s'agit pas d'une de leurs missions *stricto sensu*, le SNVEL se veut tout de même partenaire des vétérinaires en ce qui concerne leur avenir comme en témoigne l'intitulé d'une de leurs missions « Ecouter

vos besoins, préparer votre avenir » (SYNDICAT NATIONAL DES VETERINAIRES D'EXERCICE LIBERAL, 2023). On pourrait donc s'attendre à des publications sur un sujet émergent (et connu depuis longtemps en pratique rurale). Il pourrait s'agir également d'un moyen pour faire connaître le sujet aux praticiens. En ce qui concerne l'Ordre, il y a quelques communications sur le digital en général mais assez peu sur l'IA, témoin qu'il ne s'agit pas d'un sujet prioritaire.

Enfin, en plus de l'acculturation des vétérinaires, l'acculturation de la société et l'acceptabilité de l'IA dans la prise en charge des soins peuvent aussi jouer un rôle important. Selon une étude menée par Ipsos et le LEEM en 2018, plus de 50% de la population accepterait que des modèles d'IA puissent donner des recommandations relatives à leur santé. La proportion tombe à moins de 30% lorsqu'il s'agit de donner un diagnostic ou de faire une prescription (STATISTA, 2018). On comprend donc que la population civile reste circonspecte quant à une implication croissante de l'IA dans la prise en charge de leur santé. En revanche, selon cette étude, tant que des professionnels de santé restent ultimes décisionnaires, la société civile encourage l'introduction raisonnée de l'IA dans la pratique de la médecine. On peut supposer que cette tendance devrait se confirmer en ce qui concerne la santé des animaux de compagnie. Des contrôles et des décisions éthiques concernant l'usage des données, sujet dont le secteur de la santé animale s'est peu emparé, permettraient d'encourager l'acceptation de la population vis-à-vis de ces nouvelles technologies.

C. Proposer des solutions à valeur ajoutée pour les vétérinaires

Pour faciliter la diffusion des modèles d'IA, il faut que les vétérinaires aient des besoins auxquels ces modèles puissent répondre. Nous en revenons à une idée précédente de l'importance d'une équipe pluridisciplinaire. Pour développer une solution qui se diffuserait au sein de la population vétérinaire, il faut comprendre quels besoins cette solution pourrait satisfaire. Pour ce faire, posséder des connaissances solides en médecine vétérinaire permettrait de sonder la profession et d'identifier les besoins. Identifier ces besoins n'est pas nécessairement l'obstacle le plus difficile à franchir. En effet, il existe un certain nombre de congrès en France et dans le monde où rencontrer des vétérinaires (si l'on ne vient pas de ce milieu) et donc de discuter des difficultés qu'ils rencontrent. De plus, si l'on repense à la centralisation croissante de l'activité clinique, il sera possible pour une entité comme une chaîne de cliniques d'identifier les difficultés rencontrées par les praticiens (il s'agit même parfois d'un argument mis en avant par certain groupe lors du rachat d'une clinique).

En revanche, l'implémentation de ces nouvelles solutions peut s'avérer être une difficulté. Sachant que l'introduction de nouvelles technologies telles que l'IA promet de changer la façon dont les cliniques fonctionneront, il faut tenir compte du coût de changement de la pratique. Pour donner un exemple, le système de prise de rendez-vous en ligne adopté par certaines cliniques réduit le temps passé par les vétérinaires et leurs assistants à planifier les rendez-vous. Il faut donc trouver une nouvelle organisation du temps à la place du temps qui était autrefois dévolu à la prise de rendez-vous. Cette notion de coût du changement est bien connue en marketing et peut s'avérer réhibitoire lorsqu'elle concerne des éléments importants du fonctionnement d'une clinique, particulièrement dans un contexte de sur-activité où le temps pour réfléchir à la réorganisation du travail manque. Cette situation concerne de nombreuses cliniques.

Pour pallier ce problème, il est important de réduire autant que possible ce coût du changement afin d'encourager les cliniques à adopter ces technologies. L'interopérabilité des logiciels avec les outils déjà disponibles dans la clinique est primordiale. En effet, même si le marché des logiciels de gestion de clinique est assez resserré, il en existe un certain nombre. Il faut que les nouveaux puissent être pris en charge facilement par l'environnement de la clinique. Nous avons par exemple vu que PicoxIA était pris en charge d'ores et déjà par deux systèmes, AssistoVet et VetoPartner. Il faut élargir ce champ. Une fois encore, les

groupements de cliniques sont bien placés puisqu'ils peuvent uniformiser les logiciels au sein des cliniques qu'ils possèdent. On peut envisager qu'il soit possible de développer des outils qui s'interfaçent avec leur logiciel de gestion dès la conception. On peut envisager d'autres obstacles à la mise en place de nouveaux outils comme la facturation. En effet, ces outils ont un coût pour les cliniques et s'ils permettent de mieux prendre en charge les animaux, ils doivent rester rentables pour la clinique. Ainsi, les concepteurs de solutions doivent accompagner autant que possible les cliniques à l'installation de nouveaux outils pour maximiser la valeur ajoutée offerte aux vétérinaires dans leur pratique mais aussi en trouvant un business model viable pour la clinique et ainsi faciliter la diffusion de ces solutions au sein de la population vétérinaire.

