



HAL
open science

L'intelligence artificielle dans l'industrie des produits de santé : de la recherche à la commercialisation

Inès Hamtache

► **To cite this version:**

Inès Hamtache. L'intelligence artificielle dans l'industrie des produits de santé : de la recherche à la commercialisation. Sciences pharmaceutiques. 2024. dumas-04747628

HAL Id: dumas-04747628

<https://dumas.ccsd.cnrs.fr/dumas-04747628v1>

Submitted on 22 Oct 2024

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.

THÈSE D'EXERCICE
pour le
DIPLÔME D'ÉTAT DE DOCTEUR EN PHARMACIE

Présentée et soutenue publiquement
le 24 mai 2024
par
Inès HAMTACHE

L'intelligence artificielle dans l'industrie des produits de santé : de la recherche à la commercialisation

Directeur de thèse : **Mme Magali VIVIER**

Jury

Président : **Mr David Balayssac**

Professeur des universités, HDR,
Docteur en pharmacie
UFR Pharmacie de Clermont-Ferrand

Membres : **Mme Magali VIVIER**

Maître de conférences, Docteur en pharmacie
UFR Pharmacie de Clermont-Ferrand

Mme Ghislain GARRAIT

Professeur des universités, HDR,
UFR Pharmacie de Clermont-Ferrand

Mme Sandrine BAGEL

Docteur en pharmacie,
CHU de Clermont-Ferrand

THÈSE D'EXERCICE
pour le
DIPLÔME D'ÉTAT DE DOCTEUR EN PHARMACIE

Présentée et soutenue publiquement
le 24 mai 2024
par
Inès HAMTACHE

L'intelligence artificielle dans l'industrie des produits de santé : de la recherche à la commercialisation

Directeur de thèse : **Mme Magali VIVIER**

Jury

Président : **Mr David Balayssac**

Professeur des universités, HDR,
Docteur en pharmacie
UFR Pharmacie de Clermont-Ferrand

Membres : **Mme Magali VIVIER**

Maître de conférences, Docteur en pharmacie
UFR Pharmacie de Clermont-Ferrand

Mme Ghislain GARRAIT

Professeur des universités, HDR,
UFR Pharmacie de Clermont-Ferrand

Mme Sandrine BAGEL

Docteur en pharmacie,
CHU Gabriel Montpied, Clermont-Ferrand

Remerciements

A Monsieur le Professeur David Balayssac, président du jury,

Pour m'avoir fait l'honneur d'accepter avec enthousiasme de présider le jury de ma thèse. Je vous adresse également toute ma reconnaissance pour m'avoir offert l'opportunité de réaliser un stage de recherche au Danemark et une année d'étude aux États-Unis, deux expériences qui ont profondément enrichi mon parcours académique et professionnel.

A Madame le Docteur Magali Vivier, directeur de thèse,

Pour avoir accepté d'encadrer mon travail et de faire partie de ce jury. Votre soutien, à travers vos lettres de recommandation et vos précieux conseils ont joué un rôle important dans mon parcours académique.

A Monsieur le Professeur Ghislain Garrait,

Pour vos enseignements et votre bienveillance tout au long de mes études. Je vous remercie sincèrement d'avoir accepté de faire partie de ce jury et de l'attention particulière portée à mon sujet.

A Madame le Docteur Sandrine Bagel,

Pour le stage très formateur que j'ai eu la chance de réaliser à vos côtés au sein de la PUI du CHU Gabriel-Montpied. Vos conseils avisés ont été un catalyseur précieux dans le développement de mes compétences tant professionnelles que personnelles. Je vous remercie chaleureusement de me faire l'honneur de faire partie de ce jury.

A mes parents,

Pour les efforts que vous avez consentis afin de me permettre de poursuivre des études aussi longues, et d'avoir cru en moi dans tous les projets que j'ai entrepris. Maman, tes encouragements constants, ton amour sans limite et ton soutien indéfectible ont été ma source de force et de courage. Papa, ta sagesse et ta présence bienveillante ont été des phares dans les moments où j'avais le plus besoin de recul et de relativiser. Votre soutien a été un pilier essentiel tout au long de mon parcours académique, et je vous en suis profondément reconnaissante. Cette réussite est aussi la vôtre.

A Alexandre,

Pour ta présence constante et ton soutien inébranlable. Merci d'être toujours là, à mes côtés, et de croire en moi. Tout au long de cette thèse et plus généralement de ces études, tu as été mon pilier, m'accompagnant et me soutenant, même dans les moments les plus difficiles. Tes encouragements, ta confiance et tes attentions au quotidien ont été précieux dans la réussite de cette thèse. Une nouvelle vie commence pour nous, remplie de nouvelles aventures à partager.

A Auréa,

Pour ta bonne humeur contagieuse et ta présence rassurante. Nos fous rires et les précieux moments que nous avons partagés ensemble ont largement contribué à embellir ces années d'études. Ta fidélité et le fait que nous puissions toujours compter l'une sur l'autre sont des trésors inestimables. Merci pour ces instants magiques et pour ceux qui viendront encore enrichir notre amitié.

A Christine et Christophe,

Pour votre accueil chaleureux et votre hospitalité lors de mon stage parisien qui resteront gravés dans ma mémoire. Votre générosité et votre altruisme sont pour moi une véritable source d'inspiration. Grâce à vous, j'ai pu me concentrer pleinement sur mes études tout en me sentant comme à la maison, et je vous en serai éternellement reconnaissante.

A **mes collègues** des différentes entreprises dans lesquelles j'ai eu le privilège de travailler, Je vous exprime ma profonde gratitude pour m'avoir toujours accueilli à bras ouverts au sein de vos équipes, et pour tout ce que vous m'avez appris.

A ma famille,

Mes taties, mes tontons, mes cousins, mes cousines, et ma belle-famille, pour votre soutien depuis toujours. Votre convivialité, votre bonne humeur et vos encouragements ont été très importants. Merci de faire partie intégrante de ma vie depuis tant d'années.

A mes amis,

Qu'elles soient des amitiés de longue date ou des rencontres plus récentes, ainsi qu'à toutes les personnes que j'ai eu le plaisir de croiser sur mon chemin, que ce soit aux Etats-Unis ou ailleurs, lors de ces voyages enrichissants qui ouvrent tant l'esprit, sans oublier ceux dont je me suis éloignée avec le temps et la distance, merci d'avoir illuminé mon parcours de votre amitié.

L'intelligence artificielle dans l'industrie des produits de santé : de la recherche à la commercialisation

REMERCIEMENTS	5
TABLE DES ILLUSTRATIONS	10
LISTE DES ABREVIATIONS	12
LEXIQUE	15
INTRODUCTION	17
PARTIE 1 : L'INTELLIGENCE ARTIFICIELLE : DEFINITIONS, HISTOIRE ET BASES THEORIQUES	18
I. DEFINITIONS	18
1. <i>La paradoxale "intelligence artificielle"</i>	<i>18</i>
2. <i>Définitions historiques</i>	<i>19</i>
3. <i>Définition actuelle</i>	<i>19</i>
II. HISTOIRE DE L'INTELLIGENCE ARTIFICIELLE	21
1. <i>Les prémices de l'IA : 1943 – 1956</i>	<i>21</i>
2. <i>L'âge d'or de l'IA : 1957 – 1974</i>	<i>25</i>
3. <i>La traversée de l'hiver : 1975-1980</i>	<i>27</i>
4. <i>Un regain d'intérêt : 1980-1987</i>	<i>27</i>
5. <i>Le second hiver de l'IA : 1987 – 1993</i>	<i>28</i>
6. <i>Les succès de l'IA : 1993 à aujourd'hui</i>	<i>28</i>
III. LES BASES THEORIQUES DE L'IA	29
1. <i>Les algorithmes : fondements incontournables de l'IA</i>	<i>29</i>
2. <i>L'apprentissage automatique</i>	<i>30</i>
3. <i>Les réseaux neuronaux et l'apprentissage profond (24) (45)</i>	<i>41</i>
4. <i>La vision par ordinateur</i>	<i>45</i>
5. <i>Le traitement automatique du langage naturel</i>	<i>45</i>
IV. LES DONNEES, SOURCE INDISPENSABLE POUR TOUT PROJET D'IA	46
1. <i>La révolution du 'Big Data'</i>	<i>46</i>
2. <i>Définition des mégadonnées</i>	<i>46</i>
3. <i>Les caractéristiques des mégadonnées</i>	<i>47</i>
4. <i>Les sources des mégadonnées</i>	<i>48</i>
5. <i>IA et données : une collaboration indispensable pour l'industrie des produits de santé</i>	<i>48</i>
PARTIE 2 : LES APPLICATIONS DE L'IA DANS L'INDUSTRIE DES PRODUITS DE SANTE	50
I. LES APPLICATIONS DE L'IA DANS LE CYCLE DU MEDICAMENT	50
1. <i>IA et recherche préclinique</i>	<i>52</i>
1.1. <i>Les différentes approches de la découverte d'un médicament</i>	<i>52</i>
1.2. <i>Recherche de la cible moléculaire</i>	<i>52</i>
1.3. <i>Identification et optimisation de candidats-médicaments</i>	<i>66</i>

2.	<i>IA et essais cliniques</i>	77
2.1.	Optimisation de la sélection et du recrutement des patients participants	77
2.2.	Suivi des patients	80
2.3.	Une meilleure valorisation des données d'essais cliniques	82
3.	<i>IA et production pharmaceutique</i>	83
3.1.	Surveillance des paramètres de production en temps réel	83
3.2.	Optimisation des processus de fabrication.....	84
3.3.	Amélioration de la planification de la production	85
4.	<i>IA et logistique pharmaceutique</i>	86
4.1.	Une chaîne logistique plus performante	86
4.2.	Lutte contre les ruptures d'approvisionnement.....	88
4.3.	Lutte contre les médicaments contrefaits	89
5.	<i>IA et pharmacovigilance</i>	93
6.	<i>IA et promotion pharmaceutique</i>	95
II.	L'IA, PILIER DE LA MEDECINE DU FUTUR.....	98
1.	<i>Des diagnostics plus précis et plus précoces</i>	98
2.	<i>Une prise en charge thérapeutique personnalisée</i>	100
3.	<i>Une médecine plus proactive : Prédire pour mieux prévenir</i>	105
4.	<i>Une médecine plus participative</i>	107
PARTIE 3 : LES PRINCIPAUX ENJEUX ET DEFIS LIES A L'INTELLIGENCE ARTIFICIELLE DANS L'INDUSTRIE DES PRODUITS DE SANTE		112
I.	LES ENJEUX ETHIQUES.....	112
1.	<i>Protection des données personnelles et de la vie privée des patients</i>	112
2.	<i>Opacité des décisions de l'IA</i>	115
3.	<i>Biais algorithmiques et équité dans les décisions de santé</i>	116
4.	<i>Utilisation non éthique des systèmes d'IA</i>	117
5.	<i>Enjeux éthiques liés au développement d'agents moraux artificiels</i>	118
6.	<i>Enjeux éthiques liés au pouvoir des entreprises commercialisant les systèmes d'IA</i>	119
II.	LES DEFIS REGLEMENTAIRES ET JURIDIQUES	119
1.	<i>Cadre réglementaire pour la protection des données</i>	120
2.	<i>Cadre réglementaire pour l'approbation des dispositifs médicaux basés sur l'IA</i>	120
3.	<i>Enjeux juridiques de la responsabilité décisionnelle</i>	124
4.	<i>Enjeux juridiques des collaborations pharmaceutiques - IA : questions de la propriété intellectuelle</i>	125
5.	<i>Harmonisation des normes réglementaires au niveau national et international</i>	126
III.	LES ENJEUX TECHNIQUES	127
1.	<i>Les défis liés à l'accès aux données</i>	127
2.	<i>Les défis liés à la standardisation des données et à l'interopérabilité des systèmes</i>	128
3.	<i>Sécurité et protection contre les cyberattaques</i>	130

IV. LES ENJEUX ECONOMIQUES.....	131
1. <i>Tarifification et modèle économique.....</i>	<i>131</i>
2. <i>Remboursement des soins impliquant l'IA par les assurances santé.....</i>	<i>133</i>
V. LES DEFIS ORGANISATIONNELS ET SOCIETAUX.....	133
1. <i>Impact sur les professionnels de santé.....</i>	<i>133</i>
2. <i>Enjeux organisationnels et managériaux.....</i>	<i>134</i>
3. <i>Les inégalités d'accès à l'IA et leur implication dans les injustices sociales.....</i>	<i>135</i>
4. <i>Impact de l'IA en santé sur la société et ses valeurs.....</i>	<i>136</i>
VI. LES ENJEUX ECOLOGIQUES.....	137
CONCLUSION.....	138
BIBLIOGRAPHIE.....	139

Table des illustrations

Figure 1 : Chronologie de l'IA, de ses origines à aujourd'hui.....	21
Figure 2 : Représentation mathématique / informatique d'un neurone biologique (5).....	22
Figure 3 : Photographie du SNARC (8)	24
Figure 4 : Photographie d'ELIZA, premier agent conversationnel de l'histoire (11).....	26
Figure 5 : Diagramme de Venn des relations entre IA, ML et apprentissage profond (21).....	30
Figure 6 : Les étapes clés pour développer un algorithme de ML (23).....	31
Figure 7 : Exemple d'un arbre de décision à partir de l'exemple de Quinlan (27) pour prédire si «oui ou non » le temps permettra de jouer à l'extérieur (28).....	33
Figure 8 : Illustration du fonctionnement de l'algorithme des k-voisins les plus proches (31).....	35
Figure 9 : Exemple de modèle SVM linéaire (35).....	37
Figure 10 : Illustration d'une fonction noyau (35).....	37
Figure 11 : Exemple de représentation graphique de la méthode des k-moyennes (43)	40
Figure 12 : Exemple d'un réseau de type MLP (47)	42
Figure 13 : Architecture d'un réseau de type CNN (47).....	43
Figure 14 : Architecture d'un RNN (48).....	44
Figure 15 : Le processus de recherche et développement (R&D) de médicaments (56).....	50
Figure 16 : Contributions de l'IA dans le cycle du médicament.....	51
Figure 17 : Arbre de décision pour la sélection de l'ensemble de données BeFree sur les associations gènes-maladies (61).....	55
Figure 18 : Chevauchement des associations gènes-maladies identifiées par BeFree avec les associations disponibles sur DisGeNET (61).....	55
Figure 19 : Gènes de la dépression identifiés par BeFree et leur chevauchement avec les gènes disponibles dans d'autres référentiels (61)	56
Figure 20 : Sous-typage des lignées cellulaires du cancer du pancréas et prédiction des réponses aux médicaments (62).....	57
Figure 21 : Illustration du MOGONET	58
Figure 22 : Courbes démontrant la supériorité de GCNLDA dans la prédiction des associations lncRNA-maladie pour 405 pathologies (64).....	59
Figure 23 : Diagramme de Venn des cibles prédites par le classificateur SVM pour le cancer du sein (BrCa), du pancréas (PaCa) et de l'ovaire (OvCa).....	59
Figure 24 : Arbre de décision généré par l'apprentissage de l'algorithme J48 pour identifier les gènes pharmacologiquement accessibles (druggable) (67)	60
Figure 25 : Flux de travail de la plateforme Open Targets (68)	62
Figure 26 : Processus de conception de médicaments basée sur la structure (58)	63
Figure 27 : Exemples de prédictions de structures protéiques générées par AlphaFold (74).....	64
Figure 28 : Page d'accueil de la base de données AlphaFold Protein Structure Database.....	65
Figure 29 : Domaines thérapeutiques du pipeline de R&D d'Atomwise (78)	67
Figure 30 : Outils de toxicologie in silico et étapes pour générer des modèles de prédiction (84).....	68
Figure 31 : Aperçu de l'ensemble de données du défi Tox21 et des 12 paramètres toxiques (87).....	69
Figure 32 : Construction de pharmacophores par des réseaux neuronaux (87)	69
Figure 33 : Prédiction des propriétés ADMET des médicaments par un réseau neuronal à base de graphes (65).....	73
Figure 34 : Architecture globale de Chemi-Net (98).....	74

<i>Figure 35 : Représentation du flux de travail pour une expérience générative de novo avec la plateforme Chemistry42 (103)</i>	75
<i>Figure 36 : Illustration de la conception de l'étude utilisant l'IAG pour réutiliser / réaffecter des médicaments (104)</i>	76
<i>Figure 37 : Illustration de l'utilisation d'Aicure (111)</i>	81
<i>Figure 38 : Configuration de la plateforme Data42 de Novartis (113)</i>	82
<i>Figure 39 : Interface de planification de Rombio (116)</i>	85
<i>Figure 40 : Infographie sur les pénuries de médicaments en 2022 (120)</i>	88
<i>Figure 41 : Exemple de boîte de médicaments avec un datamatrix (124)</i>	91
<i>Figure 42 : Photographie du RxScanner (126)</i>	92
<i>Figure 43 : Exemples d'utilisation de l'IA pour évaluer la toxicité en préclinique et la sécurité post AMM (91)</i>	93
<i>Figure 44 : Illustration du fonctionnement de Medication Shield (129)</i>	94
<i>Figure 45 : Exemple de tableau de bord de la plateforme Aktana (130)</i>	96
<i>Figure 46 : Parcours de soin avec et sans l'IA de Deepmind. AKI : Acute Kidney Injury (IRA) (133)</i>	99
<i>Figure 47 : Synergies entre l'IA et la médecine de précision (137)</i>	101
<i>Figure 48 : Résumé des principaux éléments de la biobanque britannique (139)</i>	102
<i>Figure 49 : Illustration du processus de radiogénomique (145)</i>	104
<i>Figure 50 : Comparaison des réponses prédites par l'algorithme et observées de deux patientes atteintes d'un cancer de l'ovaire aux traitements par carboplatine et paclitaxel (147)</i>	104
<i>Figure 51 : Indice de concordance pour chacun des modèles testés au cours des années (148)</i>	106
<i>Figure 52 : Augmentation des patients atteints de diabètes en France de 2015 à 2021 (153)</i>	108
<i>Figure 53 : Illustration de la K'watch Glucose et de son application (155)</i>	109
<i>Figure 54 : Exemples de capture d'écran de l'application Hillo (156)</i>	110
<i>Figure 55 : Photographie de Mallya installé sur un stylo Solostar (157)</i>	110
<i>Figure 56 : Enjeux éthiques et cycle de vie des données (158)</i>	113
<i>Figure 57 : Classification des DM</i>	121
<i>Figure 58 : Étendue du marquage CE médical (173)</i>	122
<i>Figure 59 : Fonctionnement du Health Data Hub (177)</i>	128
<i>Figure 60 : HL7, Pilier de l'interopérabilité dans le secteur de la santé (178)</i>	129

Liste des abréviations

ACC: American College of Cardiology

ADMET : Absorption, Distribution, Métabolisme, Excrétion et Toxicité

AHA : American Heart Association

AKI : Acute Kidney Injury

AMM : Autorisation de Mise sur le Marché

ANFR : Agence Nationale des FRéquences

ANSM : Agence Nationale de Sécurité du Médicament et des produits de santé

APHP : Assistance Publique - Hôpitaux de Paris

AUC : Area Under the Curve

AVC : Accident Vasculaire Cérébral

AWS : Amazon Web Services

BBMRI : Biobanking and Biomolecular Resources Research Infrastructure

BrCa : Breast Cancer

CDC : Centers for Disease Control and Prevention

CE : Conformité Européenne

CNEDiMTS : Commission Nationale d'Evaluation des Dispositifs Médicaux et des Technologies de Santé

CNN : Convolutional Neural Network

CTM : Clinical Trial Matching

CRPV : Centres Régionaux de PharmacoVigilance

DL : Deep Learning

DL50 : Dose Létale à 50%

DNN : Deep Neural Network

DM : Dispositif Médical

DME : Dossiers Médicaux Électroniques

EI : Effet Indésirable

EMBL : European Molecular Biology Laboratory

ERP : Enterprise Resource Planning

FPR : False Positive Rate

GCN : Graph Convolutional Network

GRU : Gated Recurrent Unit

HAS : Haute Autorité de Santé

HDAC1 : Histone DésACétylase 1
HIPAA : Health Insurance Portability and Accountability Act
IBM : International Business Unit
IA : Intelligence Artificielle
IAG : Intelligence Artificielle Générative
IoT : Internet of Things
IRA : Insuffisance Rénale Aiguë
IRM : Imagerie par Résonance Magnétique
LBDD : Ligand Based Drug Design
LMA : Leucémie Myéloïde Aiguë
lncARN : ARN longs non codants
LSTM : Long Short-Term Memory
MESA : Multi-Ethnic Study of Atherosclerosis
MIT : Massachusetts Institute of Technology
ML : Machine Learning
MLP : Multi-Layer Perceptron
MOGONET : Multi-Omics Graph cOnvolutional NETworks
NLP : Natural Language Processing
OCDE : Organisation de Coopération et de Développement Économiques
OMS : Organisation Mondiale de la Santé
ONUDC : Office des Nations Unies contre la Drogue et le Crime
OTC : Over-The-Counter
OvCa : Ovarian Cancer
PaCa : Pancreatic Cancer
PMIDs : PubMed IDs
PR : Precision-Recall
PV : PharmacoVigilance
QSAR : Quantitative Structure-Activity Relationship
RF : Random Forest
RGPD : Règlement Général sur la Protection des Données
RNA : Réseau de Neurones Artificiels
RMN : Résonance Magnétique Nucléaire
RNN : Recurrent Neural Network
ROC : Receiver Operating Characteristic

SBDD : Structure-Based Drug Design
SMILES : Simplified Molecular Input Line Entry System
SNARC : Stochastic Neural Analog Reinforcement Calculator
SNIA : Stratégie Nationale pour l'Intelligence Artificielle
SVM : Support Vector Machine
SVR : Support Vector Regression
TALN : Traitement Automatique du Langage Naturel
TIC : Technologies de l'Information et de la Communication
TPR : True Positive Rate
UE : Union Européenne
VCDN : View Correlation Discovery Network

Lexique

La complexité inhérente de l'intelligence artificielle implique l'utilisation de nombreux acronymes et termes techniques, souvent anglophones. Pour faciliter la navigation du lecteur, un lexique a été établi, présentant ces termes dans l'ordre alphabétique des abréviations utilisées et permettant ainsi des références rapides aux concepts clés tout au long de la lecture. En complément, cette thèse est agrémentée de notes de bas de page, qui fournissent des définitions approfondies des concepts complexes et les replacent dans le contexte de leur utilisation, contribuant ainsi à enrichir la compréhension du sujet.