Enfin, la performance des algorithmes ne doit pas être dégradée par rapport aux méthodes classiques des vétérinaires. Il faut notamment comparer les performances des algorithmes avec les méthodes de référence si elles sont à disposition des vétérinaires en pratique courante ou avec les méthodes couramment employées par les vétérinaires. Gardons à l'esprit que les études publiées sur tel ou tel sujet se font souvent dans des contextes restreints à ceux de l'étude. L'ensemble des publications que nous avons pu détailler préalablement dans ce travail constituent des Proofs of Concept (POCs) du potentiel des méthodes d'IA dans le domaine. Selon Pascal Gêné (alors responsable des ventes chez Watson Health), lors d'une présentation lors du congrès E-vet 2020 ayant eu lieu le 30/01/2020 à Paris, lorsqu'une étude est publiée, elle utilise un jeu de données particulier dans un environnement particulier. Si on essaye de l'appliquer à un autre jeu de données, dans un autre environnement avec des outils différents, les résultats ne devraient pas être concluants. Cela ne signifie pas que le modèle n'est pas utilisable mais plutôt qu'il faut continuer son entraînement sur des données plus représentatives des méthodes et de la nouvelle population cible. On comprend donc la complexité du procédé si l'on veut faire passer à l'échelle (au moins nationale) des modèles au vu de la diversité des pratiques, des populations cibles, ...

En conclusion de cette dernière partie, nous voyons qu'il y a un certain nombre d'obstacles à la diffusion de l'IA (notamment en France) dans le milieu vétérinaire. Il y a tout d'abord le manque de disponibilité des données qui sont soit des données à destination de la recherche soit des données qui appartiennent à des organisations privées. Il y a ensuite le sujet de la création des algorithmes qui nécessitent les connaissances d'équipes pluridisciplinaires dans un milieu où les professionnels de santé sont peu acculturés à l'utilisation de l'IA. Enfin, il y a le sujet de la démonstration de la valeur ajoutée des solutions proposées. En quoi est-ce que cela va aider les vétérinaires dans leur exercice de la médecine vétérinaire ? Comment se comporte l'outil par rapport au gold standard pour un sujet donné ? Comment aider les vétérinaires à changer leur pratique pour s'adapter aux solutions avec un minimum d'effort ? Autant de questions que doivent se poser les concepteurs de solutions afin qu'elles soient adoptées par les vétérinaires. Les groupements de cliniques détiennent certaines réponses à toutes ces questions.

Prenons l'exemple de Mars Veterinary Health pour illustrer ce dernier propos. Ce groupe possède un réseau de plus de 2500 cliniques localisées dans plus de 20 pays à travers le monde. Nous comprenons donc qu'ils sont en mesure de construire et de maintenir des bases de données sur les animaux traités dans ces cliniques. Ils disposent donc de la capacité d'exploiter ces bases de données grâce à l'expertise d'ingénieurs en science de la donnée et de vétérinaires de leur propre groupement. De plus, ils sont en mesure de créer des solutions qui s'interfaçeront avec les logiciels déjà en place au sein des cliniques. Ainsi, ils seront en mesure de propager des solutions et d'accompagner les vétérinaires à les mettre en place, apportant ainsi de la valeur pour les praticiens mais aussi pour les propriétaires d'animaux *in fine*. Nous avons vu précédemment qu'ils ont développé RenalTech™, qui permet de prédire l'apparition d'une MRC dans les deux ans malgré des analyses rénales dans les normes grâce à l'IA. Ils disposent aussi de Wisdom Panel, une entreprise spécialisée dans le génotypage pour animaux de compagnie. Leur outil permet de détecter (pour moins de \$130) les prédispositions génétiques des chiens et des chats. En mettant en place ces outils dans leurs cliniques, ils peuvent « prédire » les affections que risquent de développer les animaux. Enfin, le groupe Mars possède Royal Canin® et leur collier intelligent Whistle Fit qui permet de suivre le comportement du chien et

donc d'alerter en cas de problèmes. On comprend donc que le groupe Mars peut révolutionner la pratique courante en permettant la prédiction des maladies et un suivi amélioré des animaux en appliquant leurs solutions dans leurs cliniques. Cette diffusion facile au sein de leurs structures leur conférerait un avantage compétitif certain vis-à-vis d'autres cliniques ne les ayant pas adoptées.

CONCLUSION

Les technologies numériques en général et les technologies d'Intelligence Artificielle en particulier reposent sur l'exploitation des données. Ces données se doivent d'être en quantité mais aussi en qualité suffisantes afin de bâtir les modèles les plus fiables possibles. La quantité de données n'est pas un problème. Avec l'émergence des technologies Big Data, nous produisons toujours plus de données chaque année. En revanche, leur qualité dépend du choix de collecte et de sa rigueur. Le stockage demande aussi des capacités de plus en plus importantes.

Cependant, le secteur de la santé a intégré le potentiel de l'IA dans l'amélioration de la prise en charge des patients, via l'amélioration du système de santé en général ainsi que via la prise en charge individuelle et personnalisée (médecine des « 4P » ou des « 6P »). En ce sens, de nombreuses publications démontrent l'efficacité de ces technologies dans tous les aspects de la médecine. On en retrouve des exemples en chirurgie, en diagnostic, en éducation des populations, en prévention, ... En revanche, le passage de POC à l'échelle rencontre des difficultés. En premier lieu, se posent des questions d'éthique et de réglementation. Les instances politiques à l'échelle nationale et internationale tentent de mettre en place un cadre légal qui permettrait d'encadrer les pratiques pour la protection des patients (et particulièrement de leurs données individuelles). Ensuite, se posent des questions liées à l'interopérabilité des modèles avec l'environnement de santé existant. De nombreux acteurs produisent et utilisent les données, de nombreux acteurs produisent également des solutions mais la difficulté de lier tous ces modèles demeure. Les possibilités de l'IA en santé restent néanmoins certaines et les professionnels de santé en prennent conscience.