Abréviation utilisée	Notion correspondante	Définition
CNN	Convolutional Neural Network (Réseau de neurones convolutifs)	Type spécifique de réseau de neurones profonds utilisé principalement dans le traitement d'images
DL	Deep Learning (Apprentissage profond)	Branche de l'IA qui se concentre sur l'apprentissage à partir de données non structurées ou peu structurées
DNN	Deep Neural Network (Réseau de neurones profond)	Réseau de neurones comportant plusieurs couches cachées
DME	Dossier Médical Électronique	Dossier médical stocké électroniquement, facilitant l'accès, la gestion et le partage des informations médicales des patients
GCN	Graph Convolutional Network (Réseau de convolution sur graphe)	Type de réseau de neurones utilisé pour le traitement des données structurées sous forme de graphes
IAG	Intelligence Artificielle Générative	IA capable de générer de nouvelles données réalistes, souvent utilisé dans la création d'images, de textes ou de sons
k-means	Méthode des k-moyennes	Algorithme de clustering non supervisé qui sépare un ensemble de données en k groupes en fonction de leurs similarités

Abréviation utilisée	Notion correspondante	Définition
kNN	k-Nearest Neighbors (k plus proches voisins)	Algorithme de classification qui utilise les étiquettes des k points les plus proches pour prédire la classe d'un nouvel échantillon
LSTM	Long Short-Term Memory	Type de réseau de neurones récurrents (RNN) spécialisé dans le traitement de séquences de données, capables de conserver des informations sur de longues périodes de temps.
ML	Machine Learning (Apprentissage automatique)	Branche de l'IA qui permet aux ordinateurs d'apprendre à partir de données sans être explicitement programmés.
MLP	Multilayer Perceptron (Perceptron multicouche)	Type de réseau de neurones artificiels composé de plusieurs couches de neurones dont au moins une couche cachée.
NLP	Natural Language Processing (Traitement du langage naturel)	Domaine de l'IA qui se concentre sur la compréhension et la manipulation du langage humain par les ordinateurs.
RF	Random Forest (Forêt d'arbres décisionnels)	Modèle d'apprentissage automatique utilisé pour la classification et la régression, composé de plusieurs arbres de décision.
RNA	Réseau de Neurones Artificiels	Modèle informatique inspiré du cerveau humain, composé de neurones interconnectés et utilisé pour effectuer des tâches d'apprentissage et de reconnaissance de motifs.
RNN	Recurrent Neural Network (Réseau de neurones récurrent)	Type de réseau de neurones qui prend en compte les séquences et l'ordre des données en entrée, souvent utilisé dans le traitement de données séquentielles, en particulier dans le NLP.
SVM	Support Vector Machine (Machine à vecteurs de support)	Algorithme d'apprentissage supervisé utilisé pour la classification qui trouve un hyperplan optimal pour séparer les différentes classes.
SVR	Support Vector Regression (Régression par machine à vecteurs de support)	Variante des machines à vecteurs de support utilisée pour la régression au lieu de la classification.

Introduction

Dans un monde en perpétuelle évolution, où la technologie joue un rôle central, l'intelligence artificielle (IA) prend une importance croissante. Au fil des décennies, l'IA a évolué d'un concept futuriste à une réalité omniprésente, transformant de nombreux secteurs de la société. Son impact se démarque particulièrement dans l'industrie des produits de santé, où cette révolution technologique redéfinit les frontières entre la science et la technologie, ouvrant ainsi la voie à des approches médicales novatrices et à des progrès sans précédent. De la recherche préclinique à la commercialisation des médicaments, l'IA réinvente chaque étape du processus. Si l'IA est appelée à jouer un rôle central dans la médecine de demain, elle soulève également une série de défis et d'interrogations, notamment en matière d'éthique, de réglementation et de sécurité des données.

Face à ces mutations profondes, l'IA suscite des opinions divergentes : d'un côté, ses partisans voient en elle une révolution aux multiples promesses, tandis que de l'autre, ses détracteurs redoutent les conséquences néfastes qu'elle pourrait engendrer. Au cœur de ce débat, cette thèse adopte une perspective équilibrée en examinant de manière approfondie les applications prometteuses de l'IA dans l'industrie des produits de santé, mais aussi les enjeux et les défis qui en découlent.

L'objectif principal de cette recherche est donc d'analyser en profondeur comment l'IA transforme l'industrie pharmaceutique et médicale, ainsi que les implications de cette transformation pour l'avenir de la santé. En nous appuyant sur des études scientifiquement validées et des cas concrets d'utilisation, nous nous efforçons d'apporter un éclairage nouveau sur ce sujet complexe et de contribuer aux débats académiques et sociétaux en cours.

Dans cette perspective, cette thèse se structure en trois grandes parties. La première partie commencera par définir l'IA, abordera son évolution historique puis ses bases théoriques et enfin l'importance des données. La deuxième partie se concentrera sur les multiples applications de l'IA dans les industries de santé, en mettant en lumière son rôle tout au long du cycle du médicament et de la médecine du futur. Enfin, la troisième partie explorera les principaux défis et enjeux liés à l'intégration de l'IA dans l'industrie des produits de santé, en examinant notamment les questions éthiques, réglementaires et sociales qui en découlent.

PARTIE 1 : L'intelligence artificielle : Définitions, histoire et bases théoriques

L'intelligence artificielle (IA), à la fois fascinante et omniprésente, a déjà et va continuer à profondément remodeler notre façon de vivre. De la fascination suscitée par les machines intelligentes dans les œuvres de science-fiction à la réalité tangible des systèmes d'IA qui font aujourd'hui partie de notre quotidien, cette technologie est devenue incontournable. Mais qu'englobe véritablement le concept d'IA ? Cette première partie de notre exploration définit l'IA, retrace son histoire captivante, pose les bases théoriques de cette technologie et enfin met en lumière l'importance des données dans la mise en place de systèmes d'IA.

I. Définitions

L'IA a toujours captivé, suscitant débats et discussions animées. Son évolution constante et ses caractéristiques techniques complexes rendent sa définition délicate, nécessitant une compréhension approfondie de ses divers aspects. Notre cherchons ici à lever le voile sur l'IA, et à en saisir sa véritable signification. Ainsi, cette première section se penchera sur les multiples dimensions de l'intelligence artificielle, en explorant le caractère paradoxal de cette expression et en retraçant ses définitions historiques et contemporaines.

1. La paradoxale "intelligence artificielle"

L'expression "intelligence artificielle" semble paradoxale à première vue, combinant deux concepts en apparence contradictoires : "l'intelligence", qui englobe la conscience, la réflexion et la résolution de problèmes, et le qualificatif "artificiel", qui renvoie à ce qui est créé par l'homme plutôt que par la nature. Cette dualité suscite des débats philosophiques et conceptuels sur la nature réelle de l'IA, remettant en question la possibilité d'attribuer une forme d'intelligence à des entités artificielles dénuées de l'intelligence innée des êtres vivants. Ainsi, la question fondamentale demeure : Comment les entités artificielles peuvent-elles être qualifiées d'intelligentes ?

2. Définitions historiques

2.1. Définition de John McCarthy

L'expression « Intelligence Artificielle » a été introduite lors d'un colloque scientifique en 1956 par le mathématicien américain **John McCarthy** qui la définit alors comme la science de « **rendre des machines intelligentes** ». (1)

2.2. Définition de Marvin Minsky

Marvin Minsky, professeur émérite au MIT (Massachusetts Institute of Technology) et l'un des pères fondateurs de l'IA, a apporté une précision à cette définition en décrivant l'IA comme « *la construction de programmes informatiques capables d'accomplir des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisantes par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique* ». (2) L'ambition était donc de concevoir des appareils dotés d'une intelligence leur permettant de **reproduire les comportements humains** dans des activités de raisonnement. Toutefois, il est important de noter que cette définition tend à devenir partiellement **obsolète** car l'IA a depuis surpassé l'Homme dans certains domaines spécifiques.

3. Définition actuelle

Selon l'Encyclopédie du Larousse, l'IA est définie de nos jours comme « *l'ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine* ». (3) Ici, la notion de **simulation** fait son entrée, laissant transparaître le fait que les machines ne sont pas réellement intelligentes.

Les définitions historiques et contemporaines partagent un socle conceptuel commun : l'hypothèse que l'intelligence humaine et toutes ses fonctions cognitives (telles que l'apprentissage, le raisonnement, la perception, la mémorisation ou encore le calcul) peuvent être **décrites avec une précision suffisante** pour concevoir des machines capables de les reproduire.

L'IA ne fait donc pas seulement référence à la réalisation de machines capables de simuler l'intelligence humaine mais également à un **domaine scientifique** visant à **comprendre le fonctionnement de cette intelligence**. Cette perspective est clairement exprimée dans le Rapport Villani rendu public en 2018 : « L'intelligence artificielle désigne [en effet] moins un champ de recherches bien défini qu'un programme, fondé autour d'un objectif ambitieux : comprendre comment fonctionne la cognition humaine et la reproduire ; créer des processus cognitifs comparables à ceux de l'être humain. » (4)

Ainsi, l'IA ne se limite pas à une simple technologie, mais représente plutôt un **ensemble de technologies interconnectées** qui constituent un domaine scientifique à la convergence des sciences cognitives, de l'informatique, des mathématiques, de la neurobiologie et les sciences du langage. Cette **transdisciplinarité** offre un terreau fertile pour **l'innovation** mais contribue également à la **complexité** de l'IA. Il est important de souligner qu'il n'existe pas une seule intelligence artificielle, mais bien **plusieurs facettes de l'IA**, chacune étant spécialisée dans un domaine particulier.

Pour conclure cette partie sur la définition de l'IA, il est essentiel de comprendre que l'IA ne prétend pas à la possession d'une intelligence authentique ou de conscience. Elle se réfère plutôt à la capacité des systèmes informatiques à **imiter** certaines formes d'intelligence humaine en **traitant des informations de manière complexe**. L'IA se fonde sur des méthodes et des techniques sophistiquées pour permettre aux machines de **traiter des données**, de **décoder des motifs** et de **prendre des décisions**, mais cela ne signifie pas qu'elles possèdent une véritable compréhension ou une conscience. Par conséquent, les scénarios catastrophes souvent envisagés dans les œuvres de science-fiction semblent peu plausibles face aux progrès réels de l'IA.

II. Histoire de l'intelligence artificielle

L'histoire de l'IA (*Figure 1*) est jalonnée de découvertes majeures et de périodes de ralentissement qui ont profondément influencé son développement. Ces moments clés ont pavé la voie à la multitude d'applications dans l'industrie de la santé développées tout au long de cette thèse.

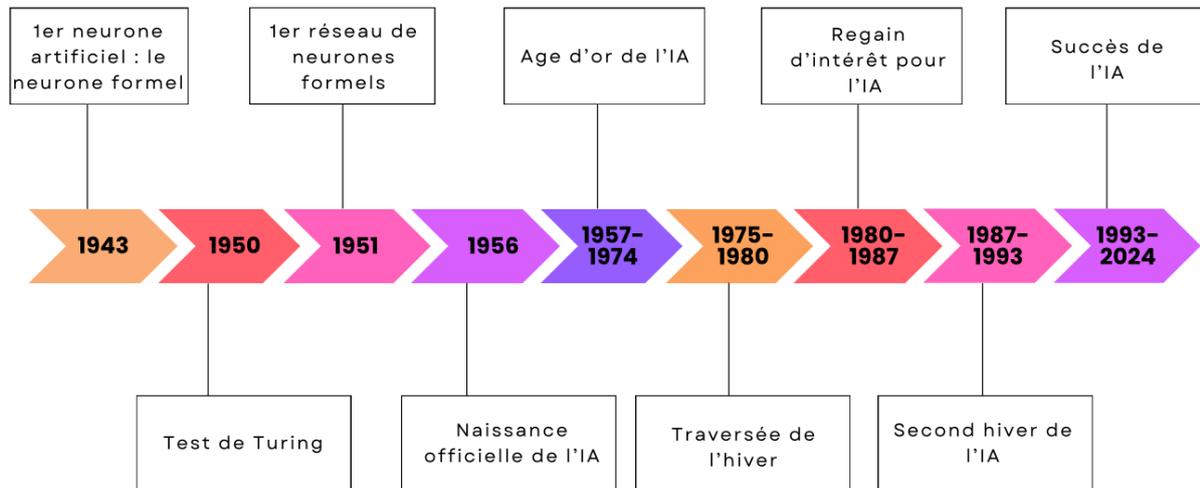


Figure 1 : Chronologie de l'IA, de ses origines à aujourd'hui

1. Les prémices de l'IA : 1943 – 1956

1.1. Le premier neurone artificiel : le neurone formel

Dès 1943, bien avant l'émergence du terme "Intelligence Artificielle", le neurophysiologiste américain Warren **McCulloch** et le logicien Walter **Pitts** ont jeté les bases du domaine en proposant le premier modèle de neurone artificiel : le "**neurone formel**". Ce concept novateur, toujours pertinent aujourd'hui, s'inspire largement du fonctionnement des neurones biologiques. McCulloch et Pitts affirment que le neurone constitue l'élément fondamental de toute opération logique et que des réseaux de neurones soigneusement conçus peuvent être capables d'apprentissage.

Dans le système nerveux biologique, les neurones émettent des signaux électriques appelés **influx nerveux**, qui sont reçus par les neurones au niveau de leurs **dendrites**. Chaque neurone possède un **axone** chargé de transmettre ces influx au neurone voisin, favorisant l'activité cérébrale. La synapse, zone de contact entre deux neurones, joue un rôle clé dans ce processus. L'activation des neurones voisins suit une **règle binaire** : elle se déclenche selon un principe de « tout ou rien », en fonction de la nature des signaux reçus (Figure 2).

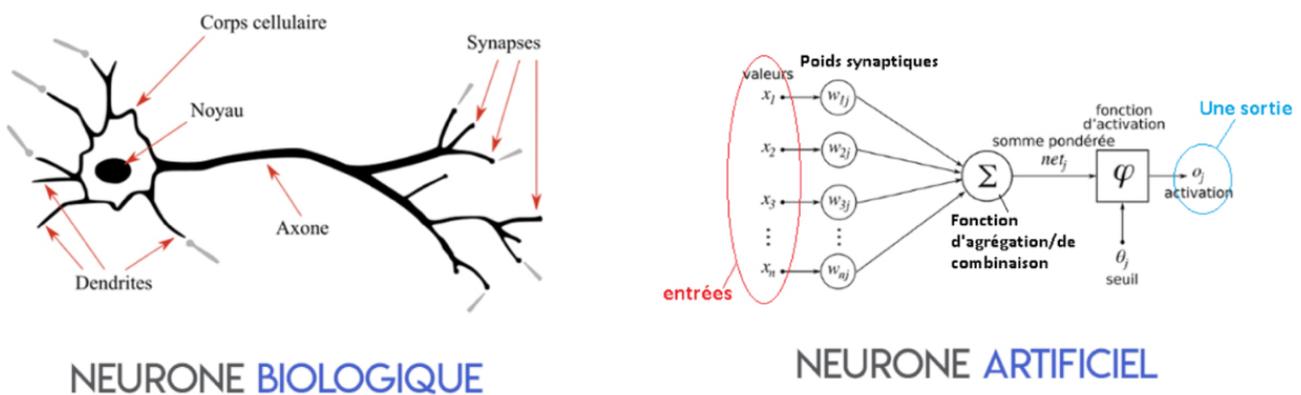


Figure 2 : Représentation mathématique / informatique d'un neurone biologique (5)

De manière analogue au neurone biologique, un neurone artificiel reçoit une multitude de signaux d'entrée, comparables aux influx reçus par les dendrites des neurones biologiques. Ensuite, le neurone artificiel effectue une **somme pondérée** de ces signaux d'entrée, attribuant des poids à chacun en fonction de leur importance relative. Cette somme pondérée est alors soumise à une **fonction d'activation**, qui agit comme le corps cellulaire des neurones biologiques pour réguler la réponse du neurone. Une fois activé, le neurone transmet les signaux de sortie aux neurones connectés suivants, agissant ainsi comme des axones dans le système biologique.

Cette représentation mathématique et informatique simplifiée du neurone physiologique, fonctionnant en mode **binaire** où la sortie est évaluée à 0 ou 1, a jeté les bases de l'IA moderne.

(6)

1.2. Le test de Turing

En **1950**, dans son article *Computing Machinery and Intelligence* (7), l'éminent mathématicien britannique **Alan Turing** tente de répondre à la question centrale de l'IA : « Peut-on attribuer aux machines la capacité de penser ? ». Il développe alors le « **test de Turing** », qui vise à évaluer la capacité d'une machine à **simuler de manière convaincante l'intelligence humaine**. Trois participants sont impliqués : un juge humain, un autre être humain et une machine. Le juge, isolé dans une pièce, dialogue par écrit avec les deux interlocuteurs sans les voir. La machine cherche à répondre de manière à ce que le juge ne puisse pas facilement distinguer la machine de l'être humain. Si le juge ne peut pas faire la distinction après cinq minutes de dialogue, la machine réussit le test, démontrant sa capacité à communiquer de manière similaire à un être humain. Ce test a eu un impact significatif sur la réflexion en IA et continue d'alimenter les débats dans ce domaine.

1.3. SNARC : le premier réseau de neurones formels

En 1951, **Marvin Minsky** et **Dean Edmonds**, deux doctorants en mathématiques américains, ont construit le premier **réseau de neurones formels** : le Stochastic Neural Analog Reinforcement Calculator (SNARC).

Composé de **quarante neurones artificiels** interconnectés par des **synapses**, le SNARC **ajuste le poids** de ces synapses en fonction des succès ou des échecs dans l'exécution de tâches spécifiques. L'expérience fondamentale du SNARC simulait le cerveau d'un rat apprenant à se diriger dans un labyrinthe, avec la progression du rongeur fictif suivie sur un réseau lumineux.

Chaque neurone du SNARC possède une **mémoire à court terme** grâce à un condensateur stockant l'énergie, lui permettant de se souvenir pendant quelques secondes si le neurone a récemment relayé un signal. De plus, chaque neurone dispose d'une **mémoire à long terme** générée par un potentiomètre permettant d'ajuster la probabilité de transmission d'un signal sortant en fonction de la réussite ou non de la tâche. Cette combinaison de paramètres, analogue aux poids dans les réseaux de neurones numériques actuels, a permis au SNARC **d'apprendre à résoudre un labyrinthe** (*Figure 3*).

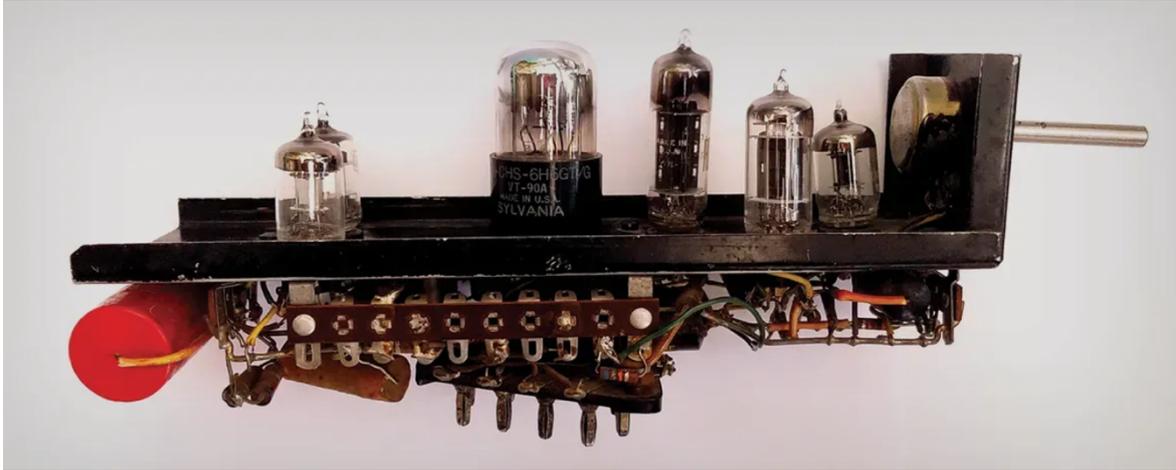


Figure 3 : Photographie du SNARC (8)

L'importance historique du SNARC réside dans le fait qu'il a été l'une des premières incarnations concrètes des concepts d'IA. Bien que ses capacités aient été limitées par rapport aux systèmes modernes, le SNARC a jeté les bases de la recherche en IA, montrant que les machines pouvaient **apprendre** et **s'adapter** à partir d'expériences.

1.4. Naissance officielle de l'IA : 1956

L'IA a été officiellement initiée en **1956** lors du **Dartmouth Summer Research Project on Artificial Intelligence**, organisé par **Marvin Minsky** et **John McCarthy** au Dartmouth College, dans le New Hampshire (États-Unis). Cette conférence de huit semaines a réuni vingt chercheurs et pionniers de l'IA, définissant **les objectifs et les contours** de ce nouveau domaine de recherche. Les participants ont affirmé qu'une fois qu'une machine pouvait accomplir une tâche, il était théoriquement possible de programmer une calculatrice automatique pour reproduire son comportement. Cet événement historique a consacré l'émergence de l'IA en tant que domaine autonome, axé sur la **compréhension et la simulation des processus cognitifs**, englobant divers aspects de l'intelligence humaine tels que la maîtrise du langage et le raisonnement. (9)

2. L'âge d'or de l'IA : 1957 – 1974

2.1. Le premier algorithme d'apprentissage automatique : le perceptron

En 1957, **Frank Rosenblatt**, scientifique américain, révolutionne le domaine de l'IA en introduisant le **perceptron**, un modèle de réseau de neurones artificiels conçu pour la **classification automatique des données**. Le perceptron se démarque par sa capacité à résoudre des **problèmes de classification binaire de manière autonome**, après avoir été entraîné avec un ensemble de données. Cette innovation marque le début de **l'apprentissage automatique** en tant que discipline distincte, explorée plus en détail dans le chapitre III.