En médecine vétérinaire, le problème est plus complexe. Peu d'initiatives sont disponibles sur le marché et tout particulièrement en France. De plus, les solutions disponibles sont encore peu démocratisées. Les concepteurs se heurtent d'abord à des problèmes d'acculturation des vétérinaires. Nous voyons souvent des vétérinaires, dont des spécialistes, mal appréhender les potentialités de l'IA. Les technologies à base d'IA ne remplaceront pas les vétérinaires, ils ont encore moins vocation à remplacer les spécialistes dont l'expertise est essentielle. En revanche, ils peuvent aider les vétérinaires à être plus efficaces et à renforcer leurs relations avec les propriétaires. Ces systèmes peuvent notamment permettre des gains de temps qui permettraient aux vétérinaires de mieux prendre en charge les animaux. Se pose néanmoins le sujet de la mise en place de ces systèmes au sein des cliniques. Pour espérer une diffusion de la profession, il faut accompagner les vétérinaires dans la mise en place des systèmes. Les groupements de cliniques semblent bien placés pour encourager la diffusion de l'IA au sein de la profession. Ils disposent de la main d'œuvre pour développer les algorithmes mais aussi de l'environnement informatique pour les mettre en place et accompagner les vétérinaires dans leur adoption. A terme, cela pourrait leur donner un avantage compétitif certain vis-à-vis de leurs concurrents, moins bien équipés. En effet, via des outils de prédiction et de diagnostic, les vétérinaires pourraient suivre rigoureusement des animaux susceptibles de développer des pathologies à l'avenir permettant une prise en charge plus rapide et plus adaptée. Cette meilleure prise en charge se poursuivrait par une surveillance plus rapprochée de l'animal dans la gestion de sa maladie. Il ne faut pas non plus sous-estimer le gain de temps administratif potentiellement apporté par l'IA. Les vétérinaires pourraient utiliser leur temps autrement, pour être davantage au contact des patients et des propriétaires.

Les acteurs publics comme l'Ordre, les ENV ou le SVEL ont aussi l'opportunité de favoriser le développement de l'IA via la formation qu'ils peuvent proposer aux futurs et actuels vétérinaires. Leur implication est d'autant plus importante que le milieu vétérinaire français est encore peu exposé à ces technologies. Pourtant, dans d'autres pays comme aux Etats-Unis, l'IA a montré pouvoir améliorer la prise en charge des animaux de compagnie, ce qui est finalement la raison d'être des vétérinaires.

BIBLIOGRAPHIE

- AGENCE DU NUMÉRIQUE EN SANTÉ (2022) Référencement des services numériques au catalogue d'applications de Mon espace santé : une nouvelle étape est franchie ! *In esante.gouv.* [https://esante.gouv.fr/espace-presse/referencement-des-services-numeriques-au-catalogue-dapplications-de-mon-espace-sante-une-nouvelle-etape-est-franchie] (consulté le 17/09/2022).
- AGENCE NATIONALE DU NUMERIQUE EN SANTE (2022) Mon espace santé présente son catalogue de service. *In Esanté.gouv.* [https://esante.gouv.fr/actualites/mon-espace-sante-presente-son-catalogue-de-service] (consulté le 17/09/2022).
- AI FOR HEALTH (2023) AI for Health European Mapping. *In Aiforhealth.* [https://www.aiforhealth.fr/] (consulté le 14/04/2023).
- ATHER, S., KADIR, T., GLEESON, F. (2020) Artificial intelligence and radiomics in pulmonary nodule management: current status and future applications. *Clinical Radiology* vol. 75, n° 1, p. 13-19. [https://doi.org/10.1016/j.crad.2019.04.017]
- BÆRØE, K., MIYATA-STURM, A., HENDEN, E. (2020) How to achieve trustworthy artificial intelligence for health. *Bulletin of the World Health Organization* vol. 98, n° 4, p. 257-262. [https://doi.org/10.2471/BLT.19.237289]
- BERKELEY LAW (2020) Student Summer Series: Chris Gronseth '22 Finds Real Meaning in Artificial Intelligence - Berkeley Law. *In berkeley.edu.* [https://www.law.berkeley.edu/article/student-summer-series-chris-gronseth-22-finds-real-meaning-in-artificial-intelligence/] (consulté le 22/05/2023).
- BERNELL, S., HOWARD, S.W. (2016) Use Your Words Carefully: What Is a Chronic Disease? *Frontiers in Public Health* vol. 4, p. 159. [https://doi.org/10.3389/fpubh.2016.00159]
- BIOURGE, V., DELMOTTE, S., FEUGIER, A., *et al.* (2020) An artificial neural network-based model to predict chronic kidney disease in aged cats. *Journal of Veterinary Internal Medicine* vol. 34, n° 5, p. 1920-1931. [https://doi.org/10.1111/jvim.15892]
- BISCHOFF, S.C., BOIRIE, Y., CEDERHOLM, T., *et al.* (2017) Towards a multidisciplinary approach to understand and manage obesity and related diseases. *Clinical Nutrition* vol. 36, n° 4, p. 917-938. [https://doi.org/10.1016/j.clnu.2016.11.007]
- BOHANNAN, Z., PUDUPAKAM, R.S., KOO, J., *et al.* (2021) Predicting likelihood of in vivo chemotherapy response in canine lymphoma using ex vivo drug sensitivity and immunophenotyping data in a machine learning model. *Veterinary and Comparative Oncology* vol. 19, n° 1, p. 160-171. [https://doi.org/10.1111/vco.12656]
- BOISSADY, E., DE LA COMBLE, A., ZHU, X., *et al.* (2021) Comparison of a Deep Learning Algorithm vs. Humans for Vertebral Heart Scale Measurements in Cats and Dogs Shows a High Degree of Agreement Among Readers. *Frontiers in Veterinary Science* vol. 8, p. 764570. [https://doi.org/10.3389/fvets.2021.764570]
- BRADLEY, R., TAGKOPOULOS, I., KIM, M., *et al.* (2019) Predicting early risk of chronic kidney disease in cats using routine clinical laboratory tests and machine learning. *Journal of Veterinary Internal Medicine* vol. 33, n° 6, p. 2644-2656. [https://doi.org/10.1111/jvim.15623]
- CAPGEMINI INVENT, AI FOR HEALTH (2022) HEALTH & AI : Now and Next. [https://www.capgemini.com/fr-fr/perspectives/publications/health-ai-now-and-next/] (consulté le 17/04/2023).
- CAVAZZA, M. (2023) ChatGPT : quand l'IA répond aux QCM de Médecine. *In esanum.fr.* [https://www.esanum.fr/today/posts/quand-lia-repond-aux-qcm-de-medecine] (consulté le 14/04/2023).
- CHU LYON (2022) Le pancréas artificiel expérimenté sur des patients diabétiques. *In CHU Lyon.* [https://www.chu-lyon.fr/le-pancreas-artificiel-experimente-sur-des-patients-diabetiques] (consulté le 14/11/2022).