La publication de Rosenblatt en 1958 intitulée *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain*, et la construction du **Mark 1 Perceptron** achevée en 1960 matérialisent cette avancée. La machine utilise **un réseau de neurones** disposés en une couche unique et dont les poids synaptiques peuvent être modifiés par rétroaction, permettant ainsi à la machine **d'apprendre de ses réussites et de ses échecs**.

Malgré certaines critiques sur sa simplicité, le perceptron pose les fondements du développement ultérieur des réseaux de neurones profonds et d'autres techniques d'apprentissage automatique, incarnant ainsi l'esprit pionnier de l'IA et ouvrant de nouvelles perspectives pour l'avenir de cette discipline. (10)

2.2. Le premier agent conversationnel : ELIZA

En **1966**, ELIZA, un programme d'IA développé par **Joseph Weizenbaum**, chercheur au MIT, a réussi pour la première fois le test de Turing. Ce programme, bien que rudimentaire par rapport aux normes actuelles, est remarquable pour son époque. ELIZA est un programme qui simule un psychothérapeute (*Figure 4*). L'utilisateur peut interagir avec ELIZA en utilisant des affirmations ou des questions, et le programme analyse le texte pour fournir des réponses cohérentes. Si ELIZA se contente souvent de faire répéter son interlocuteur, de lui renvoyer des questions ou encore de donner des réponses génériques, elle a réussi à convaincre certains utilisateurs qu'ils dialoguaient avec un être humain.



Figure 4 : Photographie d'ELIZA, premier agent conversationnel de l'histoire (11)

ELIZA est donc considérée comme l'un des premiers **agents conversationnels (chatbots)** de l'histoire et le pionnier du **Traitement Automatique du Langage Naturel (TALN)**, ou **Natural Language Processing (NLP)** en anglais. Bien qu'aujourd'hui ELIZA aurait du mal à duper un être humain, ce programme reste une prouesse technique pour son époque et a ouvert la voie à des développements ultérieurs dans le domaine du NLP. (12)

2.3. MYCIN, un jalon significatif dans l'utilisation de l'IA en médecine

En 1974, MYCIN est lancé pour assister les professionnels de santé dans le diagnostic des infections bactériennes et la prescription d'antibiotiques et constitue un exemple éloquent de ce que sont les systèmes experts, définis comme des « *ensembles de logiciels dont les capacités de résolution de problèmes nouveaux dans un domaine donné sont assimilables à celle d'un expert humain spécialiste de ce domaine* » (13). Élaboré sous la direction d'**Edward Shortliffe**, MYCIN repose sur une **base de connaissances** contenant des **règles cliniques** pour recommander des **traitements antibiotiques appropriés** en fonction des symptômes du patient. La forte valeur ajoutée de MYCIN était de fournir aux médecins des recommandations avisées dans les **situations d'urgence** où un diagnostic rapide était essentiel, avant même que les informations complètes sur l'organisme ne soient disponibles, et **en ajustant la posologie** en fonction du poids du patient. (14)

Malgré son importance, MYCIN a également soulevé des défis en matière de **gestion des connaissances médicales** et de **fiabilité des systèmes experts**, ouvrant ainsi la voie à de futures recherches en IA médicale.

3. La traversée de l'hiver : 1975-1980

La période allant de **1975 à 1980** est souvent qualifiée de "**traversée de l'hiver**" pour l'IA. Durant cette phase, l'IA est confrontée à un **désintérêt** marqué et à un **désengagement financier** significatif. Les promesses trop ambitieuses des débuts se transforment en déceptions, principalement en raison de **capacités matérielles et de puissance de calculs des ordinateurs insuffisantes** ainsi que des performances limitées des logiciels de cette époque. De nombreux projets sont alors abandonnés, suscitant des **critiques** de la part de la communauté scientifique et du grand public. Malgré cela, cette période n'a pas été stérile, car elle a permis de réfléchir, de réévaluer les objectifs et d'affiner les approches, ouvrant la voie à une nouvelle ère de progrès dans les années suivantes. (15)

4. Un regain d'intérêt : 1980-1987

Les années 1980 marquent **un retour en force** de l'IA. De nombreux projets en veille sont réactivés, entraînant des développements significatifs. Parmi ces avancées notables, la **théorie du connexionnisme** gagne en crédibilité, explorant la **modélisation des processus mentaux** au moyen de **réseaux de neurones formels**. Les travaux du physicien **John Hopfield** en **1982** ont particulièrement contribué à cette évolution, en démontrant la capacité des réseaux neuronaux à **apprendre** et à **traiter des informations** (16). En parallèle, les **systèmes experts** contribuent à rendre l'IA plus utile pour la société. L'**essor de l'informatique**, notamment grâce aux ordinateurs de cinquième génération et à l'augmentation de la puissance de calcul, a favorisé le lancement de nombreux projets. Les financements reprennent et les grandes entreprises commencent à percevoir des opportunités d'applications commerciales dans les progrès de l'IA.

5. Le second hiver de l'IA : 1987 – 1993

Après cette période de succès, l'espoir cède à nouveau sa place à la désillusion. La **puissance de calcul** des ordinateurs a surpassé celle des systèmes d'IA, et les **coûts de maintenance** des équipements d'IA s'avèrent plus élevés que prévu, créant de nouveaux défis financiers. Cette phase difficile, souvent qualifiée de deuxième "hiver de l'IA", se caractérise par une réduction des financements gouvernementaux et des investissements privés, ce qui ralentit considérablement les progrès de la recherche dans le domaine.

6. Les succès de l'IA : 1993 à aujourd'hui

A partir de 1993, l'évolution technologique a déclenché une véritable révolution, marquée par la diffusion généralisée des ordinateurs personnels et la démocratisation d'internet, suivie de l'avènement des smartphones. Ces avancées ont offert un terrain propice à l'essor de l'IA qui a réalisé à son tour des exploits remarquables. En 1997, **Deep Blue**, un superordinateur développé par IBM (International Business Machines) spécialisé dans le jeu d'échecs, a triomphé du champion du monde de la discipline, Garry Kasparov, démontrant ainsi la capacité de l'IA à surpasser l'intelligence humaine dans des domaines intellectuellement exigeants. En 2016, **AlphaGo**, développé par la société britannique DeepMind Technologies, a créé l'événement en battant les meilleurs joueurs mondiaux de Go, un jeu réputé pour sa complexité. (17). Les années 2020 sont témoins de la démocratisation de l'IA, symbolisée notamment par l'avènement de ChatGPT.

III. Les bases théoriques de l'IA

Avant d'explorer les bénéfices potentiels de l'IA dans les industries de santé, il est important de comprendre certains **concepts fondamentaux** de ce domaine à la fois complexe et dynamique. Ainsi, cette section se concentre sur les bases théoriques indispensables pour se familiariser avec le vocabulaire et les notions clés associés à l'IA, en mettant particulièrement l'accent sur celles pertinentes pour les industries de la santé. La compréhension de ces principes facilitera l'exploration des applications concrètes de l'IA dans le domaine de la santé ainsi que des défis qui en découlent. Nous débuterons par une définition de la notion d'algorithme et son importance pour l'IA, puis nous plongerons dans les diverses méthodes d'apprentissage automatique, explorant l'univers de l'apprentissage profond et des réseaux de neurones artificiels, ainsi que du traitement automatique du langage naturel et de la vision par ordinateur.

1. Les algorithmes : fondements incontournables de l'IA

L'IA repose sur l'utilisation d'**algorithmes** pour **analyser** des données, en **tirer des enseignements**, puis **effectuer des prédictions** sur de nouveaux ensembles de données. Un algorithme est une **séquence d'instructions** logiques, structurées de manière systématique, conçues pour résoudre un problème spécifique ou effectuer des tâches déterminées. Ce nombre finis d'opérations peut être traduit, grâce à un langage de programmation, en un programme exécutable par un ordinateur. (18) Ainsi, en suivant ces **étapes spécifiques et reproductibles**, l'ordinateur traite des **données d'entrée** et produit **une sortie**.

Les algorithmes servent **de pierre angulaire** à de nombreuses techniques d'IA que nous développerons dans cette thèse, tels que l'apprentissage automatique, l'apprentissage profond, ou encore les réseaux neuronaux. Dans le vaste spectre de l'IA, divers types d'algorithmes existent, chacun présentant des caractéristiques et applications distinctes. Leur utilisation permet l'**automatisation rapide de calculs**, favorisant ainsi la **valorisation des données**, y compris dans le domaine de la santé.

2. L'apprentissage automatique

L'apprentissage automatique, également connu sous l'appellation d'apprentissage machine ou *Machine Learning* (ML) en anglais, est une branche de l'IA (Figure 5) fondée sur l'idée que «les systèmes peuvent apprendre à partir des données, identifier des modèles et prendre des décisions avec un minimum d'intervention humaine ». (19) (20)

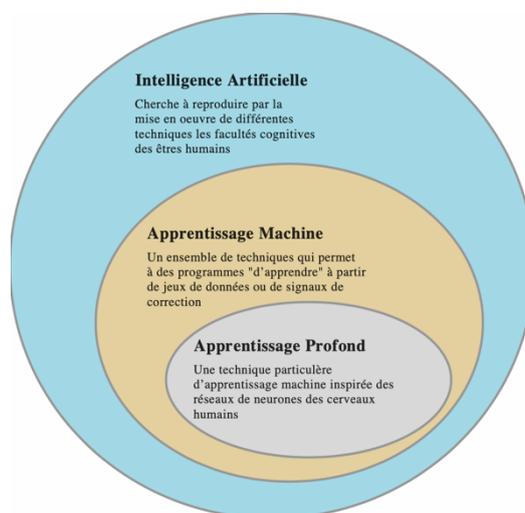


Figure 5 : Diagramme de Venn des relations entre IA, ML et apprentissage profond (21)

Le ML se distingue de la programmation classique par son approche qui consiste à **entraîner la machine** à partir de vastes ensembles de données et d'algorithmes. Contrairement à la programmation traditionnelle, qui implique que les instructions soient définies par l'humain et dictées à la machine, le ML implique que le programmeur code l'**algorithme utilisé pour entraîner le réseau**, plutôt que des règles spécifiques. Ainsi, les systèmes informatiques **évoluent de manière incrémentale, s'adaptant de manière autonome** en se basant sur les données qu'ils reçoivent. Cette approche vise à permettre aux machines d'**acquérir des connaissances à partir des données**, améliorant leur performance dans des tâches spécifiques sans nécessiter une programmation explicite pour chaque instruction. Ainsi, le ML est particulièrement efficace pour résoudre des problèmes impliquant une **grande quantité de données et plusieurs variables sans qu'un modèle ou une formule les reliant ne soient préalablement établi(e)**. (22) Cette capacité révolutionne divers secteurs, notamment l'industrie pharmaceutique, en ouvrant la voie à des percées significatives dans la recherche, la production ou encore la pharmacovigilance.

Pour illustrer cette notion complexe, prenons l'exemple fictif d'une entreprise pharmaceutique développant un traitement contre l'hypertension. Après des essais cliniques sur 1000 patients, cette entreprise a collecté des données sur l'efficacité du traitement pour des doses allant de 25 mg à 200 mg. Les variables étudiées ici sont donc la dose du médicament (donnée d'entrée) et la diminution de la pression artérielle (donnée de sortie). Le ML permettrait d'identifier une équation mathématique décrivant la relation entre ces variables, représentée par une courbe. Par extrapolation, cette équation prédirait la baisse de pression pour des doses non testées, guidant ainsi les essais cliniques et accélérant le développement du médicament.

Dans le cadre du ML, la machine suit généralement deux phases : la première est la **phase d'apprentissage**, visant à **établir un lien entre une entrée et une sortie** en utilisant des **données d'observations**. Cette phase comprend trois étapes fondamentales (*Figure 6*) :

1. A partir d'un jeu de **données d'entraînement**, les données sont **annotées** pour entraîner l'algorithme sur des caractéristiques spécifiques
2. A partir d'un jeu de **données de validation**, les performances de l'algorithme intermédiaire sont contrôlées sur des données annotées. En fonction des résultats, un **ajustement des paramètres** (variables internes à l'algorithme) et **hyperparamètres** (variables externes à l'algorithme définies manuellement par le *data scientist*) est effectué pour garantir une plus grande précision des prédictions. (23)
3. A partir des **données de test**, la performance de l'algorithme final est évaluée à l'aide de données annotées.

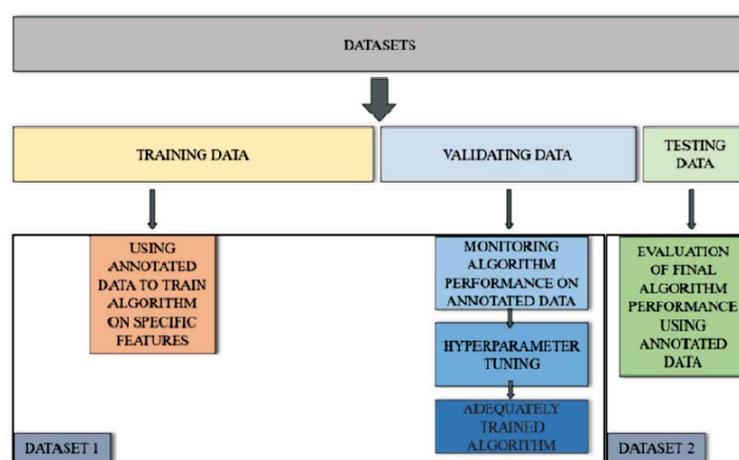


Figure 6 : Les étapes clés pour développer un algorithme de ML (23)

La seconde étape, la phase de **prédiction**, intervient une fois que le modèle est établi. Elle permet d'estimer les résultats pour de **nouvelles données** dans des contextes inconnus.

Les méthodes d'apprentissage automatique les plus courantes peuvent être classées en deux catégories principales : **l'apprentissage supervisé et non supervisé**. Nous développerons ici les méthodes les plus utilisées dans les industries de santé et notamment dans le développement des médicaments.

2.1. L'apprentissage supervisé

Dans l'apprentissage supervisé, les données d'entraînement comprennent à la fois des variables et des étiquettes, associant chaque exemple à un résultat attendu. On parle alors d'ensemble de **données étiquetées**. L'objectif principal est de permettre au modèle d'apprendre à partir de ces données étiquetées afin **de réaliser des prédictions** sur de nouvelles données. La **régression** et la **classification** sont deux méthodes supervisées d'analyse de données fréquemment utilisées en ML. (24) (22)

2.1.1. La régression

La **régression** est une méthode statistique utilisée pour **prédire une valeur continue**. Par exemple, elle pourrait permettre de prédire la concentration d'un médicament dans le sang d'un patient en fonction du temps écoulé depuis son administration. La régression peut aussi être utilisée pour **mesurer l'influence** d'une ou plusieurs variables sur une autre variable, comme évaluer l'impact de l'âge du patient sur une durée d'hospitalisation pour une pathologie particulière. (25) Les algorithmes de régression les plus utilisés dans l'industrie des produits de santé sont la **régression linéaire** et les **forêts aléatoires**.

2.1.1.1. La régression linéaire (26)

La régression linéaire est utilisée pour évaluer la relation entre une ou plusieurs variables indépendantes et une variable dépendante continue (le résultat). La régression linéaire est une technique extrêmement polyvalente qui peut être utilisée pour répondre à une variété de questions de recherche. Elle permet d'explorer les relations entre une **variable catégorielle** (comme le groupe de traitement ou le sexe du patient) et un **résultat quantitatif** (comme la pression artérielle). Le test t à deux échantillons et l'analyse de variance, souvent utilisés à cette fin, sont des cas spécifiques de régression linéaire.

Dans sa forme simple, nommée **régression linéaire simple**, une seule variable indépendante est considérée. Elle ajuste ainsi une **ligne droite** aux points de données pour caractériser au mieux la relation entre la variable dépendante (Y) et la variable indépendante (X), avec l'ordonnée à l'origine (b_0) et le coefficient de régression comme pente (b_1) de cette droite :

$$E(Y) \text{ (expected value of } Y) = b_0 + b_1 X$$

La **régression linéaire multiple** intègre quant à elle **plusieurs variables indépendantes** et peut, malgré son nom, modéliser des **relations non linéaires**. Ainsi, l'inclusion de plusieurs variables permet d'évaluer l'effet de chacune tout en maintenant constantes les autres, facilitant **l'identification des facteurs influençant le résultat**. La régression linéaire multiple permet alors de quantifier **la direction** et **la force** des relations.

2.1.1.2. Les forêts aléatoires

Les **forêts aléatoires** ou **Random Forest (RF)** en anglais sont une méthode d'apprentissage basée sur un **mécanisme de vote majoritaire** dans laquelle **plusieurs arbres de décisions** sont construits à partir des données d'apprentissage. Les RF peuvent être utilisées pour faire de la **régression** mais aussi de la **classification**.

Les arbres de décision (*Figure 7*), largement utilisés dans les domaines des statistiques et de l'ingénierie, doivent leur popularité à leur simplicité et à leur polyvalence.

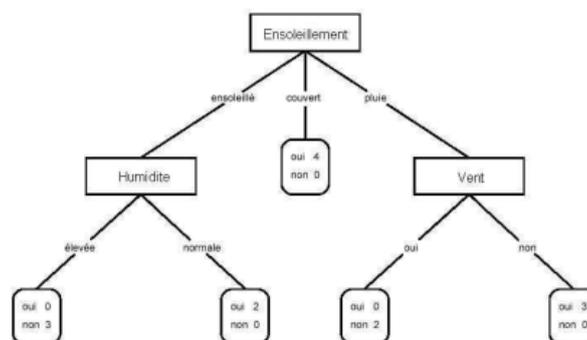


Figure 7 : Exemple d'un arbre de décision à partir de l'exemple de Quinlan (27) pour prédire si «oui ou non » le temps permettra de jouer à l'extérieur (28)

Les RF appartiennent aux **méthodes d'apprentissage d'ensemble**. Dans ce contexte, le terme "ensemble" évoque un regroupement : une forêt aléatoire standard implique la création d'un grand nombre d'arbres de décision lors de la phase d'apprentissage. Les RF sont des méthodes d'apprentissage **non linéaires et non paramétriques**. Elles peuvent donc traiter efficacement des ensembles de données contenant un grand nombre de variables (on parle d'ensemble de données à **haute dimension**), qu'elles soient **numériques ou catégorielles**, sans nécessiter de **conditions** spécifiques sur les données. (29)

Les arbres de décision sont créés en utilisant des ensembles de données générés par **échantillonnage aléatoire** de l'ensemble d'apprentissage (appelés *bootstraps*). Chaque arbre est construit à partir d'un sous-ensemble différent des données d'apprentissage. Ensuite, chaque arbre est évalué sur un ensemble de données de test distinct, excluant certains enregistrements de l'ensemble d'apprentissage (appelés "échantillons hors sac"). L'erreur de classification de cet ensemble de test est utilisée pour estimer l'erreur de généralisation du modèle. Après l'entraînement, chaque arbre de décision dans une RF est capable de cartographier les données d'entrée. Pour prédire de nouvelles données, chaque arbre vote pour une classe, et la **classe la plus fréquemment votée** est choisie comme prédiction finale pour la classification. (30)

2.1.2. La classification

La **classification** est une technique utilisée pour **attribuer des catégories prédéfinies** (on parle d'étiquettes) à des données en fonction de leurs caractéristiques. La classification est donc utilisée lorsque la sortie souhaitée est une **catégorie discrète ou une classe**. Un exemple d'application est le classement des images radiologiques en différentes pathologies, permettant ainsi une amélioration du diagnostic médical. La méthode des **k plus proches voisins**, les **machines à vecteur de support** et la **classification naïve bayésienne** comptent parmi les algorithmes de classification les plus utilisés dans l'industrie des produits de santé.

2.1.2.1. La méthode des k plus proches voisins (22,24,31)

La méthode des k plus proches voisins ou *k-Nearest Neighbors* en anglais (**kNN**) classe les données non étiquetées en les affectant à la classe des exemples les plus similaires. En effet, dans l'algorithme des kNN, les données sont visualisées dans un espace à plusieurs dimensions où chaque point représente une entité avec des caractéristiques spécifiques. Ici, le paramètre k fait référence au **nombre de voisins les plus proches** participant au processus de **vote majoritaire**. Ainsi, pour classer de nouvelles données, l'algorithme mesure la proximité entre ces points dans cet espace et utilise les k voisins les plus proches pour déterminer la classe de la donnée en question.

Prenons un exemple concret pour illustrer cette notion. Les fruits, les légumes et les céréales se distinguent par deux paramètres que sont leur texture (plus ou moins croquants) et leur saveur (plus ou moins sucrés) (*Figure 8*).