- CNIL (2022) IA : comment être en conformité avec le RGPD ? In *CNIL*. [<https://www.cnil.fr/fr/intelligence-artificielle/ia-comment-etre-en-conformite-avec-le-rgpd>] (consulté le 20/07/2022).
- COMMISSION EUROPEENNE (2021) Proposal for a Regulation of the European Parliament and of the council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts. In *EUR-Lex*. [<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>] (consulté le 21/07/2022).
- COTE, C. (2021) 3 Applications of Data Analytics in Healthcare. In *Harvard Business School Online - Business Insights Blog*. [<https://online.hbs.edu/blog/post/data-analytics-in-healthcare>] (consulté le 20/07/2022).
- COULET, A., JAY, N. (2019) Contre les effets indésirables, un algorithme pour personnaliser les doses de médicaments. In *The Conversation*. [<http://theconversation.com/contre-les-effets-indesirables-un-algorithme-pour-personnaliser-les-doses-de-medicaments-112438>] (consulté le 20/07/2022).
- CURRAN, K., THAMM, D.H. (2016) Retrospective analysis for treatment of naïve canine multicentric lymphoma with a 15-week, maintenance-free CHOP protocol. *Veterinary and Comparative Oncology* vol. 14 Suppl 1, p. 147-155. [<https://doi.org/10.1111/vco.12163>]
- DAVENPORT, T., KALAKOTA, R. (2019) The potential for artificial intelligence in healthcare. *Future Healthcare Journal* vol. 6, n° 2, p. 94-98. [<https://doi.org/10.7861/futurehosp.6-2-94>]
- DAVENPORT, T.H., HONGSERMEIER, T.M., CORD, K.A.M. (2018) Using AI to Improve Electronic Health Records. In *Harvard Business Review*. [<https://hbr.org/2018/12/using-ai-to-improve-electronic-health-records>] (consulté le 19/04/2023).
- DEFORTESCU, G., TILLOU, X., BADET, L., *et al.* (2016) La chirurgie robot-assistée en transplantation rénale. *Progrès en Urologie - FMC* vol. 26, n° 4, p. F67-F72. [<https://doi.org/10.1016/j.fpurol.2016.06.001>]
- DE ROOIJ, M., VAN POPPEL, H., BARENTSZ, J.O. (2021) Risk Stratification and Artificial Intelligence in Early Magnetic Resonance Imaging-based Detection of Prostate Cancer. *European Urology Focus* p. S2405-4569(21)00305-9. [<https://doi.org/10.1016/j.euf.2021.11.005>]
- DIAS, R., TORKAMANI, A. (2019) Artificial intelligence in clinical and genomic diagnostics. *Genome Medicine* vol. 11, n° 1, p. 70. [<https://doi.org/10.1186/s13073-019-0689-8>]
- DRIOT, C. (2023) L'intelligence artificielle au service des vétérinaires. *La Semaine Vétérinaire* n° 1979, p. 36-41.
- ERNST & YONGS, CHRU NANCY (2019) Baromètre de maturité de l'IA dans les CHU. [<https://www.chu-nancy.fr/index.php/chru-et-ey-1er-barometre-de-maturite-de-l-ia-dans-les-chu>] (consulté le 14/04/2023).
- ERTEL, W. (2017) Introduction to Artificial Intelligence. In Ertel, W. (Éd.), *Introduction to Artificial Intelligence, Undergraduate Topics in Computer Science*. Cham, Springer International Publishing, p. 1-21. [https://doi.org/10.1007/978-3-319-58487-4_1]
- FALLERY, B. (2020) Données de santé : l'arbre StopCovid qui cache la forêt Health Data Hub. In *The Conversation*. [<http://theconversation.com/donnees-de-sante-larbre-stopcovid-qui-cache-la-foret-health-data-hub-138852>] (consulté le 22/05/2023).
- FLORIDI, L., COWLS, J., BELTRAMETTI, M., *et al.* (2018) AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines* vol. 28, n° 4, p. 689-707. [<https://doi.org/10.1007/s11023-018-9482-5>]
- FRANCE CULTURE (2021) Une IA peut désormais vous vaincre à un jeu, même si elle n'en connaît pas les règles. In *France Culture*. [<https://www.radiofrance.fr/franceculture/une-ia-peut-desormais-vous-vaincre-a-un-jeu-meme-si-elle-n-en-connaît-pas-les-regles-3961209>] (consulté le 20/07/2022).
- FUTURE OF LIFE INSTITUTE (2017) ASILOMAR AI Principles. In *Future of Life*. [<https://futureoflife.org/2017/08/11/ai-principles/>] (consulté le 03/08/2022).
- GAO, S., LIMA, D. (2022) A review of the application of deep learning in the detection of Alzheimer's disease. *International Journal of Cognitive Computing in Engineering* vol. 3, p. 1-8. [<https://doi.org/10.1016/j.ijcce.2021.12.002>]
- GREENE, J.P., LEFEBVRE, S.L., WANG, M., *et al.* (2014) Risk factors associated with the development of chronic kidney disease in cats evaluated at primary care veterinary hospitals. *Journal of the*

- American Veterinary Medical Association* vol. 244, n° 3, p. 320-327.
[<https://doi.org/10.2460/javma.244.3.320>]
- GÜNERI, P., EPSTEIN, J.B. (2014) Late stage diagnosis of oral cancer: components and possible solutions. *Oral Oncology* vol. 50, n° 12, p. 1131-1136. [<https://doi.org/10.1016/j.oraloncology.2014.09.005>]
- GUTIERREZ C, N. (2022) L'intelligence artificielle nous rapproche du pancréas artificiel. In *Sciences et Avenir*. [https://www.sciencesetavenir.fr/sante/un-pas-de-plus-vers-le-pancreas-artificiel-pour-les-enfants-diabetiques_161063] (consulté le 20/07/2022).
- HAUTE AUTORITE DE SANTE (2012) Comprendre l'évaluation des dispositifs médicaux. In *HAS-sante*. [https://www.has-sante.fr/jcms/c_928541/fr/comprendre-l-evaluation-des-dispositifs-medicaux] (consulté le 14/11/2022).
- HAUTE AUTORITE DE SANTE (2011) La prise en charge des dispositifs médicaux par l'assurance maladie. In *Haute Autorité de Santé*. [https://www.has-sante.fr/jcms/c_930603/fr/la-prise-en-charge-des-dispositifs-medicaux-par-l-assurance-maladie] (consulté le 14/11/2022).
- HEALTHLEADERS (2010) Nurses Say Distractions Cut Bedside Time by 25%. In *Health Leaders Media*. [<https://www.healthleadersmedia.com/nursing/nurses-say-distractions-cut-bedside-time-25>] (consulté le 19/04/2023).
- HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE (2019) Ethics Guidelines for Trustworthy AI. [<https://ec.europa.eu/futurium/en/ai-alliance-consultation>] (consulté le 03/08/2022).
- HWANG, E.J., PARK, S., JIN, K.-N., *et al.* (2019) Development and Validation of a Deep Learning-Based Automated Detection Algorithm for Major Thoracic Diseases on Chest Radiographs. *JAMA Network Open* vol. 2, n° 3, p. e191095. [<https://doi.org/10.1001/jamanetworkopen.2019.1095>]
- HYUN, C.M., KIM, H.P., LEE, S.M., *et al.* (2018) Deep learning for undersampled MRI reconstruction. *Physics in Medicine and Biology* vol. 63, n° 13, p. 135007. [<https://doi.org/10.1088/1361-6560/aac71a>]
- IDEXX CANADA (2023) L'intelligence artificielle appliquée à la médecine vétérinaire mène à une efficacité et une précision accrues. In *Idexx*. [<https://ca.idexx.com/fr-ca/veterinary/analyzers/artificial-intelligence-veterinary-medicine-leads-to-efficiency/>] (consulté le 13/04/2023).
- ILHAN, B., GUNERI, P., WILDER-SMITH, P. (2021) The contribution of artificial intelligence to reducing the diagnostic delay in oral cancer. *Oral Oncology* vol. 116, p. 105254. [<https://doi.org/10.1016/j.oraloncology.2021.105254>]
- INSERM (2023) ChatGPT : l'IA prête à remplacer les chercheurs et les médecins, vraiment ? *Salle de presse de l'Inserm*. [<https://presse.inserm.fr/canal-detox/chatgpt-lia-prete-a-remplacer-les-chercheurs-et-les-medecins-vraiment/>] (consulté le 14/04/2023).
- JOHN-MATHEWS, J.-M. (2019) Interprétabilité en Machine Learning, revue de littérature et perspectives. In *Séminaire Good In Tech. "Développement de technologies responsables"*. [<https://hal.science/hal-02540761>] (consulté le 14/09/2023).
- KATZMAN, J.L., SHAHAM, U., CLONINGER, A., *et al.* (2018) DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Medical Research Methodology* vol. 18, n° 1, p. 24. [<https://doi.org/10.1186/s12874-018-0482-1>]
- LEBRATY, J.-F., LOBRE-LEBRATY, K., TRÉBUCQ, S. (2018) Chapitre 16. Logiciels et analyse de données qualitatives. In *Les méthodes de recherche du DBA, Business Science Institute*. Caen, EMS Editions, p. 283-305. [<https://doi.org/10.3917/ems.cheva.2018.01.0283>]
- LEE, A., MAVADDAT, N., WILCOX, A.N., *et al.* (2019) BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genetics in Medicine* vol. 21, n° 8, p. 1708-1718. [<https://doi.org/10.1038/s41436-018-0406-9>]
- LEE, Y.H. (2018) Efficiency Improvement in a Busy Radiology Practice: Determination of Musculoskeletal Magnetic Resonance Imaging Protocol Using Deep-Learning Convolutional Neural Networks. *Journal of Digital Imaging* vol. 31, n° 5, p. 604-610. [<https://doi.org/10.1007/s10278-018-0066-y>]
- LEEM (2019) La médecine : vision d'aujourd'hui et de demain. In *Leem - Les Entreprises du Médicament*. [<https://www.leem.org/la-medecine-vision-daujourd'hui-et-de-demain>] (consulté le 20/07/2022).
- LEGIFRANCE (2022) Sous-section 2 : Contenu de l'espace numérique de santé (Article R1111-27) - Légifrance. In *legifrance.gouv*.