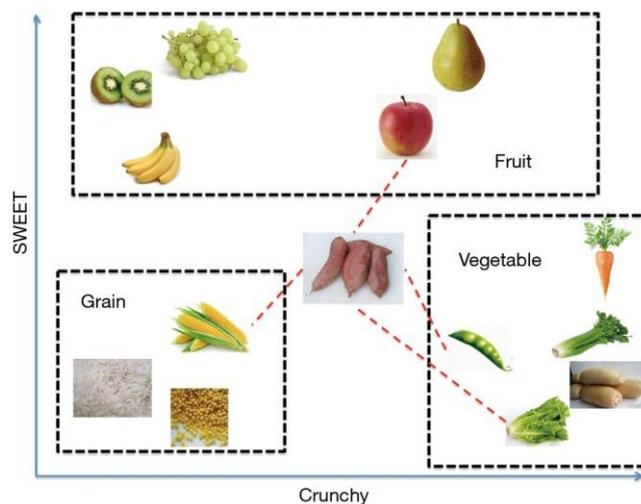


Figure 8 : Illustration du fonctionnement de l'algorithme des k-voisins les plus proches (31)

Pour les représenter graphiquement en deux dimensions, seules deux caractéristiques sont utilisées ici, mais le modèle peut être étendu pour inclure n'importe quel nombre de caractéristiques. Notre objectif est de déterminer la catégorie à laquelle appartient la patate douce (fruit, légume ou céréales). Dans cet exemple, $k=4$: nous sélectionnons les quatre types d'aliments les plus proches : la pomme, le haricot vert, la laitue et le maïs. Puisque le légume obtient le plus grand nombre de votes, la patate douce est classée dans la catégorie des légumes.

Comme le démontre l'exemple précédent, deux concepts sont très importants dans le KNN :

- **La méthode de calcul de la distance** entre l'élément à déterminer et les autres éléments. Bien qu'il existe plusieurs méthodes pour calculer cette distance, la manière la plus fréquente de la calculer est la suivante :

$$D(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

Dans cette formule, p et q sont les éléments à comparer et n le nombre de caractéristiques pris en considération.

- **Le paramètre k qui détermine le nombre de voisins** : ce paramètre a un impact significatif sur les performances de l'algorithme. Une valeur de k trop élevée ou trop faible peut conduire à des taux de faux positifs ou de faux négatifs indésirables. Il est donc nécessaire de choisir une valeur k appropriée pour éviter le surajustement et le sous-ajustement (32). Certains auteurs suggèrent de définir k comme étant égal à la racine carrée du nombre d'observations dans le jeu de données d'apprentissage (33).

2.1.2.2. Les machines à vecteur de support

Les machines à vecteur de support dites **Support Vector Machine (SVM)** en anglais sont une méthode de ML supervisé qui effectue des tâches de **classification**. Elles construisent des **lignes de séparation** dans un espace multidimensionnel pour distinguer les objets appartenant à des **classes différentes**. (22)

Les SVM projettent les données dans un hyperplan où elles sont séparées en classes. L'objectif est de créer un **plan de séparation (dit hyperplan séparateur)** qui maximise la distance entre lui-même et les échantillons les plus proches de chaque classe, situés sur les limites de la **marge de séparation** (*Figure 9*). (34)

Ainsi, l'équation de cette droite séparatrice optimale, $\mathbf{w}\mathbf{x}^T + \mathbf{b} = 0$ constitue la base du modèle SVM linéaire. Dans cette équation :

- w est un vecteur qui définit la direction de la séparation entre les catégories
- x représente les données d'entrée
- b est un terme qui permet de déplacer l'hyperplan de séparation (le biais)

Il s'agit donc de trouver la meilleure combinaison de w et b pour diviser les données de manière optimale entre les deux catégories tout en minimisant les erreurs de classification et ainsi assurer une classification précise des nouvelles données en fonction de leur position par rapport à l'hyperplan de séparation.

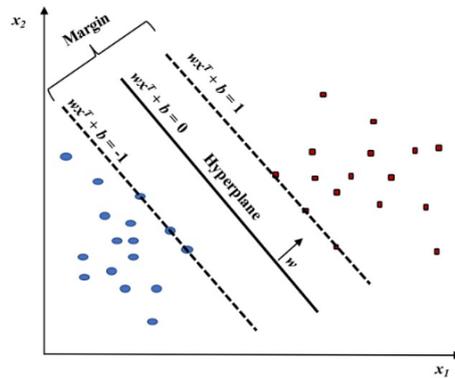


Figure 9 : Exemple de modèle SVM linéaire (35)

Deux classes (rouge et bleue) ont été séparées. La droite continue représente l'hyperplan séparateur tandis que les lignes discontinues représentent les hyperplans de soutien, indiquant la marge de séparation entre deux classes.

Lorsque les données ne sont pas linéairement séparables, comme c'est le cas dans la plupart des applications du monde réel, les SVM utilisent des noyaux non linéaires, également appelés **fonctions noyau** ou **Kernel** en anglais (Figure 10). Ces fonctions mathématiques permettent de projeter et de séparer les données dans un espace vectoriel de dimensions plus élevées, en ajoutant ainsi des dimensions supplémentaires aux données brutes pour rendre la séparation réalisable et permettre la classification (36). Cette méthode permet une classification précise même pour des données non incluses dans l'ensemble d'apprentissage, assurant des performances de généralisation élevées.

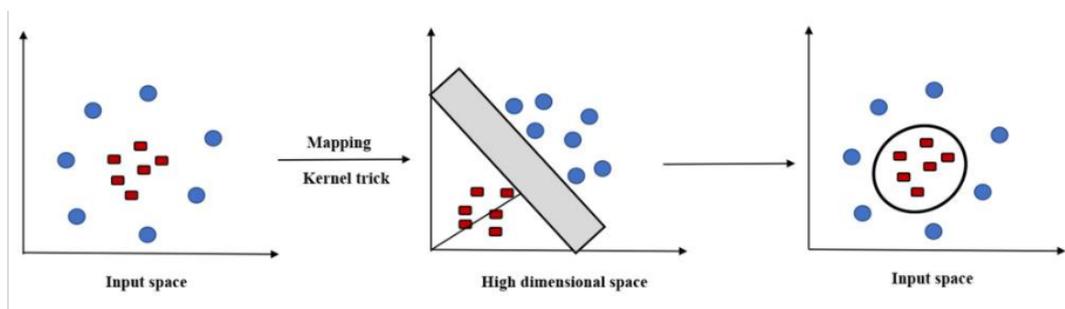


Figure 10 : Illustration d'une fonction noyau (35)

Les données qui ne peuvent être séparées par un SVM linéaire sont transformées et séparées par une fonction noyau dans un espace vectoriel de dimensions plus élevées.

Initialement conçus pour la classification de variables qualitatives binaires, les SVM ont été adaptées pour la prédiction de valeurs numériques, c'est-à-dire de **régression**. On parle alors de Support Vector Regression (SVR). Rapidement adoptée en raison de sa capacité à traiter des données de grandes dimensions ainsi que sa simplicité d'usage, la SVM/SVR est devenue l'une des approches d'apprentissage automatique les plus populaires en découverte de médicaments.

2.1.2.3. La classification naïve bayésienne (37)

La méthode de classification naïve bayésienne est la troisième méthode de classification très utilisée dans la recherche médicale. Basée sur le théorème de Bayes et les probabilités conditionnelles, elle est simple mais peut être extrêmement efficace. Le terme « naïf » fait référence à l'hypothèse selon laquelle toutes les caractéristiques des données sont **indépendantes** les unes des autres. Bien que cette hypothèse soit irréaliste (toutes les variables n'ont pas la même importance), Domingos et Pazzani (38) ont démontré que la méthode bayésienne peut rester optimale même en présence de violations de cette indépendance car cette hypothèse d'indépendance n'est qu'un aspect de l'estimation de la fonction de classification.

Les classificateurs bayésiens attribuent **la classe la plus probable** à chaque échantillon, en se basant sur les valeurs vectorielles de ses variables. L'apprentissage probabiliste de Bayes consiste à évaluer la probabilité qu'un échantillon appartienne à une certaine classe en se basant sur la distribution de probabilité a priori, distribution correspondant aux proportions relatives des différentes classes dans les données d'apprentissage. (39)

Ce modèle simple et rapide est capable de travailler avec **des données bruitées** et d'apprendre à partir de **petits ensembles de données**, ce qui est un avantage. En revanche, ce n'est pas l'algorithme idéal pour les problèmes impliquant de nombreuses caractéristiques car il utilise des tables de fréquence pour extraire des connaissances des données, ce qui le rend moins efficace dans les cas de grande dimensionnalité. En outre, la classification naïve bayésienne considère chaque caractéristique comme si elle était catégorique, ce qui peut poser problème lorsqu'on traite avec des données numériques. L'algorithme doit alors les convertir de manière qui peut ne pas être optimale. (24)

2.1.3. L'apprentissage semi-supervisé

L'apprentissage semi-supervisé est une variante de l'apprentissage supervisé intégrant une **petite quantité de données étiquetées** avec un **grand nombre de données non étiquetées** dans le processus d'apprentissage. Cette approche vise à **améliorer la précision** des modèles d'apprentissage, surtout lorsque les ensembles de données sont **petits ou déséquilibrés**. En combinant les informations provenant à la fois des données étiquetées et non étiquetées, l'apprentissage semi-supervisé permet aux modèles d'apprentissage automatique **de généraliser plus efficacement** et de **mieux capturer la complexité** des données réelles. (40)

2.2. L'apprentissage non supervisé

L'apprentissage non supervisé explore et met en évidence les modèles et les relations sous-jacents présents dans des données de **manière autonome**. Contrairement aux techniques supervisées, les algorithmes d'apprentissage non supervisé sont entraînés à partir de **données non étiquetées**, ce qui signifie qu'ils n'ont pas de réponses attendues associées aux entrées.

2.2.1. Le clustering

Le clustering est une technique non supervisée largement utilisée visant à **regrouper** des données similaires en clusters ou en groupes, **sans recourir à des étiquettes préalables**. L'objectif ici est de découvrir des structures intrinsèques dans les données et de les regrouper en fonction de leurs similarités. Dans le domaine médical, par exemple, le clustering peut être appliqué pour regrouper des patients présentant des caractéristiques médicales similaires, permettant l'identification de groupes de patients homogènes. Nous allons ici décrire la méthode de clustering la plus courante dans la littérature : la méthode des k-moyennes.

2.2.1.1. La méthode des k-moyennes (41) (42) (43)

Le principal objectif de l'**algorithme des k-moyennes** ou *k-means* en anglais est de regrouper un ensemble de données en un nombre prédéfini de **clusters** (k), chaque cluster étant représenté par un centre appelé **centroïde** (*Figure 11*).

Le processus de l'algorithme débute par l'assignation aléatoire de **centroïdes** dans l'espace des données. Ensuite, chaque point de données est assigné au cluster dont le centroïde est le plus proche. Les centroïdes sont ensuite mis à jour en calculant les **moyennes** des points de données au sein de chaque cluster. Les étapes se répètent jusqu'à ce que les centroïdes convergent vers des positions où les changements deviennent négligeables. L'objectif est de minimiser la somme des carrés des distances entre chaque point de données et son centroïde associé, une mesure appelée **inertie intra-cluster**. Cette inertie intra-cluster assure une certaine **cohésion** au sein de chaque groupe (44).

Bien que très largement utilisé en pratique, l'algorithme des k-moyennes présente quelques limites. En effet, le **choix du nombre optimal de clusters (k)** peut influencer significativement les résultats du clustering et implique la nécessité d'avoir une idée de ce k en amont.

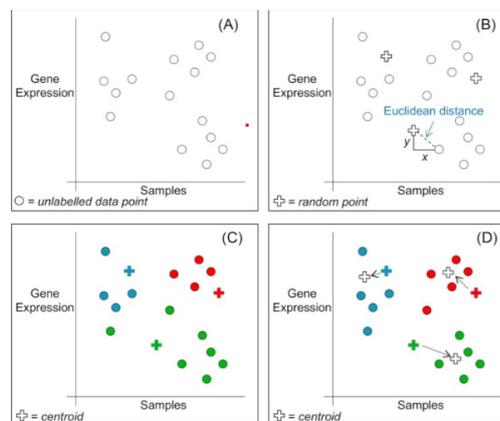


Figure 11 : Exemple de représentation graphique de la méthode des k-moyennes (43)

(A) Visualisation de l'expression génique à partir d'échantillons microbiens simulés comprenant 15 points de données non étiquetés. L'objectif est de regrouper ces échantillons pour une analyse ultérieure. La première étape du k-moyenne consiste à déterminer le nombre k de clusters souhaités. (B) Trois points ont été choisis aléatoirement comme centroïdes initiaux pour les trois clusters prévus ($k = 3$). Une mesure de distance est calculée entre les centroïdes et chaque point de données. (C) Chaque point est assigné au cluster dont le centroïde est le plus proche (D) Les nouveaux centroïdes sont recalculés en prenant la moyenne des points de données de chaque cluster, puis chaque point est réassigné au centroïde le plus proche. Ce processus est répété jusqu'à ce que les centroïdes ne bougent plus, indiquant la convergence du clustering.

2.3. Sélection du modèle adapté et mesure de la performance

La sélection du modèle adapté et l'évaluation de sa performance sont des étapes cruciales dans le développement de modèles de ML. Pour choisir le modèle le plus approprié pour un problème spécifique, il est essentiel de comprendre la **nature des données**, la **complexité du problème** et les **objectifs** de l'application. Les chercheurs doivent explorer différents algorithmes, architectures et paramètres pour trouver celui qui offre le meilleur compromis entre précision, généralisation et complexité du modèle.

Une fois le modèle sélectionné, il est essentiel **d'évaluer sa performance** à l'aide de mesures appropriées telles que la **précision**, le **taux d'erreur** ou **l'aire sous la courbe ROC**, en fonction de la nature du problème. La validation et l'interprétation des résultats garantissent que le modèle **généralise efficacement** à partir des données d'entraînement.

3. Les réseaux neuronaux et l'apprentissage profond (24) (45)

L'apprentissage profond ou *Deep Learning* en anglais (DL), est une sous-catégorie de l'apprentissage automatique (*Figure 5*) qui s'inspire du cerveau humain pour créer des modèles capables d'apprendre à partir de données. Cette technologie repose sur l'utilisation de **réseaux de neurones profonds**, structures composées de **plusieurs couches** de neurones artificiels interconnectés : une **couche d'entrée**, qui reçoit directement les informations, plusieurs **couches cachées** interconnectées et une **couche de sortie**, qui transmet les informations directement à l'extérieur.

Les réseaux neuronaux apprennent automatiquement à partir de vastes ensembles de données, évitant ainsi le besoin de programmation explicite. Cette caractéristique leur permet de discerner des motifs et des relations complexes dans les données. Grâce à leurs architectures complexes, parfois constituées de millions de paramètres, les réseaux neuronaux profonds peuvent modéliser des relations entre les entrées et les sorties, même en l'absence de modèle préétabli. Ainsi, en utilisant **des exemples** pour générer des modèles, les réseaux neuronaux profonds sont des outils polyvalents pour la prédiction, pouvant être utilisés pour la détection, la régression et la classification. Ils fournissent une approche efficace pour résoudre une variété de problèmes complexes du monde réel. (34)

Le fonctionnement du DL repose sur le concept de **propagation avant** (*forward propagation*) et de **rétropropagation** (*backpropagation*). La propagation avant **introduit les données** dans le réseau, tandis que la rétropropagation **ajuste les poids** des connexions entre les neurones pour minimiser l'écart entre les prédictions et les valeurs cibles, améliorant ainsi la **performance** du modèle au fil du temps. Dans le domaine de la santé, le DL peut avoir des applications diverses allant de **l'interprétation d'images médicales**, à la **prédiction de l'efficacité d'un composé chimique**, passant par **l'identification de nouvelles cibles moléculaires**.

3.1. Les perceptrons multicouches

Un **perceptron multicouche** ou *Multi Layer Perceptron* en anglais (MLP) est un type de réseau de neurones artificiels capable de modéliser des relations complexes et non linéaires dans les données. Contrairement au perceptron simple, qui comporte une seule couche de neurones, le MLP intègre au moins une couche cachée entre la couche d'entrée et la couche de sortie. (Figure 12)

Les MLP sont des réseaux neuronaux relativement simples dits à **propagation directe**, où les informations circulent dans **une seule direction**, de l'entrée à la sortie. Ainsi, chaque neurone reçoit des données provenant des neurones de la couche précédente (ou des données d'entrée), les combine puis applique une **fonction d'activation** pour produire une sortie. Ces entrées sont associées à des poids ajustant l'importance des connexions. Pour améliorer les prédictions, le MLP utilise la rétropropagation, qui corrige les erreurs en ajustant les poids des connexions.

Les neurones de la couche cachée introduisent de la **non-linéarité** dans le modèle en utilisant des fonctions d'activation telles que la sigmoïde, la tangente hyperbolique (tanh) ou la fonction ReLU. C'est cette non-linéarité qui permet aux MLP de capturer des relations **complexes** entre les entrées et les sorties. Les neurones de la dernière couche cachée transmettent les résultats à la couche de sortie, qui calcule la sortie finale. Le défi principal de la formation des MLP réside dans le **choix de l'architecture du réseau**, notamment le nombre de couches cachées et le nombre de neurones dans chaque couche. (46) (34)

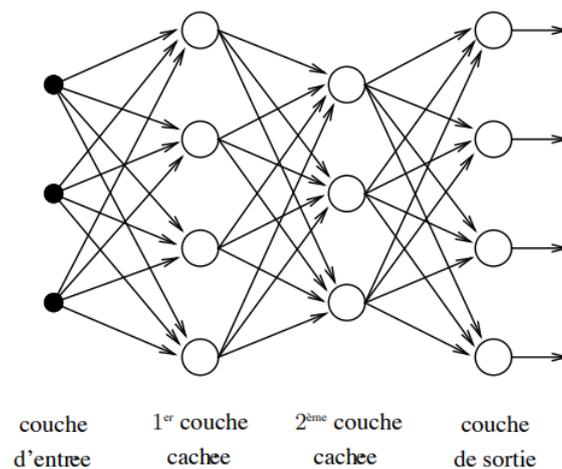


Figure 12 : Exemple d'un réseau de type MLP (47)

3.2. Les réseaux de neurones convolutifs profonds (24)

Les **réseaux de neurones convolutifs profonds** ou *Convolutional Neural Network (CNN)* sont une sous-catégorie de réseaux neuronaux d'apprentissage profond conçus spécifiquement pour le traitement des **données structurées en grille**, telles que les **images**. Ils se distinguent par leur capacité à extraire automatiquement des caractéristiques pertinentes des images.

La clé des CNN réside dans leur architecture en **couches spécialisées** (Figure 13). Les **couches de convolution** agissent comme des **filtres** pour détecter des motifs visuels tels que des contours ou des textures. Ces filtres sont ajustés pendant l'entraînement pour améliorer la reconnaissance des motifs pertinents. Les **couches de pooling** réduisent la dimension spatiale des données, ce qui accélère le processus d'apprentissage tout en préservant les caractéristiques importantes.

Une fois que les caractéristiques ont été extraites par les couches de convolution et de pooling, elles sont agrégées par des **couches entièrement connectées**, similaires aux perceptrons multicouches, pour effectuer la classification ou la prédiction finale.

Grâce à cette structure hiérarchique, les CNN sont en mesure de capturer avec précision des informations complexes, leur conférant une grande efficacité dans une variété de tâches de **vision par ordinateur**, notamment la reconnaissance et la classification d'images, ainsi que la détection d'objets.

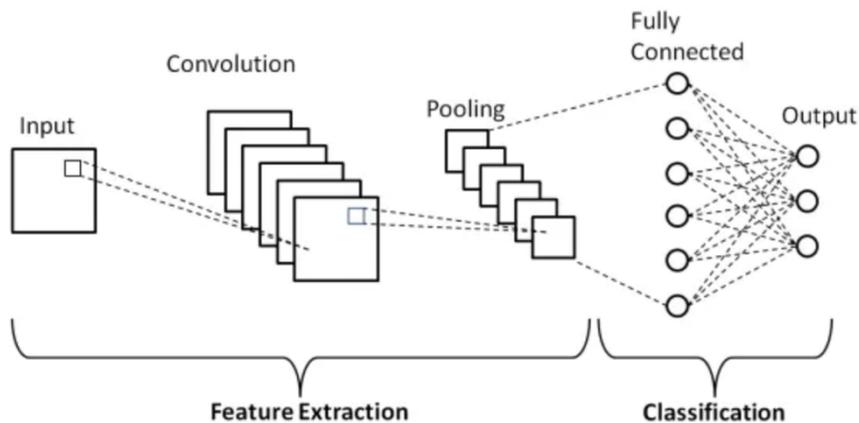


Figure 13 : Architecture d'un réseau de type CNN (47)

Les **réseaux convolutifs de graphes** ou *Graph Convolutional Network* en anglais (*GCN*) sont une variante des CNN adaptée au traitement des données structurées sous forme de **graphes**, comme les graphes moléculaires.

3.3. Les réseaux de neurones récurrents (48)

Les **réseaux de neurones récurrents** ou *Recurrent Neural Network* en anglais (*RNN*) constituent une classe de réseaux neuronaux conçus pour analyser des **données séquentielles**. Contrairement aux réseaux neuronaux traditionnels, les RNN peuvent conserver une **mémoire interne**, leur permettant ainsi de comprendre le contexte antérieur lors du traitement de séquences de données.

Leur architecture (*Figure 14*) se compose de modules de neurones répétitifs disposés en **une séquence**, permettant à chaque neurone de tenir compte des informations des étapes précédentes. Cette capacité à capturer les dépendances séquentielles en fait un choix privilégié pour le **traitement du langage naturel**, avec notamment des tâches telles que la **traduction automatique** ou la **génération de texte** (49). Les RNN incluent des variantes telles que les Long Short-Term Memory (LSTM) et les Gated Recurrent Unit (GRU), qui surmontent les limitations de la mémoire à court terme des RNN classiques pour modéliser des séquences plus longues et complexes.