- [<https://www.legifrance.gouv.fr/codes/id/LEGISCTA000043926406/2022-01-01>] (consulté le 17/09/2022).
- LIAW, S.-T., GUO, J.G.N., ANSARI, S., *et al.* (2021) Quality assessment of real-world data repositories across the data life cycle: A literature review. *Journal of the American Medical Informatics Association* vol. 28, n° 7, p. 1591-1599. [<https://doi.org/10.1093/jamia/ocaa340>]
- LIFELearn ANIMAL HEALTH (2019) LifeLearn Sofie: The Evolution of the Lightning-Fast Veterinary Medical Search Tool. *LifeLearn Inc.* [<https://www.lifelearn.com/2019/05/28/sofie-the-evolution/>] (consulté le 17/02/2023).
- MINISTÈRE DE LA SANTÉ ET DE LA PRÉVENTION (2023) Vivre avec une maladie chronique - Ministère de la Santé et de la Prévention. In *solidarites-sante.gouv.fr*. [<https://solidarites-sante.gouv.fr/soins-et-maladies/maladies/vivre-avec-une-maladie-chronique/>] (consulté le 05/08/2022).
- MINISTÈRE DE L'ÉCONOMIE, DES FINANCES ET DE LA SOUVERAINETÉ INDUSTRIELLE ET NUMÉRIQUE (2019) Le règlement général sur la protection des données (RGPD), mode d'emploi. In *Ministère de l'Économie, des Finances et de la Souveraineté Industrielle et Numérique*. [<https://www.economie.gouv.fr/entreprises/reglement-general-sur-protection-des-donnees-rgpd>] (consulté le 20/07/2022).
- MOREIRA, J., CARVALHO, A., HORVATH, T. (2018) A General Introduction to Data Analytics. Hoboken, NJ, John Wiley & Sons.
- MORRIS ANIMAL MORRIS ANIMAL FOUNDATION (2023) Golden Retriever Lifetime Study RFP. In *Morris Animal Foundation*. [<https://www.morrisanimalfoundation.org/golden-retriever-lifetime-study-rfp>] (consulté le 13/05/2023).
- NAGAMORI, Y., HALL SEDLAK, R., DEROSA, A., *et al.* (2020) Evaluation of the VETSCAN IMAGYST: an in-clinic canine and feline fecal parasite detection system integrated with a deep learning algorithm. *Parasites & Vectors* vol. 13, p. 346. [<https://doi.org/10.1186/s13071-020-04215-x>]
- NAM, J.G., KIM, M., PARK, J., *et al.* (2021) Development and validation of a deep learning algorithm detecting 10 common abnormalities on chest radiographs. *The European Respiratory Journal* vol. 57, n° 5, p. 2003061. [<https://doi.org/10.1183/13993003.03061-2020>]
- ODOXA (2018) La formation en santé des soignants est insatisfaisante. In *Odoxa*. [<http://www.odoxa.fr/sondage/a-lheure-digital-de-lia-formation-sante-professionnels-insatisfaisante-luniversite-laider-a-prendre-nouvelle-dimension/>] (consulté le 08/04/2022).
- OECD (2019) Applications de l'intelligence artificielle. In *L'intelligence artificielle dans la société*. [https://www.oecd-ilibrary.org/science-and-technology/l-intelligence-artificielle-dans-la-societe_47f9bf3a-fr] (consulté le 20/07/2022).
- OPPERMANN, A. (2021) Stochastic-, Batch-, and Mini-Batch Gradient Descent Demystified. In *Medium*. [<https://towardsdatascience.com/stochastic-batch-and-mini-batch-gradient-descent-demystified-8b28978f7f5>] (consulté le 22/05/2023).
- ORACLE (2022) Qu'est-ce que le Big Data Analytics ? In *Oracle*. [<https://blogs.oracle.com/oracle-france/post/quest-ce-que-le-big-data-analytics>] (consulté le 23/05/2023).
- ORACLE FRANCE (2023) Qu'est-ce que le Big Data ? In *Oracle*. [<https://www.oracle.com/fr/big-data/what-is-big-data/>] (consulté le 20/07/2022).
- ORDRE NATIONAL DES VÉTÉRINAIRES (2021) Atlas démographique de la profession vétérinaire 2021.
- ORGANISATION MONDIALE DE LA SANTÉ (2018) Maladies non transmissibles. In *World Health Organization*. [<https://www.who.int/fr/news-room/fact-sheets/detail/noncommunicable-diseases>] (consulté le 05/08/2022).
- OZER, E.M., ROWE, J., TEBB, K.P., *et al.* (2020) Fostering Engagement in Health Behavior Change: Iterative Development of an Interactive Narrative Environment to Enhance Adolescent Preventive Health Services. *The Journal of Adolescent Health* vol. 67, n° 2S, p. S34-S44. [<https://doi.org/10.1016/j.jadohealth.2020.04.022>]
- PAITRAUD, D. (2021) Diabète de type 1 : le système DIABELOOP désormais inscrit sur la LPPR. In *VIDAL*. [<https://www.vidal.fr/actualites/27892-diabete-de-type-1-le-systeme-diabeloop-desormais-inscrit-sur-la-lppr.html>] (consulté le 14/11/2022).