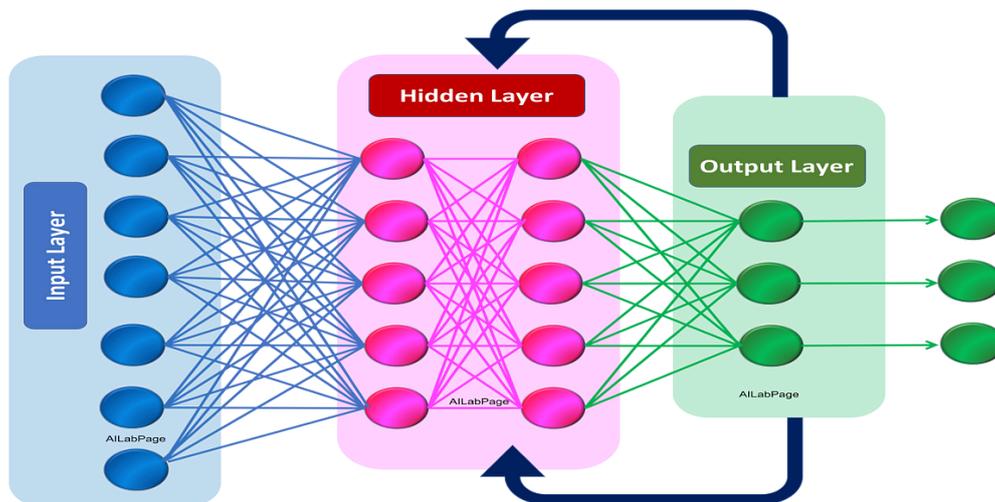


Figure 14 : Architecture d'un RNN (48)

4. La vision par ordinateur

La vision par ordinateur est un domaine de l'IA qui se concentre sur la compréhension et l'interprétation des informations visuelles extraites d'**images** ou de **vidéos**. Parmi ses principales applications, on trouve :

- **La reconnaissance d'images**, qui vise à identifier des objets, des scènes ou des motifs spécifiques dans une image, permettant ainsi à un système informatique de comprendre et d'interpréter son contenu visuel
- **La classification d'images**, qui consiste à attribuer une étiquette ou une catégorie prédéfinie à une image en fonction de son contenu, facilitant ainsi l'organisation et le tri des images en fonction de leurs caractéristiques communes
- **La détection d'objets**, qui vise à localiser et à identifier des objets spécifiques dans une image en les encadrant.

Ces techniques sont largement utilisées dans de nombreux domaines, tels que la surveillance vidéo, la reconnaissance faciale, la médecine et l'automobile autonome, contribuant ainsi à automatiser des tâches auparavant réservées aux humains et à améliorer l'efficacité des systèmes informatiques dans le traitement et l'analyse d'informations visuelles.

5. Le traitement automatique du langage naturel

Le **TALN** ou **NLP** est un domaine clé de l'IA qui se concentre sur la compréhension et la manipulation du langage humain par les ordinateurs. Il englobe un ensemble de techniques et d'outils visant à analyser, interpréter et générer du texte en langage naturel.

Ses applications sont vastes, allant de la **traduction automatique** à la **génération de texte**, en passant par la **classification de texte**, l'**extraction d'informations** et le **résumé automatique** de documents. Les systèmes de NLP utilisent des algorithmes de ML pour traiter et comprendre le langage humain dans ses différentes dimensions, y compris la syntaxe et la sémantique. Grâce aux progrès récents dans le domaine du DL et du traitement des données massives, le NLP a connu des avancées significatives, permettant le développement de systèmes de traduction automatique plus précis, de chatbots conversationnels plus sophistiqués et d'outils d'analyse de texte plus performants.

Les industries de santé bénéficient largement des applications du NLP, notamment avec **l'extraction automatique d'informations** à partir de **documents cliniques**, **l'analyse de la littérature médicale** et la **surveillance des réactions indésirables** aux médicaments.

IV. Les données, source indispensable pour tout projet d'IA

1. La révolution du 'Big Data'

Tous les domaines et secteurs d'activité sont désormais concernés par ce qu'on appelle la « **mise en données du monde** » (ou *datafication* en anglais). Dans leur ouvrage intitulé *Big Data : A Revolution that Will Transform how We Live, Work, and Think*, Viktor Mayer-Schönberger et Kenneth Cukier décrivent la mise en donnée d'un phénomène comme sa transformation en « **données quantifiées, tabulées et analysables** ».

Les nouvelles Technologies de l'Information et de la Communication (TIC) ont joué un rôle essentiel en permettant de collecter, de stocker et d'analyser des données à des échelles sans précédent. L'adoption généralisée d'**internet** à travers le monde et son intégration dans de nombreux produits du quotidien (**les objets connectés**) ont notamment créé de nombreuses sources de données. Entre 2010 et 2020, le volume de données numériques a augmenté de **3100%**, passant de deux à soixante-quatre zettaoctets. (50) Cette croissance exponentielle du volume et de la qualité des données collectées, couplée à l'augmentation des capacités de stockage, a catalysé l'essor récent de l'IA.

2. Définition des mégadonnées

Les **mégadonnées**, également appelées « Big Data » ou « données massives » désignent, d'après le Parlement Européen « *des ensembles de données collectées si **gigantesques et complexes** qu'ils nécessitent de nouvelles technologies, telles que **l'intelligence artificielle**, pour être traités.* » (51) La frontière entre données massives et données traditionnelles réside donc dans les **ressources nécessaires au traitement** des données.

3. Les caractéristiques des mégadonnées

Les mégadonnées sont décrites par les « **5V** », correspondant à cinq caractéristiques essentielles (52) : Volume, Variété, Vitesse, Vérité, Valeur.

Le volume fait référence à la quantité de données, qui, comme évoqué précédemment, est trop volumineuse pour être traitée par des outils traditionnels.

La variété met en lumière la diversité des formats (image, texte, audio...) et des types de données, incluant **des données structurées, semi-structurées ou non structurées** (brutes). Les données structurées sont formatées selon une structure spécifique, tandis que les données non structurées, qui représentent **plus de 80%** des informations collectées, peuvent prendre différentes formes, ce qui les rend plus complexes à exploiter et sont donc un enjeu majeur pour les entreprises.

La vitesse des données renvoie à la **vitesse de génération, de collecte et d'analyse**. Ces données sont souvent générées à un rythme rapide, parfois sous forme de **flux continus**, nécessitant un traitement en temps réel.

La vérité souligne l'importance de la **fiabilité** des données, nécessitant une vérification minutieuse de la **qualité** du contenu et de la **crédibilité de la source** avant l'exploitation des données. Résoudre les problèmes de valeurs aberrantes ou manquantes est crucial pour renforcer la confiance dans l'utilisation des données

Enfin, **la valeur** des mégadonnées réside dans leur capacité à apporter une réelle valeur ajoutée dans ce contexte de profusion de données. En effet, les technologies de stockage et d'analyse des mégadonnées n'ont de sens que si elles permettent de mener des **analyses pertinentes** et contribuent à la prise de décision en fournissant des informations précieuses.

4. Les sources des mégadonnées

4.1. Des données générées par les internautes

Dans la routine quotidienne des individus, la génération de données est omniprésente, que ce soit à travers les échanges de courriels entre collègues, les appels via des plateformes telles que Skype ou Teams, ou encore les données de localisation capturées par les téléphones lors de l'utilisation du GPS. Cette production massive de données se déroule souvent de manière inconsciente. Pour illustrer l'ampleur de ce phénomène, il est intéressant de noter qu'en 2020, une estimation a révélé que chaque minute, 400 000 heures de contenu vidéo étaient visionnées sur Netflix, et que 42 millions de messages étaient échangés via WhatsApp à travers le monde. (53)

4.2. Des données générées par les objets

Les **objets connectés**, aussi appelés **Internet des Objets** ou *Internet of Things (IoT)* en anglais, se définissent comme « *Des objets ayant la particularité d'être, comme leur nom l'indique, connectés entre eux via Wi-Fi, Bluetooth ou tout autre protocole de communication leur permettant d'échanger des informations* ». (54) Ces objets font désormais partie intégrante de notre quotidien, avec des exemples comme les montres ou encore les pèse-personnes connectés. Selon l'Agence Nationale des Fréquences (ANFR), le déploiement actuel de la 5G permettra de **multiplier par 10 le nombre d'objets connectés** pouvant simultanément accéder au réseau. (55)

5. IA et données : une collaboration indispensable pour l'industrie des produits de santé

L'IA est inefficace sans des données d'entrée de qualité. En effet, la collecte de données est cruciale pour les applications de l'IA dans les industries de santé puisque ces données servent de **matière première** aux processus de **valorisation de l'information** et permettent ainsi de **générer de nouvelles connaissances**.

Les sources de données utiles pour les industries de santé sont variées et englobent notamment les **dossiers médicaux électroniques (DME)**, les **capteurs médicaux**, les **articles scientifiques**, les **dispositifs de surveillance médicale**, les **données génétiques**, les données sur les **structures chimiques**, les données **in vitro**, et bien d'autres.

L'essor des **dispositifs médicaux (DM) connectés** génère un flux continu de données en temps réel, nécessitant une harmonisation avec d'autres sources pour une prise de décision clinique cohérente. Les protocoles de collecte de données sont cruciaux pour garantir la qualité des informations recueillies, en identifiant et corrigeant les anomalies. Les normes et directives **standardisent la collecte, le stockage et le partage des données**, tandis que les techniques d'analyse avancées et de l'IA contribuent à **améliorer leur qualité** en détectant et corrigeant les erreurs.

En somme, la **collecte des données** est la pierre angulaire des projets d'IA dans les industries de santé. Les progrès dans les technologies de la science des données ont ouvert de nouvelles perspectives pour l'application de l'IA dans la découverte et le développement de médicaments. Parallèlement, l'IA a connu des avancées considérables ces dernières années, permettant l'utilisation de vastes ensembles de données à une échelle auparavant inatteignable. Cette synergie entre l'IA et les mégadonnées crée un cercle vertueux où l'ampleur et la diversité des données alimentent les algorithmes d'IA, qui à leur tour, grâce à leurs capacités d'analyse avancées, permettent une exploitation plus efficace et précise des mégadonnées. Dans la section suivante, nous développerons les applications de l'IA et notamment du ML dans l'industrie des produits de santé, de la recherche à la commercialisation.

PARTIE 2 : Les applications de l'IA dans l'industrie des produits de santé

I. Les applications de l'IA dans le cycle du médicament

Le développement d'un médicament débute par une phase initiale de recherche lors de laquelle les scientifiques identifient la pathologie à traiter et les **cibles** thérapeutiques potentielles. Des **molécules candidates** sont alors soigneusement sélectionnées et soumises à des **évaluations précliniques** visant à déterminer leur efficacité et leur innocuité avant toute administration à l'Homme. Ensuite, les **essais cliniques** approfondissent les profils d'**efficacité** et de **sécurité** du médicament chez l'être humain, et les résultats obtenus forment le socle du **dossier d'AMM** (Autorisation de Mise sur le Marché). L'octroi de cette autorisation marque le début de la **production industrielle** et de la **commercialisation** du médicament.

Ce processus, qui exige une approche rigoureuse et une réglementation stricte est traditionnellement caractérisé par des **coûts élevés**, des **délais prolongés** et des **taux de réussite relativement faibles** (Figure 15). En effet, le taux de succès des essais cliniques est inférieur à **15%** et seulement une molécule est approuvée par les autorités sur 200 testées lors des études précliniques (56) (57)

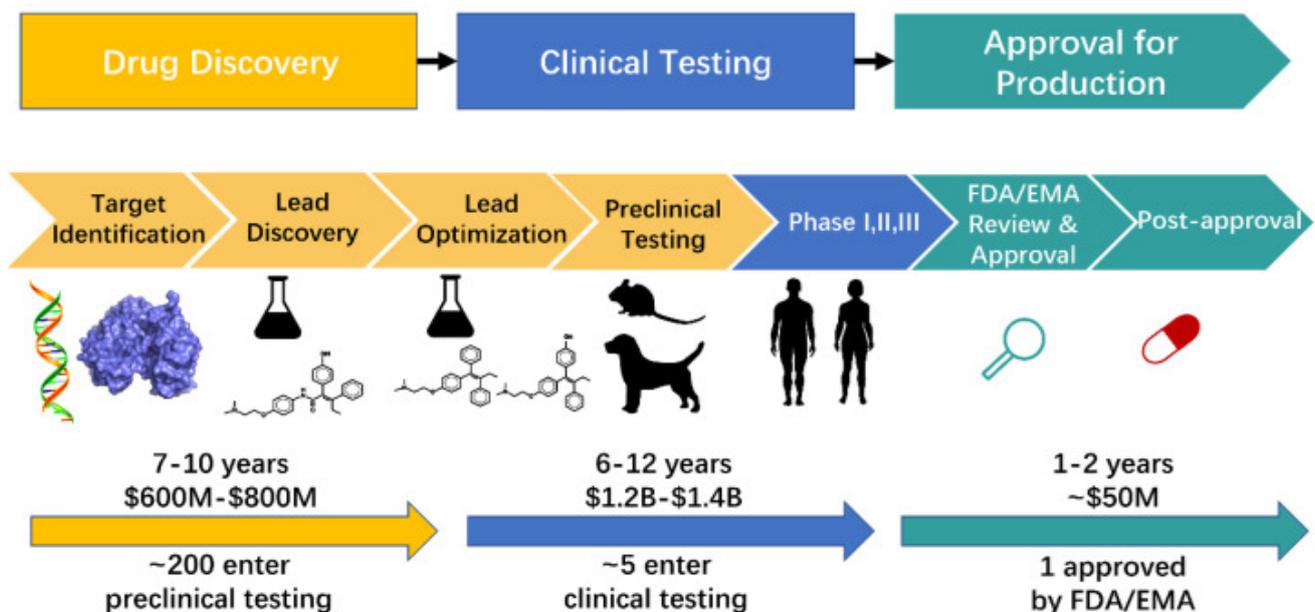


Figure 15 : Le processus de recherche et développement (R&D) de médicaments (56)

De plus, le secteur pharmaceutique connaît actuellement une période de profonds changements en matière d'**innovation**. Outre les préoccupations majeures liées aux coûts et aux délais de développement des médicaments, d'autres défis se présentent. Les médicaments phares, également connus sous le nom de **blockbusters**, caractérisés par un chiffre d'affaires annuel dépassant souvent le milliard de dollars, voient leur part de marché décliner à mesure que leurs **brevets expirent** et que des **médicaments génériques** font leur entrée sur le marché. Cette concurrence entraîne une baisse significative des prix (les génériqueurs n'ayant pas à supporter les frais de recherche), et par conséquent une diminution des revenus des laboratoires.

Dans cette conjoncture, l'IA émerge comme une solution prometteuse à ces nouveaux défis, offrant des avantages significatifs à chaque étape du cycle du médicament (*Figure 16*). Elle transforme fondamentalement les approches traditionnelles de découverte de nouvelles molécules, de mise en place d'essais cliniques, d'optimisation des processus de production et de logistique, ainsi que de surveillance post-commercialisation de la sécurité des médicaments.

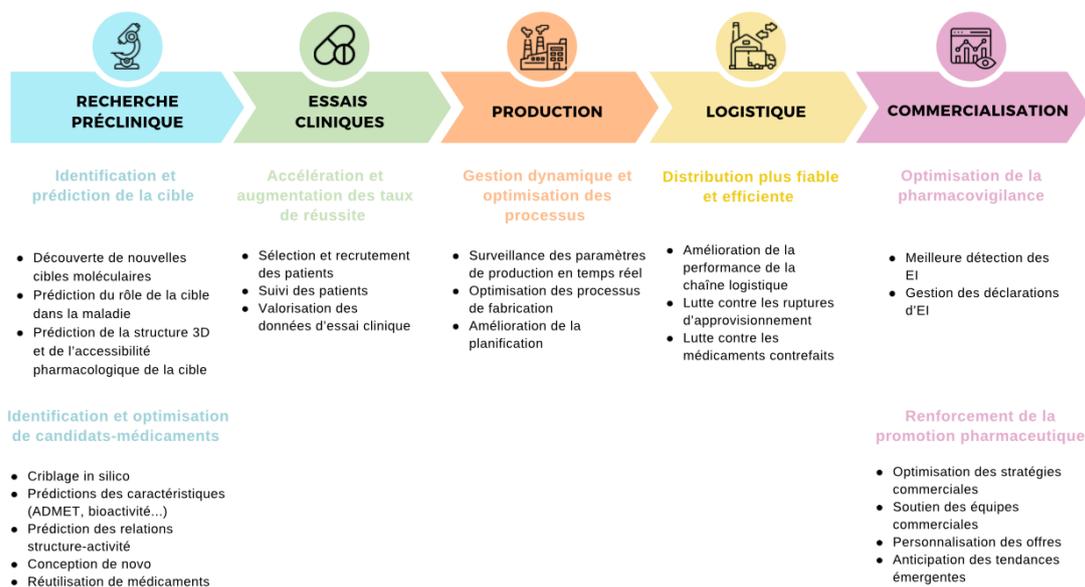


Figure 16 : Contributions de l'IA dans le cycle du médicament

Grâce à l'**analyse avancée** des données et l'**automatisation** de nombreuses tâches qu'elle permet, l'IA promet non seulement des économies substantielles en termes de temps et d'argent, mais également des améliorations significatives des résultats du développement des médicaments. Dans cette section, nous examinerons attentivement chaque étape du cycle du médicament, de l'identification de la cible thérapeutique à la commercialisation, en soulignant les avantages tangibles que les différentes techniques d'IA peuvent apporter.

1. IA et recherche préclinique

1.1. Les différentes approches de la découverte d'un médicament

Dans la découverte de médicaments, deux approches principales existent : l'approche classique et l'approche dite inverse.

L'approche classique, souvent appelée **pharmacologie traditionnelle** ou **avancée**, repose sur des méthodes **expérimentales** et **observationnelles** pour identifier des composés chimiques ayant des effets bénéfiques sur une maladie spécifique. Dans cette approche empirique, les scientifiques testent de **nombreux composés chimiques** pour évaluer leurs effets sur la maladie ciblée, sans nécessairement comprendre complètement les mécanismes sous-jacents.

À l'opposé, la **pharmacologie inverse**, également connue sous le nom de **conception rationnelle de médicaments**, repose sur une compréhension approfondie des processus biologiques sous-jacents aux maladies. Au lieu de tester des composés chimiques de manière empirique, la pharmacologie inverse identifie d'abord les **protéines cibles potentielles associées à la maladie**. Ces cibles peuvent être des protéines, des enzymes, des récepteurs, ou d'autres composants moléculaires jouant un rôle crucial dans le processus pathologique. Préférée pour sa précision dans le ciblage des processus biologiques, la pharmacologie inverse permet une conception plus efficace des médicaments. (58) (59)

1.2. Recherche de la cible moléculaire

Dans la découverte de médicaments, l'objectif est de concevoir des traitements ciblant des **mécanismes moléculaires spécifiques** pour moduler le cours des maladies. Dans cette optique, l'identification de ces cibles est une étape stratégique majeure reposant sur une recherche intensive à la croisée de la biologie moléculaire, de la génomique, de la protéomique, de la bioinformatique et de la chimie médicinale.

La sélection de cibles **pertinentes sur le plan biologique** et **accessibles aux médicaments** (*druggable*) constitue un défi majeur, mais est un point de départ essentiel pour orienter la direction des recherches. L'IA est devenue un outil incontournable dans **la recherche de nouvelles cibles thérapeutiques** potentielles. (60)

Dans cette partie, nous développerons les applications de l'IA dans cette étape clé du développement d'un médicament. Nous commencerons par évoquer l'utilisation de l'IA dans la **recherche bibliographique** pour découvrir de nouvelles cibles, puis nous aborderons son rôle dans la **prédiction du rôle** d'une cible moléculaire dans une maladie et son **accessibilité pharmacologique** et enfin nous explorerons son rôle et son importance dans la **prédiction de la structure tridimensionnelle** d'une cible d'intérêt.

1.2.1. Exploration bibliographique visant à identifier de nouvelles cibles

La **littérature scientifique** constitue la principale source de connaissances sur les liens entre une cible moléculaire et une maladie donnée. Ainsi, l'**exploration bibliographique** pour la découverte de nouvelles cibles thérapeutiques s'appuie sur un vaste réservoir de données biologiques englobant des données génétiques, transcriptomiques et protéomiques, provenant de larges cohortes d'individus en bonne santé ou atteints de maladies spécifiques. La possibilité de collecter ces données à grande échelle et de les exploiter via des **bases de données publiques** ouvre de nouvelles voies pour l'identification et la validation précoce des cibles thérapeutiques potentielles.

Cette ampleur croissante des référentiels de littérature fait émerger un fort besoin d'outils permettant d'**identifier et de rassembler** les informations pertinentes tout en les **situant dans le contexte** des connaissances biomédicales actuelles. De plus, la complexité de ces ensembles de données multidimensionnelles requiert des **méthodes analytiques robustes** pour construire des modèles statistiquement valides capables d'identifier de nouvelles cibles potentielles.

Ainsi, l'**exploration de texte** (*text mining*), ainsi que l'IA, et en particulier le **NLP**¹, jouent un rôle crucial dans ce domaine. Ces technologies permettent d'identifier les articles pertinents, et d'extraire des informations exploitables à partir de **documents textuels non structurés** tels que des articles ou des rapports. L'IA facilite donc l'accès rapide à des connaissances dispersées dans une multitude de publications, rendant ainsi possible ce qui serait fastidieux voire impossible à réaliser manuellement. (22)

¹ Domaine de l'IA qui se concentre sur la compréhension et la manipulation du langage humain par les ordinateurs. Ses applications couvrent la traduction, la génération de texte, la classification de texte, l'extraction d'informations et le résumé de documents.

Les méthodes automatisées d'exploration de texte, axées sur la détection de relations entre les entités, sont essentielles pour extraire des connaissances pertinentes à partir de **larges ensembles de documents**. Ainsi, des systèmes d'exploration de texte tels que **BeFree** (61) utilisent les méthodes NLP pour repérer les **relations gènes-maladies, cibles-médicaments et médicaments-maladies** à partir de la littérature biomédicale.

Dans une récente étude, le système BeFree a été utilisé pour analyser 737 712 résumés d'articles scientifiques portant sur des maladies humaines et leurs gènes associés. Cette analyse a révélé **530 347 associations gène-maladie**, issues de **355 976 publications**.