- PATEL, D., HALL, G.L., BROADHURST, D., *et al.* (2022) Does machine learning have a role in the prediction of asthma in children? *Paediatric Respiratory Reviews* vol. 41, p. 51-60. [https://doi.org/10.1016/j.prrv.2021.06.002]
- PAYNE, J.R., BRODBELT, D.C., LUIS FUENTES, V. (2015) Cardiomyopathy prevalence in 780 apparently healthy cats in rehoming centres (the CatScan study). *Journal of Veterinary Cardiology*, Supplement issue: The Feline Heart vol. 17, p. S244-S257. [https://doi.org/10.1016/j.jvc.2015.03.008]
- PEREZ, C., SOKOLOVA, K. (2020) Un Monde de Données: Initiation Sans Prérequis Au Domaine de la Donnée, Amazon Digital Services LLC.
- PETERSEN, R.C., SMITH, G.E., WARING, S.C., *et al.* (1999) Mild cognitive impairment: clinical characterization and outcome. *Archives of Neurology* vol. 56, n° 3, p. 303-308. [https://doi.org/10.1001/archneur.56.3.303]
- PICOXIA (2023) Interprétation radio instantanée en ligne. In *Picoxia*. [http://picoxia.com] (consulté le 13/04/2023).
- PRONZIA (2023) Pronzia Santé Animale et Technologies Numériques. In *ZAG by Domes Pharma*. [https://pronzia.fr/] (consulté le 13/04/2023).
- RAJPURKAR, P., IRVIN, J., BALL, R.L., *et al.* (2018) Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Medicine* vol. 15, n° 11, p. e1002686. [https://doi.org/10.1371/journal.pmed.1002686]
- RAO, A., PANG, M., KIM, J., *et al.* (2023) Assessing the Utility of ChatGPT Throughout the Entire Clinical Workflow. [https://www.medrxiv.org/content/10.1101/2023.02.21.23285886v1] (consulté le 14/04/2023).
- REAGAN, K.L., REAGAN, B.A., GILOR, C. (2020) Machine learning algorithm as a diagnostic tool for hypoadrenocorticism in dogs. *Domestic Animal Endocrinology* vol. 72, p. 106396. [https://doi.org/10.1016/j.domaniend.2019.106396]
- RHO, J., SHIN, S.-M., JHANG, K., *et al.* (2023) Deep learning-based diagnosis of feline hypertrophic cardiomyopathy. *PLoS One* vol. 18, n° 2, p. e0280438. [https://doi.org/10.1371/journal.pone.0280438]
- ROHMAH, R.N., SUSANTO, A., SOESANTI, I. (2013) Lung tuberculosis identification based on statistical feature of thoracic X-ray. In *2013 International Conference on QIR*. [https://ieeexplore.ieee.org/document/6632528] (consulté le 14/05/2023).
- ROWE, J.P., LESTER, J.C. (2020) Artificial Intelligence for Personalized Preventive Adolescent Healthcare. *The Journal of Adolescent Health* vol. 67, n° 2S, p. S52-S58. [https://doi.org/10.1016/j.jadohealth.2020.02.021]
- ROYAL VETERINARY COLLEGE (2023) VetCompass - Royal Veterinary College, RVC. In *rvc.ac.uk*. [https://www.rvc.ac.uk/vetcompass] (consulté le 04/04/2023).
- SAFAEI, M., SUNDARARAJAN, E.A., DRISS, M., *et al.* (2021) A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity. *Computers in Biology and Medicine* vol. 136, p. 104754. [https://doi.org/10.1016/j.compbiomed.2021.104754]
- SCHEMMEL, A., LEE, M., HANLEY, T., *et al.* (2016) Radiology Workflow Disruptors: A Detailed Analysis. *Journal of the American College of Radiology* vol. 13, n° 10, p. 1210-1214. [https://doi.org/10.1016/j.jacr.2016.04.009]
- SCHOFIELD, I., BRODBELT, D.C., KENNEDY, N., *et al.* (2021) Machine-learning based prediction of Cushing's syndrome in dogs attending UK primary-care veterinary practice. *Scientific Reports* vol. 11, n° 1, p. 9035. [https://doi.org/10.1038/s41598-021-88440-z]
- SEOANE, J., WARNAKULASURIYA, S., VARELA-CENTELLES, P., *et al.* (2006) Oral cancer: experiences and diagnostic abilities elicited by dentists in North-western Spain. *Oral Diseases* vol. 12, n° 5, p. 487-492. [https://doi.org/10.1111/j.1601-0825.2005.01225.x]
- SHIN, Y., KIM, S., LEE, Y.H. (2022) AI musculoskeletal clinical applications: how can AI increase my day-to-day efficiency? *Skeletal Radiology* vol. 51, n° 2, p. 293-304. [https://doi.org/10.1007/s00256-021-03876-8]

- SIGNALPET (2023) SignalRAY | Artificial Intelligence for Veterinary Radiology. In *signalpet.com*. [https://www.signalpet.com/products/signalray/] (consulté le 19/05/2023).
- SINGH, B., TAWFIK, H. (2020) Machine Learning Approach for the Early Prediction of the Risk of Overweight and Obesity in Young People. *Computational Science – ICCS 2020* vol. 12140, p. 523-535. [https://doi.org/10.1007/978-3-030-50423-6_39]
- SONI, D. (2019) Introduction to Bayesian Networks. In *Medium*. [https://towardsdatascience.com/introduction-to-bayesian-networks-81031eed94e] (consulté le 22/05/2023).
- STATISTA (2018) Health: openness towards AI France 2018. In *Statista*. [https://www.statista.com/statistics/1076992/medicine-opening-towards-l-intelligence-artificielle-la-france/] (consulté le 19/05/2023).
- SYNDICAT NATIONAL DES VETERINAIRES D'EXERCICE LIBERAL (2023) Nos missions. In *Vétérinaire libéral*. [https://www.veterinaireliberal.fr/qui-sommes-nous/nos-missions/] (consulté le 23/05/2023).
- THE HEALTH FOUNDATION (2015) Shaping the future. A strategic framework for a successful NHS. London, The Health Foundation.
- UNIVERSITÉ DE MONTRÉAL (2019) La Déclaration de Montréal en IA responsable. In *Déclaration de Montréal - IA Responsable*. [https://www.declarationmontreal-iaresponsable.com/la-declaration] (consulté le 03/08/2022).
- WALSH, S.L.F., CALANDRIELLO, L., SILVA, M., *et al.* (2018) Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *The Lancet. Respiratory Medicine* vol. 6, n° 11, p. 837-845. [https://doi.org/10.1016/S2213-2600(18)30286-8]
- WALTHAM (2019) AI-driven diagnostic can predict chronic kidney disease in cats two years early. In *Waltham*. [http://www.waltham.com/news-events/biomarkers/new-ai-diagnostic-tool-can-predict-chronic-kidney-disease-in-cats-two-years-early] (consulté le 19/05/2023).
- WIKIVERSITE (2017) Datawarehouse. In *wikiversity.org*. [https://fr.wikiversity.org/wiki/Datawarehouse] (consulté le 22/05/2023).
- WORLD HEALTH ORGANIZATION (2020) Urgent health challenges for the next decade. In *World Health Organization*. [https://www.who.int/news-room/photo-story/photo-story-detail/urgent-health-challenges-for-the-next-decade] (consulté le 20/07/2022).
- YALA, A., LEHMAN, C., SCHUSTER, T., *et al.* (2019) A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction. *Radiology* vol. 292, n° 1, p. 60-66. [https://doi.org/10.1148/radiol.2019182716]
- ZIMMER, M. (2018) Apprentissage par renforcement développemental. Thèse en doctorat informatique. Nancy, Université de Lorraine.