Afin de s'assurer de la fiabilité et de la validité des associations gènes-maladies découvertes, un flux de travail basé sur un **arbre de décision** (*Figure 17*) a été mis en place. Plusieurs facteurs permettant d'éliminer les cibles non pertinentes ont ainsi été pris en compte, parmi lesquels se trouvent :

- Le **nombre de publications** dans lesquels une association est citée, désigné par les PMIDs (PubMed IDs), qui sont des numéros uniques donnés à chaque article sur PubMed
- Le **recoupement avec DisGeNET** (l'un des référentiels les plus importants et exhaustifs concernant les associations gènes-maladies)
- L'**impact factor** des revues dont l'article est tiré (mesure utilisée pour évaluer la pertinence et l'influence d'une revue scientifique dans le domaine académique)

Grâce à cette sélection basée sur cet arbre de décision, le nombre d'associations gène-maladie a été réduit à **330 888**, provenant de 334 983 articles scientifiques.

Au vu du nombre massif d'articles analysés, il apparaît évident qu'une telle analyse aurait été pratiquement impossible à réaliser manuellement. La capacité de BeFree à effectuer une analyse macroscopique des associations gène-maladie illustre sa puissance dans la manipulation de vastes ensembles de données textuelles.

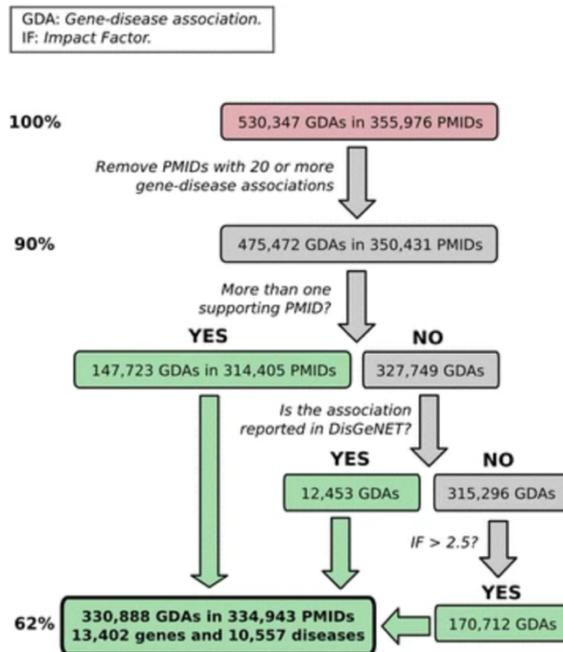


Figure 17 : Arbre de décision pour la sélection de l'ensemble de données BeFree sur les associations gènes-maladies (61)

De plus, les résultats obtenus indiquent que BeFree est capable d'identifier des associations gène-maladie **absentes des bases de données publiques** (Figure 18). Seules 7 669 associations gène-maladie soit **2 %** des associations détectées par BeFree étaient effectivement enregistrées sur DisGeNET, suscitant diverses interrogations quant à la façon dont les données sont gérées et hiérarchisées pour refléter les connaissances actuelles des scientifiques. (61)

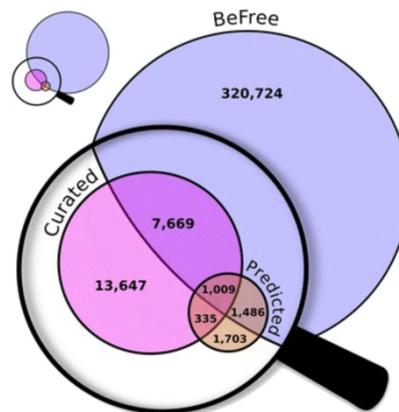


Figure 18 : Chevauchement des associations gènes-maladies identifiées par BeFree avec les associations disponibles sur DisGeNET (61)

Les associations gènes-maladies provenant de sources officiellement validées par des experts, comme UniProt sont appelées "curated", et celles provenant de prédictions basées sur des modèles animaux sont appelées "prédites".

Ainsi, BeFree aide les scientifiques à explorer les **mécanismes pathologiques** des maladies et à identifier de nouvelles cibles thérapeutiques potentielles et donc à découvrir de **nouvelles pistes de recherche**. En intégrant les informations extraites à partir d'articles scientifiques aux bases de données consolidées telles que DisGeNET, BeFree contribue à **enrichir les ressources disponibles** pour la communauté scientifique, facilitant ainsi la découverte de médicaments.

BeFree a également démontré sa capacité à réaliser une analyse **plus ciblée** et approfondie en étudiant les mécanismes pathologiques spécifiques d'une pathologie donnée. Ainsi, dans une étude de cas, la plateforme s'est basée sur **Medline**, une base de données bibliographique en ligne qui recense des articles de revues et des publications dans le domaine biomédical, pour rechercher les gènes associés à la **dépression**. S'il est largement admis que divers facteurs génétiques, épigénétiques et neurobiologiques contribuent à la dépression, les mécanismes exacts de cette pathologie et leur impact sur l'efficacité des thérapies demeurent mal compris. BeFree a réussi à **identifier** avec précision **les gènes associés à la dépression**, y compris certains qui ne sont pas répertoriés dans les ressources disponibles. (61) (Figure 19)

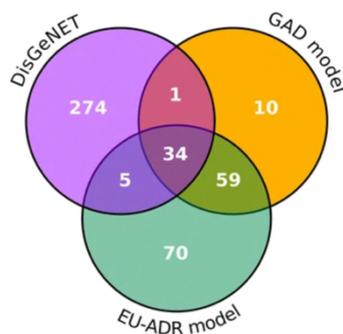


Figure 19 : Gènes de la dépression identifiés par BeFree et leur chevauchement avec les gènes disponibles dans d'autres référentiels (61)

Le diagramme de Venn illustre le recouvrement entre les gènes de la dépression identifiés par BeFree dans les corpus GAD ou EU-ADR, et ceux répertoriés dans DisGeNET

L'exemple de BeFree souligne les capacités de l'IA dans l'analyse bibliographique en démontrant leur application concrète dans l'identification de nouvelles cibles thérapeutiques.

1.2.2. Prédiction du rôle de la cible dans la maladie

Pour identifier des cibles pertinentes il est essentiel d'établir une **relation causale** entre la cible et la maladie en démontrant que la modulation de cette cible influence l'évolution de la maladie. L'IA, et en particulier le ML, permet d'analyser des données sur une cible potentielle afin de faire des prédictions sur sa causalité. (22)

Ainsi, par exemple, la **méthode des k-moyennes**² et le **kNN**³ ont été appliquées à des données protéomiques pour **classifier les sous-types de cancer du pancréas** et en révéler les caractéristiques moléculaires (*Figure 20*). Cette analyse a distingué **deux groupes** d'échantillons et identifié des **biomarqueurs spécifiques**, notamment les voies de signalisation impliquée dans l'oncogenèse. Ainsi, grâce au ML, des ensembles de protéines, d'ARNm, de miARN et de schémas de méthylation de l'ADN ont été d'identifiés comme des biomarqueurs précis permettant de différencier les deux sous-types de cancer du pancréas. La **pertinence biologique** de ces biomarqueurs a été évaluée à l'aide de programmes informatiques pour déterminer leur capacité à prédire la manière dont les cellules cancéreuses réagissent aux **traitements conventionnels**, orientant ainsi les futures stratégies thérapeutiques en identifiant de potentielles cibles thérapeutiques. (62)

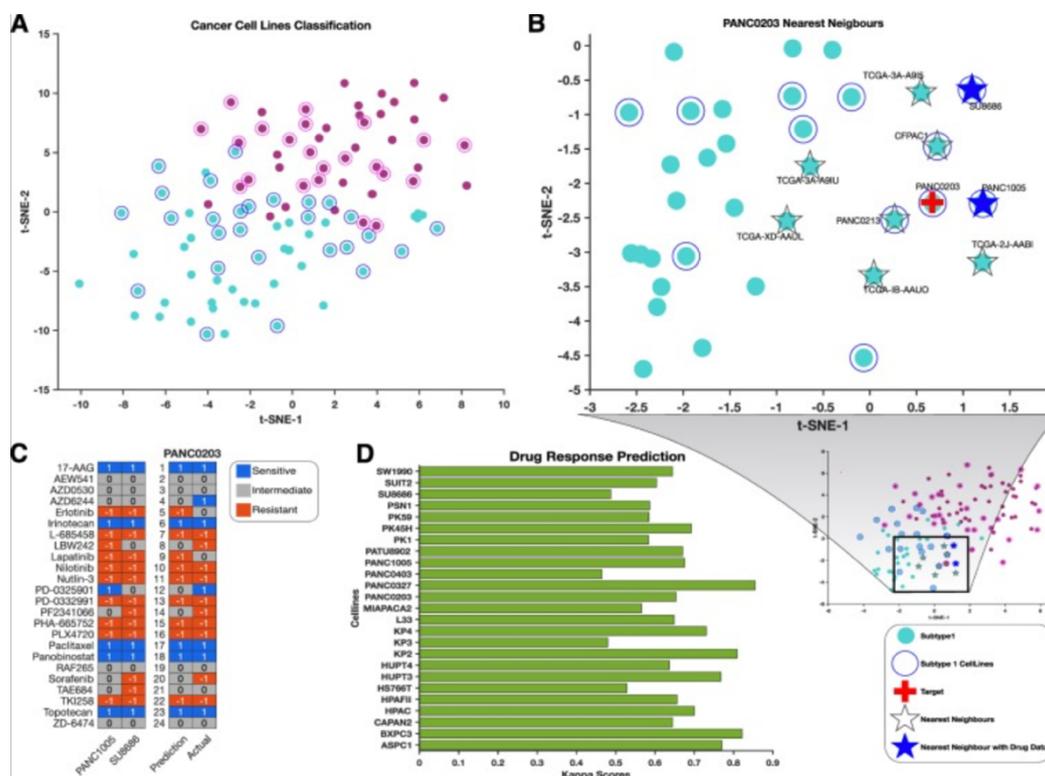


Figure 20 : Sous-typage des lignées cellulaires du cancer du pancréas et prédiction des réponses aux médicaments (62)

A) Visualisation des deux classes de tumeurs avec t-SNE (technique de réduction de dimensionnalité permettant de représenter les données complexes dans un espace en deux dimensions). Les points encerclés indiquent les lignées cellulaires classées par l'algorithme, tandis que les points non encerclés représentent les échantillons utilisés pour entraîner les algorithmes. B) Recherche des plus proches voisins de la lignée PANC0203 avec kNN C) Prédiction de la réponse à 24 médicaments anticancéreux des plus proches voisins de PANC0203 (PANC1005 et SU8686). Les deux dernières colonnes représentent les réponses médicamenteuses prédites de PANC0203 et ses réponses médicamenteuses réelles. Cela permet d'évaluer l'efficacité du modèle dans la prédiction de la réponse médicamenteuse pour PANC0203. D) Évaluation avec les scores kappa, mesure de la concordance entre les réponses médicamenteuses réelles et prédites des lignées cellulaires aux 24 anticancéreux

² Algorithme de regroupement non supervisé qui sépare un ensemble de données en k groupes en fonction de leurs similarités.

³ Algorithme de classification qui utilise les étiquettes des k points les plus proches pour prédire la classe d'un nouvel échantillon

Dans cette même optique d'identifier la causalité d'une cible potentielle dans une pathologie spécifique, plusieurs initiatives d'**apprentissage profond** ont été lancées. Parmi celles-ci, le **MOGONET** (Multi-Omics Graph cOnvolutional NETworks) se démarque. Ce cadre d'analyse des **données omiques** (génomique, protéomique et transcriptomique) se concentre sur des tâches de classification. Pour cela, MOGONET utilise les **GCN**⁴, qui lui permettent d'exploiter la **communication** entre les cibles et leurs voisins et d'intégrer les données multi-omiques afin de **classifier les cibles potentielles** et d'approfondir notre compréhension des maladies humaines. (Figure 21) (63)

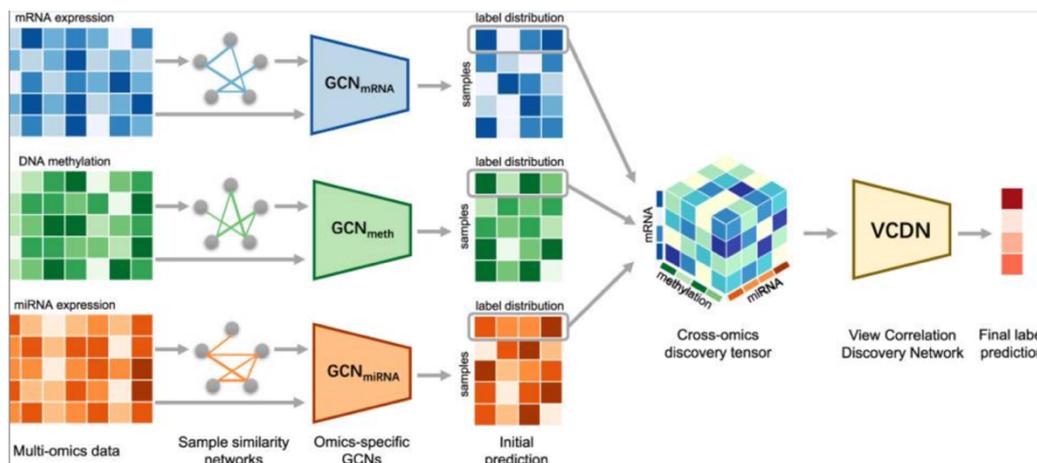


Figure 21 : Illustration du MOGONET

Chaque type de données omiques (ARN messenger, méthylation de l'ADN et micro ARN) est nettoyé pour enlever le bruit et les informations redondantes. Ensuite, un réseau GCN spécifique à chaque type de données omiques est formé pour prédire la classe de l'échantillon. Les prédictions initiales des réseaux GCN sont ensuite combinées dans un "tenseur de découverte cross-omics", qui est ensuite utilisé par le VCDN (View Correlation Discovery Network), qui est un réseau de neurones convolutifs, pour détecter des motifs ou des modèles significatifs.

Plus récemment, le GCNLDA, une méthode créée en 2019 basée sur les GCN et les réseaux de neurones convolutifs⁵, s'est révélé efficace dans l'identification de cibles thérapeutiques en examinant l'expression anormale des **ARN longs non codants (lncARN)** liés à diverses pathologies. Comparé à d'autres méthodes de prédiction à la pointe de la technologie, le GCNLDA a montré des performances supérieures (Figure 22). Des études sur le cancer de l'estomac, l'ostéosarcome et le cancer du poumon ont confirmé sa capacité à découvrir efficacement des associations potentielles entre les lncARN et les maladies. Par conséquent, le GCNLDA émerge comme un outil efficace pour pré-sélectionner des cibles fiables en vue de valider l'association ARNlnc-maladie à l'aide d'expériences biologiques. (64) (65)

⁴ Type de réseau de neurones utilisé pour le traitement des données structurées sous forme de graphes, où les nœuds représentent des entités et les arêtes des relations.

⁵ Type spécifique de réseau de neurones profonds dont l'architecture est spécifiquement conçue pour l'analyse de données structurées, telles que les images ou les séquences moléculaires (ici les lncARN)

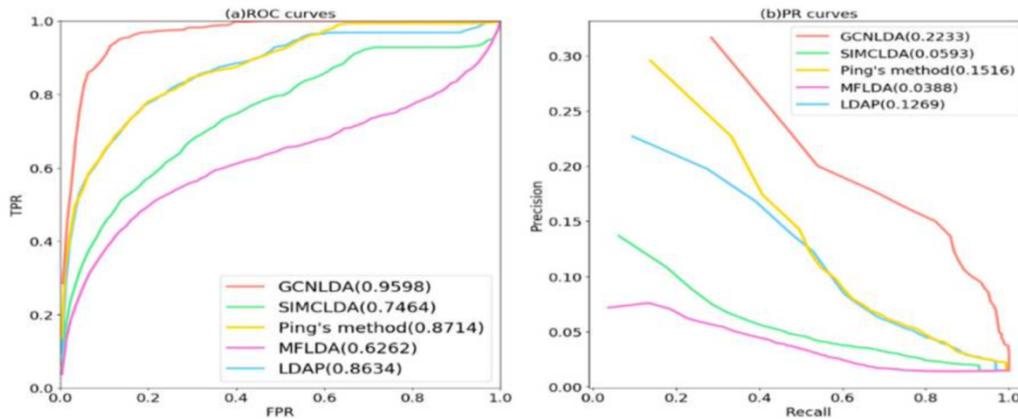


Figure 22 : Courbes démontrant la supériorité de GCNLDA dans la prédiction des associations lncRNA-maladie pour 405 pathologies (64)

La courbe ROC (Receiver Operating Characteristic) et la courbe PR (Precision-Recall) illustrent la performance du modèle GCNLDA par rapport aux autres méthodes. Ces courbes permettent de visualiser la capacité du modèle à prédire avec précision et exhaustivité les associations entre les ARN longs non codants et les pathologies.

1.2.3. Prédiction de l'accessibilité pharmacologique d'une cible

Pour identifier des cibles pertinentes, il est également nécessaire de s'assurer de leur accessibilité **pharmacologique**. Là encore, l'IA offre des solutions pour résoudre ce défi complexe. Par exemple, Jeon et ses collègues ont développé un **classificateur SVM**⁶ en 2014 pour prédire de nouvelles cibles de **médicaments anticancéreux dans les cancers du sein, du pancréas et de l'ovaire** (66). Ils ont dans un premier temps mis au point le classificateur, intégrant des ensembles de données génomiques et systématiques pour prioriser les cibles spécifiques aux cancers en question. Ensuite, ils ont étudié des stratégies pour **inhiber ces cibles** et sélectionné les cibles susceptibles d'être inhibées par de **petites molécules, des anticorps ou encore des peptides synthétiques**. Grâce à cette approche, **122 cibles universelles** (présentes dans tous les types de cancer) ont été identifiées, dont 69 se recoupant avec les 116 cibles déjà connues. De plus, respectivement **266, 462 et 355 cibles** spécifiques aux cancers du sein, du pancréas et de l'ovaire ont été identifiées (Figure 23). Deux des cibles prédites ont été **confirmées** à l'aide d'**inhibiteurs peptidiques**, montrant des effets antiprolifératifs significatifs dans des modèles de culture cellulaire. (66) (22)

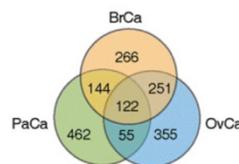


Figure 23 : Diagramme de Venn des cibles prédites par le classificateur SVM pour le cancer du sein (BrCa), du pancréas (PaCa) et de l'ovaire (OvCa)

⁶ Algorithmes d'apprentissage supervisé utilisés pour la classification, très employés dans la bioinformatique et la découverte de médicaments.

Dans une autre étude, une équipe de recherche a développé un méta-classificateur de ML pour prédire, à l'échelle du génome humain, les gènes morbides (c'est-à-dire les gènes dont les mutations provoquent des maladies héréditaires) et **pharmacologiquement accessibles**. Ce méta-classificateur a correctement identifié 65 % des gènes morbides connus avec une précision de 66 % et a repéré **78 % des gènes pharmacologiquement accessibles** avec une **précision de 75 %**. Il a ensuite été utilisé pour évaluer la morbidité et l'accessibilité pharmacologique de gènes qui non officiellement répertoriés comme morbides et / ou accessibles pharmacologiquement et les résultats obtenus ont démontré une concordance importance avec les données de la littérature.

Des arbres de décision (Figure 24) ont été générés en entraînant l'algorithme (J48) sur les ensembles de données de morbidité et d'accessibilité pharmacologique afin de découvrir des règles cellulaires pour ces deux critères. Parmi ces règles, le nombre de facteurs de transcription régulateurs et la localisation de la membrane plasmique se sont révélés être les facteurs les plus importants pour la morbidité et l'accessibilité pharmacologique, respectivement. (67)

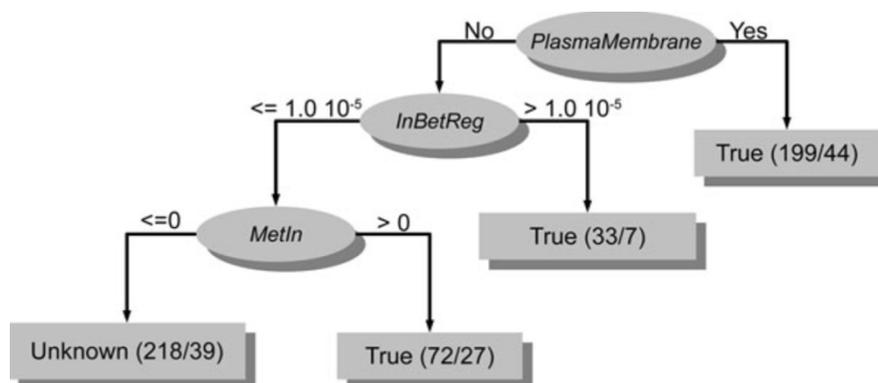


Figure 24 : Arbre de décision généré par l'apprentissage de l'algorithme J48 pour identifier les gènes pharmacologiquement accessibles (druggable) (67)

La racine de l'arbre représente la condition la plus influente pour différencier les gènes potentiellement réceptifs aux médicaments de ceux qui ne le sont pas. Ici, cette condition est la localisation des protéines codées à la membrane plasmique. Les ellipses suivantes sont des nœuds internes définissant des conditions supplémentaires pour classer un gène comme pharmacologiquement accessible ou non. Dans la branche gauche de l'arbre, ces conditions incluent la position centrale dans une voie de régulation transcriptionnelle (InBetReg) et le fait d'être une enzyme (MetIn). Les rectangles représentent les gènes classés comme accessibles pharmacologiquement (True) ou non (Unknown). Les chiffres entre parenthèses indiquent le nombre total de gènes classés correctement (avant la barre oblique) et le nombre de gènes incorrectement prédits (après la barre oblique).

Ces exemples illustrent la manière dont l'IA facilite la prédiction de l'accessibilité pharmacologique des cibles thérapeutiques, utilisant des techniques telles que les méta-classificateurs et les arbres de décision pour identifier avec précision les gènes susceptibles de répondre à des traitements pharmacologiques.

1.2.4. Accessibilité des données scientifiques sur les cibles thérapeutiques

L'IA ne se limite pas à l'identification du rôle d'une cible dans une pathologie et à l'évaluation de son accessibilité pharmacologique. Elle excelle également dans la consolidation, la visualisation et, surtout, la mise à disposition des données, rendant ainsi le savoir médical accessible à l'ensemble de la communauté scientifique.

Grâce à des méthodes sophistiquées d'analyse de données et de NLP, l'IA facilite l'accès et l'utilisation des données relatives à l'ensemble des cibles connues par les chercheurs. De plus, en utilisant des techniques de visualisation avancées, l'IA permet de représenter graphiquement des données complexes, facilitant ainsi leur compréhension et leur utilisation par un public plus large de scientifiques.

Open Targets illustre parfaitement cette approche en agrégeant, triant et organisant l'ensemble des données disponibles actuellement sur les cibles thérapeutiques. Cette plateforme est une initiative de recherche collaborative qui vise à accélérer la découverte de médicaments en **identifiant les meilleures cibles** et en permettant une meilleure compréhension des **mécanismes sous-jacents des maladies**. Pour ce faire, Open Targets utilise une approche d'**intégration** et de **visualisation des données**, facilitant ainsi la **hiérarchisation** des cibles potentielles. Cet outil, qui utilise le **ML**, répertorie toutes les **cibles** déjà connues et évalue leur **pertinence** par rapport aux pathologies. Les notions de sécurité de la cible (risque d'**effets indésirables**) ainsi que de **disponibilité / accessibilité pharmacologique** des cibles sont intégrées aux données. Les utilisateurs ont le choix entre un **flux de travail axé sur la cible** pour découvrir les maladies potentiellement associées à une cible spécifique, ou un **flux de travail centré sur la maladie** pour identifier les cibles associées à une maladie donnée (*Figure 25*). (68)

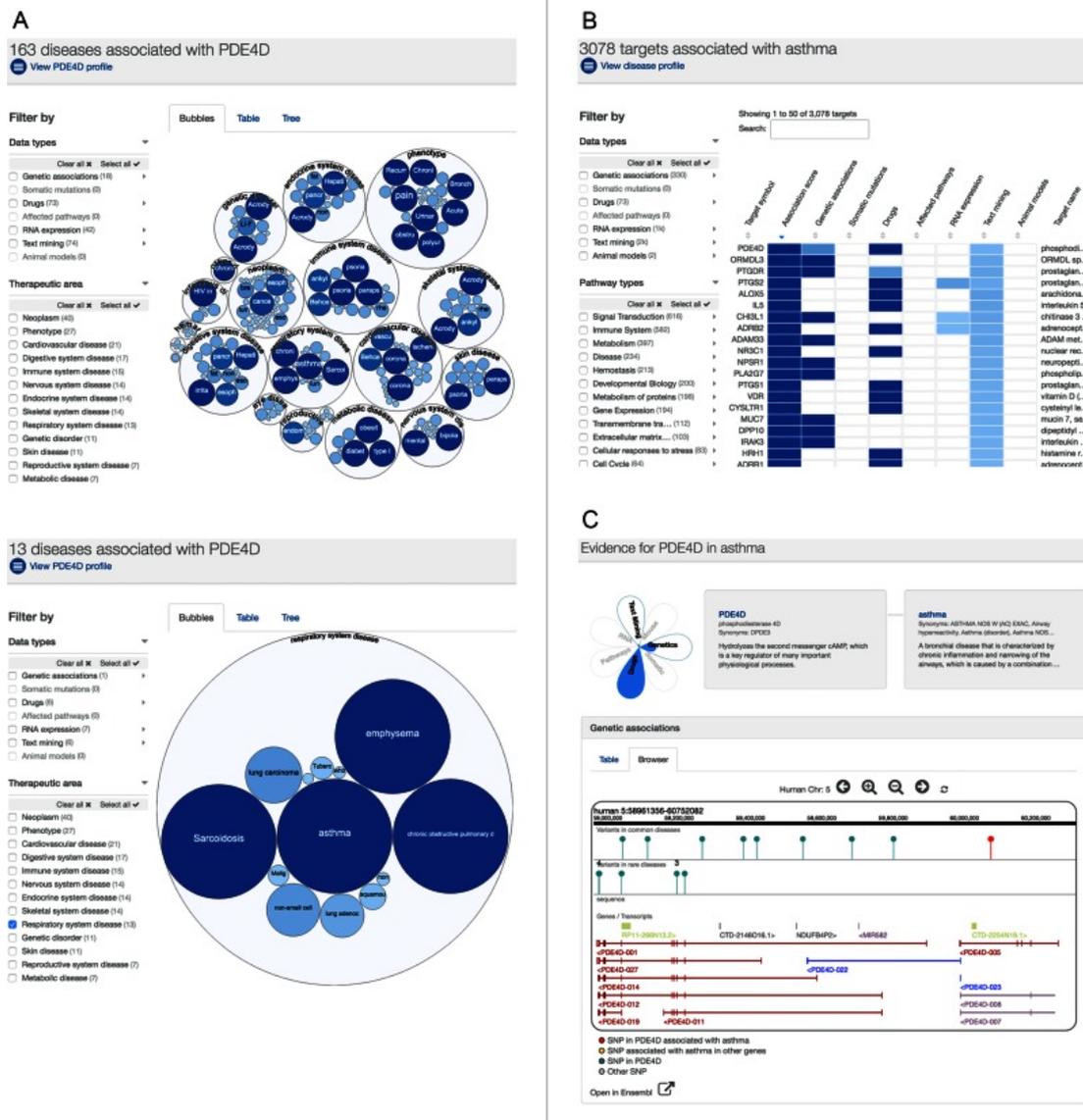


Figure 25 : Flux de travail de la plateforme Open Targets (68)

(A) Flux de travail basé sur une cible (PDE4D ici) : L'utilisateur recherche le gène et reçoit une liste de toutes les maladies associées à ce gène, représentées sous forme de "bulles" regroupées en domaines thérapeutiques. La taille et la teinte de chaque bulle varient en fonction de la force de l'association entre la maladie et la cible. L'utilisateur peut filtrer les maladies d'un domaine thérapeutique particulier associé à la cible en question. (B) Flux de travail de la plateforme basé sur une pathologie : L'utilisateur saisit le terme de la maladie (asthme ici) et reçoit une liste de toutes les cibles associées à l'asthme, y compris la PDE4D. (C) Preuves de la présence de la PDE4D dans l'asthme : En cliquant sur "asthme" dans le panneau A ou sur "PDE4D" dans le panneau B, l'utilisateur peut examiner les types de preuves qui soutiennent l'association entre la PDE4D et l'asthme. La "fleur" des preuves donne un aperçu de la force de l'association pour chaque type de preuve, et les détails des preuves disponibles sont présentés sous forme de tableaux récapitulatifs ou de graphiques.

En consolidant et mettant à disposition les données à l'ensemble des scientifiques, Open Targets et plus largement l'IA favorisent une recherche médicale plus collaborative, transparente et efficace, propulsant ainsi nos capacités de compréhension et de traitement des maladies vers de nouveaux sommets.

1.2.5. Prédiction de la structure tridimensionnelle d'une cible d'intérêt

La prédiction de la structure tridimensionnelle d'une cible potentielle est indispensable pour évaluer et exploiter efficacement son potentiel thérapeutique. En effet, la grande majorité des cibles thérapeutiques sont des **protéines** jouant des rôles importants dans divers processus biologiques tels que les réactions enzymatiques, la signalisation cellulaire et les interactions intercellulaires. Or la **fonction** de ces protéines est intimement liée à leur **structure tridimensionnelle**. Il est donc crucial d'identifier cette dernière pour les protéines ayant un intérêt thérapeutique potentiel, afin d'identifier des **sites de liaison pour de futurs médicaments**.

Cette logique établit les fondements de la **conception de médicaments basée sur la structure** ou **Structure-Based Drug Design (SBDD)** en anglais. Cette approche (*Figure 26*), essentielle dans la recherche universitaire et industrielle, est précise, rapide et efficace car elle analyse la structure tridimensionnelle des protéines cibles et son lien avec la pathologie (69) (58)

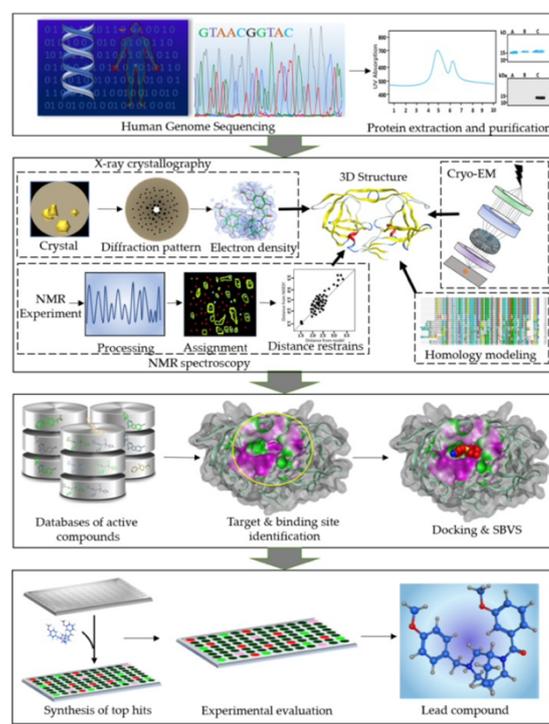


Figure 26 : Processus de conception de médicaments basée sur la structure (58)

Le SBDD commence par le séquençage du génome humain et l'extraction des protéines cibles : Le génome humain est séquencé pour identifier les gènes qui codent les protéines cibles potentielles. Ces protéines sont alors extraites et purifiées à partir des échantillons biologiques appropriés. Ensuite, la structure des protéines thérapeutiquement importantes est déterminée à l'aide de techniques de biologie structurale telles que la cristallographie aux rayons X et la RMN (Résonance Magnétique Nucléaire). Une base de données des composés actifs est ensuite préparée, contenant une variété de molécules potentielles pour le criblage. À partir de là, la protéine cible pharmacologiquement pertinente et accessible et son site de liaison sont identifiés, et les composés actifs sont passés au crible et ancrés dans la cavité de liaison de la protéine cible. Les composés les plus prometteurs obtenus par ce criblage virtuel sont ensuite synthétisés et testés *in vitro*, avec la possibilité de modifications supplémentaires pour optimiser l'efficacité du composé principal (lead compound).

Bien que des méthodes expérimentales comme la spectroscopie par RMN soient disponibles, leur coût élevé et leur longue durée d'exécution limitent leur capacité à résoudre les structures tridimensionnelles des protéines (70). Jusqu'à présent, ces méthodes n'ont permis de déterminer les structures que d'environ 100 000 protéines uniques, ce qui ne représente qu'une fraction du nombre total de protéines répertoriées. (71) Par conséquent, il est impératif de développer de nouvelles approches pour combler le fossé entre le nombre croissant de **séquences protéiques** identifiées et le nombre limité de **structures tridimensionnelles** protéiques connues. (72)

Ainsi, avec l'avancée de la puissance de calcul et des technologies d'IA, de nombreuses approches informatiques ont émergé pour **prédire ces structures tridimensionnelles**. Parmi elles, **AlphaFold**, un système de DL développé par **DeepMind**, est particulièrement performante. Cette méthode repose sur un **réseau neuronal** capable de prédire avec précision les structures tridimensionnelles des protéines à **partir de leurs séquences d'acides aminés**, avec une précision similaire à celle obtenue par les méthodes expérimentales classiques. (Figure 27) (71)

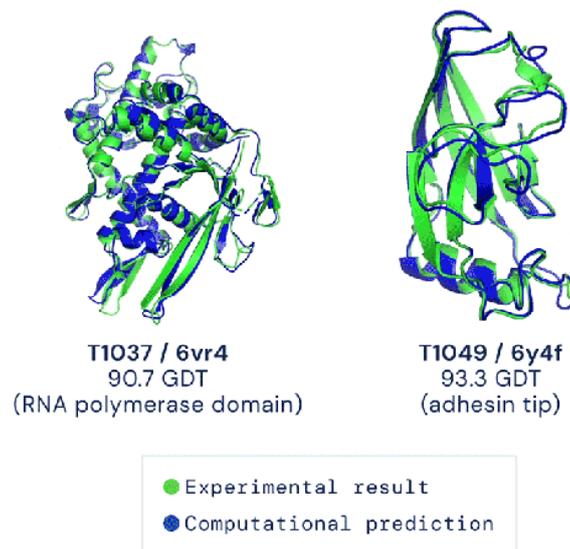


Figure 27 : Exemples de prédictions de structures protéiques générées par AlphaFold (74)

Les données expérimentales sont représentées en vert, tandis que les prédictions de AlphaFold sont en bleu. Les deux ensembles de données sont superposés à titre d'illustration, permettant ainsi de visualiser la similarité des résultats.

Ainsi, en juillet 2022, les structures de **200 millions de protéines** ont été prédites en seulement un an, grâce à AlphaFold et à l'effort collectif de la communauté scientifique. (73)

En partenariat avec le laboratoire européen de biologie moléculaire ou European Molecular Biology Laboratory (EMBL), les scientifiques ont donc constitué la base de données **AlphaFold Protein Structure Database**. Accessible librement, cette base fonctionne comme un **moteur de recherche** (Figure 28) permettant aux chercheurs du monde entier d'accéder aux prédictions de la structure 3D de la plupart des protéines connues, quelle que soit son origine (animale, bactérienne, végétale ou fongique). Ces prédictions ouvrent de nouvelles perspectives pour le traitement de diverses maladies en comprenant mieux le repliement des protéines. (75)



Figure 28 : Page d'accueil de la base de données AlphaFold Protein Structure Database

Bien que la validation finale de la cible n'intervienne qu'ultérieurement, lors des essais cliniques, cette étape de sélection et de validation précoce de la cible est cruciale pour concentrer les efforts sur les projets les plus prometteurs. Ainsi, l'intégration de l'IA dans ce processus de recherche de la cible moléculaire, que ce soit à travers l'identification de nouvelles cibles dans la littérature scientifique, la prédiction de leur rôle dans les mécanismes physiopathologiques, l'évaluation de leur accessibilité pharmacologique ou encore la projection de leur structure tridimensionnelle, optimise de manière significative la sélection et la pré-validation des cibles thérapeutiques. L'IA prépare ainsi efficacement la phase suivante, qui vise à explorer et optimiser les molécules pouvant interagir avec ces cibles.

1.3. Identification et optimisation de candidats-médicaments

L'IA joue un rôle de plus en plus important dans la conception rationnelle de médicaments, en permettant une analyse approfondie des relations entre la structure chimique des composés et leur activité biologique. Cette technologie facilite la **prédiction de l'efficacité et de la sécurité des candidats-médicaments**. Dans cette section, nous examinerons les différentes applications de l'IA à diverses étapes clés du processus de **découverte** de médicaments. Cela inclut le criblage de molécules candidates, la prédiction de leurs caractéristiques, la conception *de novo* de candidats-médicaments et la réutilisation de molécules existantes pour d'autres indications (*drug repurposing*).

1.3.1. Criblage de molécules d'intérêt

Une fois les cibles identifiées, la première étape de la découverte de médicament consiste à **cribler des bibliothèques de petites molécules**, c'est-à-dire à examiner un grand nombre de molécules pour sélectionner les plus pertinentes. L'objectif est d'identifier des composés qui peuvent interagir de manière **spécifique** avec les cibles préalablement établies, en privilégiant ceux présentant le **meilleur rapport bénéfice/risque**.

Grâce à la **modélisation moléculaire** et à la **simulation informatique**, les chercheurs peuvent concevoir des médicaments ciblant directement les mécanismes moléculaires sous-jacents de la maladie, tout en réduisant le risque d'effets indésirables (EI). (59)

Une illustration concrète de cette approche est **Atomwise**, une entreprise basée à San Francisco qui s'est spécialisée dans la découverte de médicaments alimentée par l'IA. À travers sa plateforme **AtomNet®**, qui repose sur le **DL**, Atomwise exploite les structures moléculaires pour concevoir des médicaments de manière ciblée. Cette plateforme, fondée sur un **CNN**⁷, cartographie **les sites de liaison protéine-ligand en trois dimensions, sans intervention humaine**, ce qui lui permet d'analyser des milliards de composés afin de repérer ceux qui présentent une haute spécificité pour une cible donnée.

⁷ Type spécifique de réseau de neurones profonds principalement utilisé dans le traitement d'images mais également adaptés à l'analyse de données structurées, telles que les structures moléculaires (les sites de liaison protéine-ligand ici)

Dans une étude appelée DUD-E, un groupe de l'Université de Californie à San Francisco a comparé les prédictions d'AtomNet® et d'autres plateformes à des résultats d'expériences physiques passées. Les résultats ont été très favorables pour AtomNet®, qui a démontré son efficacité en obtenant une AUC (Area Under the Curve) supérieur à **0,9** pour **57,8%** des cibles étudiées. (76)

D'après Atomwise, la technologie AtomNet® pourrait **réduire de moitié le nombre d'expériences nécessaires au criblage de médicaments** et **améliorer considérablement le taux de réussite** des autres. Depuis 2013, AtomNet® a été largement utilisé dans la recherche sur les médicaments, prédisant de nouveaux traitements potentiels dans des domaines variés tels que le cancer, les maladies neurologiques, les antiviraux, les antiparasitaires et les antibiotiques. Grâce à cette technologie, Atomwise a pu développer un **portefeuille exclusif de candidats-médicaments** (Figure 29), certains ayant déjà démontré des résultats prometteurs lors d'études animales. (77)

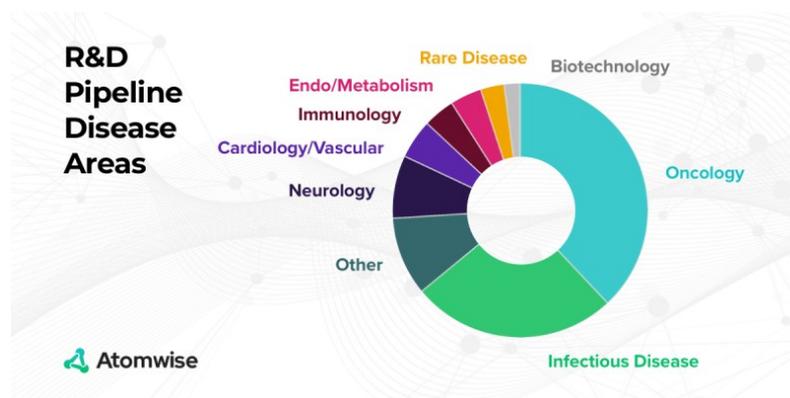


Figure 29 : Domaines thérapeutiques du pipeline de R&D d'Atomwise (78)

De nombreuses entreprises pharmaceutiques ont conclu des partenariats avec Atomwise pour stimuler leur recherche et développement. Parmi ces entreprises, on compte **Bayer** (79), **GC Pharma** (80) et **Sanofi**, qui a signé en 2022 un accord de 1,2 milliard de dollars avec Atomwise pour exploiter la plateforme AtomNet® afin de rechercher de petites molécules ciblant jusqu'à cinq cibles médicamenteuses. (81)

Les succès actuels d'Atomwise ainsi que ces collaborations prometteuses démontrent de manière éloquent l'énorme potentiel de l'IA dans le criblage de molécules.

1.3.2. Prédiction des caractéristiques des molécules d'intérêts

La prédiction des caractéristiques des molécules criblées est essentielle car elle permet d'évaluer leur **sécurité** et leur **efficacité**, guidant ainsi la sélection des candidats-médicaments les plus prometteurs pour les étapes ultérieures du développement.

a) Prédiction de la toxicité du candidat-médicament

La prédiction de la toxicité des candidats-médicaments vise à **anticiper les EI** que ces molécules pourraient entraîner chez les patients. En effet, les évaluations précliniques sont indispensables pour écarter les médicaments toxiques en prenant des décisions éclairées concernant la poursuite ou non de leur développement.

Toutefois, malgré ces précautions, la toxicité élevée reste un obstacle majeur, représentant **un cinquième des échecs au cours des essais cliniques** (82) et les **deux tiers des retraits de médicaments après la mise sur le marché**. (83)

Les tests in vivo sont **coûteux** et soulèvent des préoccupations **éthiques** en raison des expériences sur les animaux. Les méthodes de **prédiction informatique** ou **toxicologie in silico** révolutionnent l'évaluation de la sécurité des médicaments. Ce processus implique la collecte de données sur la toxicité des composés, la sélection des descripteurs moléculaires appropriés, la construction, la validation et enfin l'interprétation du modèle (*Figure 30*). Ce travail rigoureux permet de créer des outils de prédiction fiables, réduisant ainsi les risques pour la santé humaine.

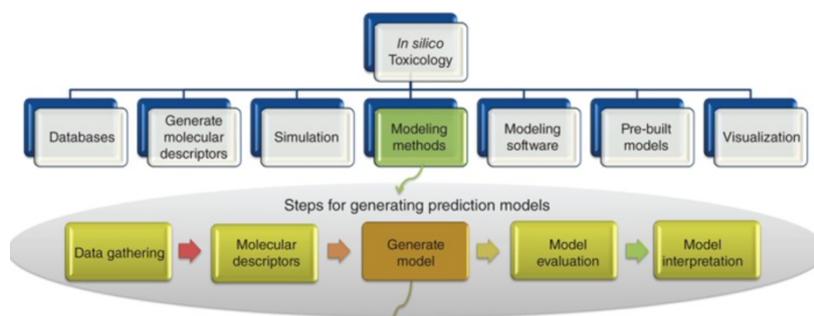


Figure 30 : Outils de toxicologie in silico et étapes pour générer des modèles de prédiction (84)

Les méthodes de modélisation computationnelle offrent donc une alternative **économique** et **efficace** pour évaluer la toxicité des candidats-médicaments.