UTILISATION DE L'INTELLIGENCE ARTIFICIELLE DANS LA GESTION DES MALADIES CHRONIQUES DES ANIMAUX DE COMPAGNIE : OPPORTUNITÉS ET FREINS À L'ADOPTION

AUTEUR : Nicolas CHAUCHART DU MOTTAY

RÉSUMÉ :

Depuis le début des années 2000, l'utilisation des données par les entreprises a connu une progression inédite dans le sillage du développement d'outil permettant de les stocker et de les exploiter. Parmi les modèles d'exploitation des données, on trouve le domaine de l'Intelligence Artificielle, technologie dont le but est de permettre à des machines de réaliser des tâches plus ou moins complexes en mimant l'intelligence humaine.

Le domaine de la santé n'est pas étranger au développement de l'Intelligence Artificielle en général et du Machine Learning, discipline de l'Intelligence Artificielle. Si de nombreuses études démontrant les meilleures performances des algorithmes par rapport à des médecins spécialistes dans certaines applications ont été publiées, et si des nombreux outils utilisables par les professionnels de santé ont été développés, la diffusion de l'Intelligence Artificielle reste limitée. Les raisons de cette limitation viennent principalement du cadre légal encore inachevé, des questions éthiques que l'utilisation de tels algorithmes soulèvent ainsi que de la difficulté à intégrer ces outils dans les parcours de soin d'aujourd'hui. Cependant, de nombreuses entreprises travaillent sur le sujet et participent à la démocratisation d'outils utilisant l'Intelligence Artificielle en médecine humaine et notamment dans la prise en charge des patients atteints de maladies chroniques.

En revanche, en médecine vétérinaire, le développement de l'IA est très limité et inégal en fonction des pays. En France notamment, à peine une dizaine de solutions sont disponibles à l'utilisation pour les vétérinaires. Pourtant, le potentiel de l'Intelligence Artificielle à aider les vétérinaires dans leur pratique quotidienne est importante (diagnostic, prévention, suivi des patients, ...). Cette faible diffusion peut s'expliquer par la méconnaissance voire la méfiance de certains vétérinaires à l'encontre de ces outils mais aussi par la difficulté à développer ces outils (par manque de données souvent) et à les intégrer dans le mode de fonctionnement des cliniques.

MOTS CLÉS :

INTELLIGENCE ARTIFICIELLE, MACHINE LEARNING, INNOVATION DIGITALE, PRATIQUE VÉTÉRINAIRE, MALADIE CHRONIQUE, ANIMAL DE COMPAGNIE

JURY :

Président : Pr Henry CHATEAU

Directrice de thèse : Dr Chantal LEGRAND

Examinatrice : Dr Aline DE PAULA REIS

USING ARTIFICIAL INTELLIGENCE IN THE MANAGEMENT OF CHRONIC PET DISEASES: OPPORTUNITIES AND OBSTACLES TO ADOPTION

AUTHOR: Nicolas CHAUCHART DU MOTTAY

SUMMARY:

Since the early 2000s, the use of data by businesses has seen unprecedented growth in the wake of the development of tools for storing and exploiting data. Among the models of data exploitation is the field of Artificial Intelligence, a technology whose aim is to enable machines to perform more or less complex tasks by mimicking human intelligence.

The healthcare sector is not alien to the development of Artificial Intelligence in general and Machine Learning, a sub-discipline of Artificial Intelligence. Although many studies have been published showing better performances from algorithms compared to specialists in certain applications, and many tools that can be used by healthcare professionals have been developed, the diffusion of Artificial Intelligence remains limited. The main reasons for this limitation are the still unfinished legal framework, the ethical issues raised by the use of such algorithms, and the difficulty of integrating these tools into today's care pathways. However, many companies are working on the subject and are participating in the democratization of tools that use Artificial Intelligence in human medicine in general and in the care of patients with chronic diseases in particular.

On the other hand, in veterinary medicine, the development of AI is very limited and uneven across countries. In France, only a dozen solutions are available for use by veterinarians. Yet, the potential of Artificial Intelligence to help veterinarians in their daily practice is important. This low dissemination can be explained by the lack of knowledge or even distrust of these tools on the part of some veterinarians, but also by the difficulty of developing these tools (often due to a lack of data) and integrating them into the way clinics operate.

KEYWORDS:

ARTIFICIAL INTELLIGENCE, MACHINE LEARNING, DIGITAL INNOVATION, VETERINARY PRACTICE, CHRONIC DISEASE, PET

JURY:

Chairperson: Pr Henry CHATEAU

Thesis Director: Dr Chantal LEGRAND

Reviewer: Dr Alline DE PAULA REIS