Dans cette optique, plusieurs méthodes basées sur des techniques d'IA ont été développées pour prédire la toxicité de substances chimiques. (85) (86) La compétition "**Toxicology in the 21st Century (Tox21)**" a été lancée pour évaluer les performances de ces différentes méthodes de prédiction. (84) Les résultats de cette compétition ont révélé la capacité des réseaux de neurones profonds (**Deep Neural Network ou DNN**) à fournir des prédictions précises sur 12 paramètres toxiques différents préalablement sélectionnés. (Figure 31).(87)

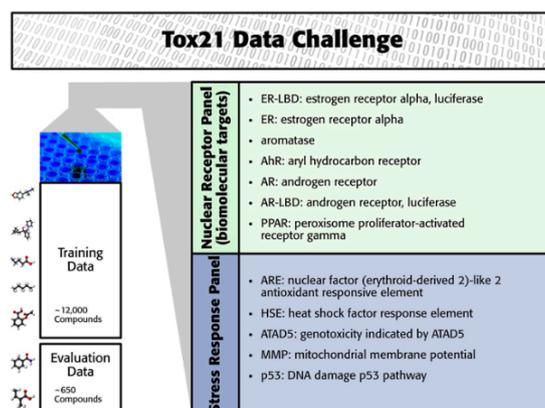


Figure 31 : Aperçu de l'ensemble de données du défi Tox21 et des 12 paramètres toxiques (87)

Ces modèles utilisent des **descripteurs moléculaires** comme la présence ou l'absence de **toxicophores** (structures moléculaires responsables de l'activité toxique) (Figure 32). Les DNN ont été capables d'extraire des caractéristiques moléculaires associées à ces paramètres toxiques, suggérant que ces réseaux apprennent des représentations abstraites dans leurs couches cachées. Toutefois, bien que prometteuses, les caractéristiques moléculaires identifiées nécessitent souvent une expertise humaine pour être directement appliquées à la découverte de médicaments. (57)

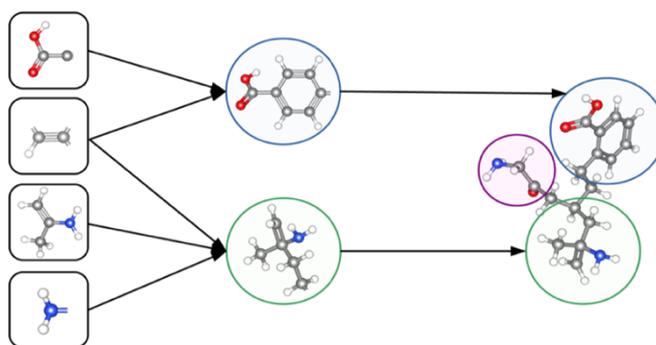


Figure 32 : Construction de pharmacophores par des réseaux neuronaux (87)

Les RNA construisent des caractéristiques complexes, comme les pharmacophores, en utilisant des sous-structures plus simples.

DeepTox est un exemple de modèle d'IA qui a obtenu de très bons résultats au challenge Tox21. Ce logiciel, spécialement conçu pour prédire la toxicité des produits chimiques, repose sur un **DNN à trois couches** (87). Après un processus de nettoyage et de contrôle de la qualité des données, les composés chimiques sont encodés à l'aide de **descripteurs moléculaires 0D à 3D** utilisés comme **entrée** pour le DNN. Puis, DeepTox entraîne différents modèles, les évalue et combine les plus performants pour prédire la toxicité de nouveaux composés. En ajustant les hyperparamètres tels que le nombre d'unités cachées, le taux d'apprentissage et le taux d'abandon, DeepTox étudie de manière précise la structure et peut **identifier d'éventuels toxicophores**, en les comparant avec une base de données contenant plus de 2500 toxicophores. Les résultats basés sur l'ensemble de données Tox21 démontrent que DeepTox surpasse ses concurrents en termes de prédiction de la toxicité.

Plus globalement, les **relations quantitatives structure-activité** (ou QSAR, pour *Quantitative Structure-Activity Relationships* en anglais) établissent, comme leur nom l'indique, des relations quantitatives entre les caractéristiques chimiques ou structurales d'un composé et son activité pharmacologique. (88) Cette approche est largement utilisée pour modéliser divers paramètres d'innocuité des médicaments, tels que la **Dose Létale à 50% (DL50)**⁸ et la **toxicité spécifique aux tissus**. L'ampleur et la complexité des données qui découlent de cette approche en font un terrain de jeu idéal pour l'IA.

b) Prédiction de la bioactivité du candidat-médicament

La prédiction de la **bioactivité** des candidats-médicaments vise à anticiper leur efficacité potentielle en tant qu'agents thérapeutiques. En évaluant cette activité biologique, il devient possible de déterminer si la molécule est susceptible d'avoir l'effet désiré sur la cible biologique spécifique préalablement définie. Cette étape est très importante étant donné que l'**inefficacité** des candidats constitue l'un des principaux obstacles dans la découverte de médicaments. En sélectionnant soigneusement les candidats-médicaments les plus prometteurs pour une évaluation plus approfondie lors des étapes ultérieures du processus de développement de médicaments, cette prédiction permet de réduire à la fois les coûts et le temps nécessaires au développement de nouveaux traitements.

⁸ Dose d'une substance d'essai administrée en une seule fois qui entraîne la mort de 50 % des animaux exposés.

Traditionnellement, les chercheurs ont utilisé des **expériences in vitro et in vivo** (coûteuses et chronophages) pour simuler les fonctions des molécules dans le corps humain. De nos jours, les chercheurs se tournent de plus en plus vers des méthodes informatiques et des modèles mathématiques basés sur des données expérimentales antérieures. Ainsi, les techniques d'IA ont été appliquées avec succès à la prédiction de la bioactivité de candidats-médicaments, notamment pour anticiper leurs **activités anticancéreuses, antivirales et antibactériennes**. Ces modèles prennent en compte diverses caractéristiques moléculaires de la substance active, telles que sa **structure chimique, sa polarité, sa charge, sa taille**, etc. pour ensuite prédire l'interaction de la molécule avec la cible biologique souhaitée et l'effet thérapeutique recherché.

La modélisation QSAR, qui utilise souvent des techniques telles que les **SVM** et les **forêts aléatoires** en raison de leur précision, leur robustesse et leur interprétabilité fournit également des informations sur la bioactivité d'un composé. (89)

Ainsi, dans une récente modélisation QSAR des inhibiteurs de l'histone désacétylase 1 (HDAC1), la **SVM**⁹ a montré de très bonnes performances dans la **prédiction de la valeur d'activité**, avec une **AUC de 0,91** (90). L'enzyme HDAC1 joue un rôle clé dans la régulation de l'expression génique en modifiant les histones, influençant ainsi divers processus biologiques tel que la différenciation cellulaire. De ce fait, certains inhibiteurs de l'HDAC1 sont prometteurs pour le développement de médicaments ciblant des maladies associées à des perturbations de l'expression génique, telles que le cancer et les maladies neurologiques. Ces performances élevées de la SVM dans la prédiction de l'activité des inhibiteurs de l'HDAC1 soulignent son potentiel prometteur dans le développement de médicaments pour ces maladies.

Par ailleurs, dans une autre étude, les **forêts aléatoires**¹⁰ se sont avérées efficaces pour **prédire la liaison aux récepteurs des œstrogènes** et d'autres bioactivités, offrant une précision interne allant jusqu'à 92 % et une validation externe atteignant 89 %. Ces résultats soulignent le potentiel des forêts aléatoires pour évaluer **l'activité endocrinienne** des produits chimiques. (91)

⁹ Méthode d'apprentissage automatique dont l'efficacité découle de sa capacité à trouver des frontières de décision optimales entre différentes classes de données, ce qui la rend particulièrement adaptée pour prédire l'activité de composés comme les inhibiteurs de l'HDAC1

¹⁰ Modèle d'apprentissage automatique composé de multiples arbres de décision, utilisé pour la classification et la régression.

En outre, Stokes et ses collaborateurs ont développé en 2020 un **DNN**¹¹ capable de prédire **l'activité antibactérienne** des composés. Ce modèle extrait des caractéristiques atomiques et de liaison (type de liaison, stéréochimie...), puis optimise ces caractéristiques pour estimer la probabilité d'activité antibactérienne de chaque molécule. Ainsi, ce travail a démontré l'utilité et l'efficacité des approches de DL pour élargir l'arsenal de médicaments antibiotiques, dans un contexte d'émergence rapide de bactéries résistantes aux antibiotiques classiques. (92) (73)

c) **Évaluation des propriétés ADME du candidat-médicament**

Cette étape consiste à évaluer les propriétés **d'absorption, de distribution, de métabolisme, d'excrétion** (ADME) des candidats-médicaments. L'objectif est de comprendre comment les médicaments interagissent avec l'organisme. En effet, les propriétés ADME jouent un rôle clé dans la détermination de l'efficacité et la sécurité des médicaments en développement et sont l'une des causes majeures de défaillance des molécules potentielles dans le processus de découverte de médicaments, engendrant des dépenses significatives en temps, en capital et en ressources humaines. Ainsi, pour accroître les probabilités de succès des composés, un intérêt grandissant émerge pour la **prédiction anticipée des propriétés ADME** des candidats-médicaments. Dans cette optique, l'IA est de plus en plus utilisée pour élaborer des modèles et des outils de prédiction afin d'évaluer plus rapidement et plus précisément ces propriétés. (93)

Les caractéristiques **physico-chimiques** des candidats-médicaments, telles que leur solubilité, leur stabilité chimique, leur polarité, leur taille moléculaire sont essentielles pour comprendre leur mode d'action et ont une forte influence sur leurs propriétés ADME. Parmi ces caractéristiques, la **solubilité** revêt une importance particulière car elle influence les **propriétés pharmacocinétiques** et la **formulation** des médicaments. (94) (95). Cependant, la prédiction précise de la solubilité a été entravée par des méthodes expérimentales longues et coûteuses, incitant ainsi le développement de modèles prédictifs basés sur l'IA. Ainsi, différentes approches d'apprentissage profond, telles que les **RNN**¹² et les **GCN**¹³ ont été évaluées pour prédire la solubilité des composés (96).

¹¹ Réseaux de neurones dits profonds car composés de plusieurs couches de neurones. Leur capacité à extraire des caractéristiques complexes à partir des données d'entrée les rend adaptés pour prédire l'activité antibactérienne des composés.

¹² Modèles d'apprentissage automatique idéaux pour anticiper la solubilité des composés, grâce à leur aptitude à traiter les données séquentielles et à capturer les relations entre les caractéristiques moléculaires

¹³ Type de réseaux de neurones efficaces pour prédire la solubilité des composés en raison de leur capacité à traiter les structures moléculaires sous forme de graphes, leur permettant de saisir les relations complexes entre les atomes et les liaisons.

La prédiction des propriétés ADME peut être abordée comme un problème de **classification ou de régression** puisqu'il s'agit soit de catégoriser les molécules en fonction de leur capacité à être absorbées, distribuées, métabolisées et excrétées, soit de prédire des valeurs numériques pour ces paramètres. Dans ce contexte, les **GCN** se révèlent particulièrement efficaces pour prédire les propriétés ADME des médicaments à partir de leurs structures moléculaires. En effet, leur forte capacité de représentation des caractéristiques permet aux GCN de capturer les descripteurs moléculaires des candidats-médicaments (propriétés physico-chimiques, représentations moléculaires et propriétés pharmacologiques) à partir des caractéristiques des sous-structures de la molécule (*Figure 33*) (94).

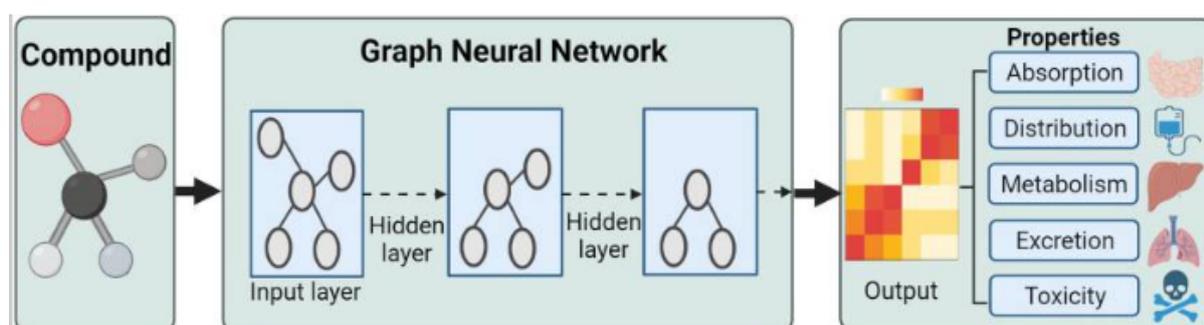


Figure 33 : Prédiction des propriétés ADMET des médicaments par un réseau neuronal à base de graphes (65)

Le réseau neuronal à base de graphes capture les caractéristiques liées aux propriétés des médicaments à partir de leur structure moléculaire pour prédire les propriétés ADMET.

Chemi-Net est un exemple de modèle de réseau neuronal qui utilise les GCN pour la prédiction des propriétés ADMET. Le processus du Chemi-Net consiste d'abord à caractériser les atomes de la molécule de médicament ainsi que les relations entre eux. Pour cela, les **SMILES** (Simplified Molecular Input Line Entry System), représentations linéaires de structures moléculaires où chaque caractère représente un atome ou une liaison chimique, servent **d'entrée** au modèle (tandis que la sortie du Chemi-Net est la prédiction des propriétés ADMET). Ce processus de prédiction des propriétés ADMET se déroule en plusieurs étapes : tout d'abord, le modèle projette **l'agencement des atomes** dans un **espace tridimensionnel** pour créer une représentation graphique de la molécule du candidat-médicament (*Figure 34*). Ensuite, le Chemi-Net utilise une série d'opérations de convolution de graphe pour extraire les caractéristiques de ces structures puis effectue un **pooling** afin de réduire la dimension spatiale des données tout en préservant les caractéristiques importantes. Enfin, les prédictions précises des propriétés ADMET sont obtenues en faisant passer les données par des **couches entièrement connectées**. (94) (97)

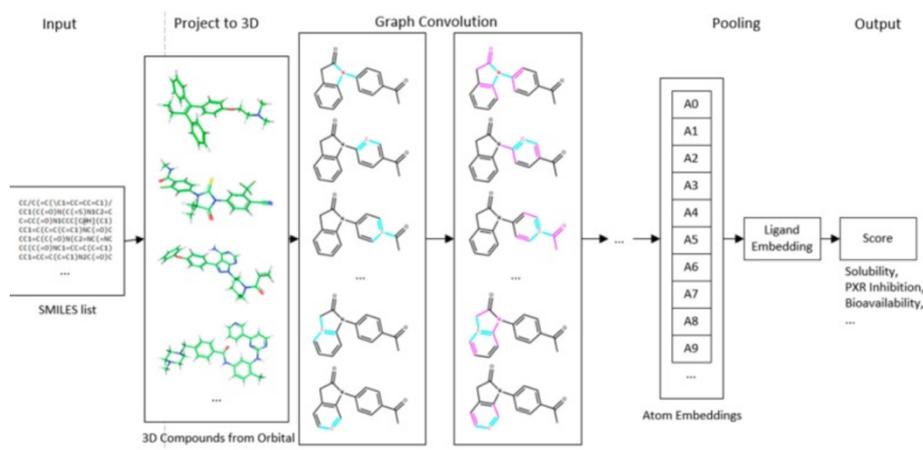


Figure 34 : Architecture globale de Chemi-Net (98)

Dans cette illustration, le rouge indique l'atome central, tandis que le cyan désigne les atomes voisins pertinents. Lors de la deuxième couche de convolution, le violet met en évidence la zone où les informations sont transmises à partir de l'étape de convolution précédente.

Pour clore cette analyse sur l'apport de l'IA dans la prédiction des propriétés des molécules d'intérêt, il apparaît évident que les progrès dans les techniques d'apprentissage automatique et d'apprentissage profond ont permis des avancées significatives. En combinant différentes approches telles que les réseaux de neurones profonds, les SVM et les forêts aléatoires, les chercheurs peuvent prédire avec précision la toxicité, la bioactivité et les propriétés ADME des candidats-médicaments. Ces avancées offrent des outils puissants pour accélérer et optimiser la sélection des composés les plus prometteurs pour les étapes ultérieures de recherche et de développement.

1.3.3. La conception de novo de candidats-médicaments

L'IA pave également la voie à la **conception de novo de candidats-médicaments** en générant de nouvelles molécules dotées des propriétés pharmacologiques souhaitées. La conception *de novo*, qui implique la création de nouveaux composés médicamenteux **sans modèle de départ**, a été traditionnellement limitée par la nécessité de connaître **le site actif** d'une cible ou les **pharmacophores**¹⁴ d'un ligand. Cependant, l'avènement des techniques d'IA a révolutionné cette approche, avec l'émergence ces dernières années de modèles prometteurs tels que **ReLeaSE** (99), **GraphINVENT** (100) et **MolRNN** (101), qui reposent sur des méthodes de DL pour accélérer le processus de découverte du médicament. Pour illustrer la puissance de l'IA dans la conception de novo de médicaments et ses résultats concrets, nous nous pencherons sur l'exemple de Chemistry42.

¹⁴ Partie d'une molécule responsable de son activité biologique produisant un effet thérapeutique.

Chemistry42 est une plateforme de ML appartenant à **Insilico Medicine** qui utilise l'**Intelligence Artificielle Générative (IAG)**¹⁵ pour la conception de médicaments *de novo*. En avril 2023, Insilico a annoncé que sa plateforme avait découvert une **petite molécule inhibitrice de CDK8** présentant une **bonne sélectivité et biodisponibilité** par voie orale, offrant ainsi un **traitement potentiel pour le cancer**. La dérégulation de CDK8/19 est fortement impliquée dans de nombreux cancers humains, en particulier dans la leucémie myéloïde aiguë (LMA) et les tumeurs solides avancées, ainsi que dans la modulation immunitaire. Le ciblage de CDK8/19 est considéré comme une **stratégie prometteuse**, offrant un double mécanisme d'action : inhibition de la croissance tumorale et stimulation de l'activité immunitaire.

En s'appuyant sur les principes de la conception de médicaments basée sur les **ligands (LBDD)** et la **SBDD** (Figure 35), Chemistry42 se repose sur des années de modélisation et de pré-entraînement sur de vastes ensembles de données biologiques, chimiques et textuelles. Le modèle intègre **42 moteurs génératifs** et plus de **500 moteurs prédictifs**, permettant aux chercheurs de **concevoir des molécules répondant à des critères spécifiques à partir de zéro**, en exploitant des technologies de DL. (102)

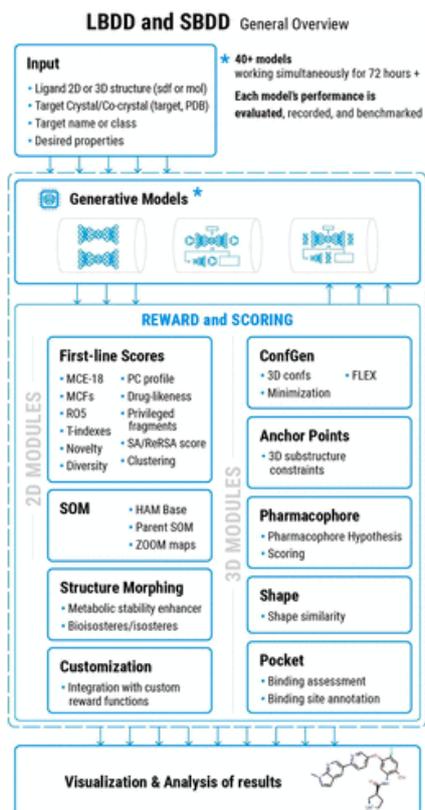


Figure 35 : Représentation du flux de travail pour une expérience générative *de novo* avec la plateforme Chemistry42 (103)

La première étape consiste à télécharger les données et configurer la plateforme avec les propriétés souhaitées pour les molécules générées. Ensuite, lors de la phase de génération, plus de 40 modèles génératifs fonctionnent simultanément pour produire de nouvelles structures. Les structures moléculaires générées sont filtrées puis évaluées de manière dynamique par des modules de récompense et de notation, en fonction de critères prédéfinis. Les scores des structures générées sont renvoyés aux modèles génératifs pour les renforcer et guider le processus génératif vers des structures à score élevé (phase d'apprentissage). Enfin, les structures sont analysées et classées selon diverses métriques, telles que le caractère innovant, l'accessibilité pharmacologique, etc.

¹⁵ Branche de l'IA qui vise à créer de nouvelles données en s'appuyant sur des modèles appris

1.3.4. La réutilisation de médicaments (*drug repurposing*)

La **réutilisation de médicaments** (ou *drug repurposing* en anglais) consiste à utiliser des médicaments **déjà approuvés** ou en cours de développement, dont **l'innocuité a été établie** lors des essais cliniques de phase I, pour traiter une **pathologie différente** de celle pour laquelle il a été initialement développé. L'application de méthodes de calcul avancées et d'algorithmes d'IA pour le criblage rapide de médicaments ayant des profils pharmacocinétiques établis offre une approche intéressante, particulièrement pertinente dans le cas des maladies graves pour lesquels les traitements efficaces sont limités tels que la **maladie d'Alzheimer**.

Ainsi, dans une étude récente, des chercheurs ont utilisé l'**IA** et notamment **ChatGPT** pour identifier les **vingt médicaments** les plus prometteurs pour le traitement de la maladie d'Alzheimer (*Figure 36*). Leurs recherches ont établi que des médicaments tels que la **metformine**, la **simvastatine** et le **losartan** étaient associés à un risque moindre de développer la maladie d'Alzheimer, ouvrant ainsi de nouvelles pistes de recherche dans la recherche de traitements efficaces contre cette pathologie neurodégénérative. (104)

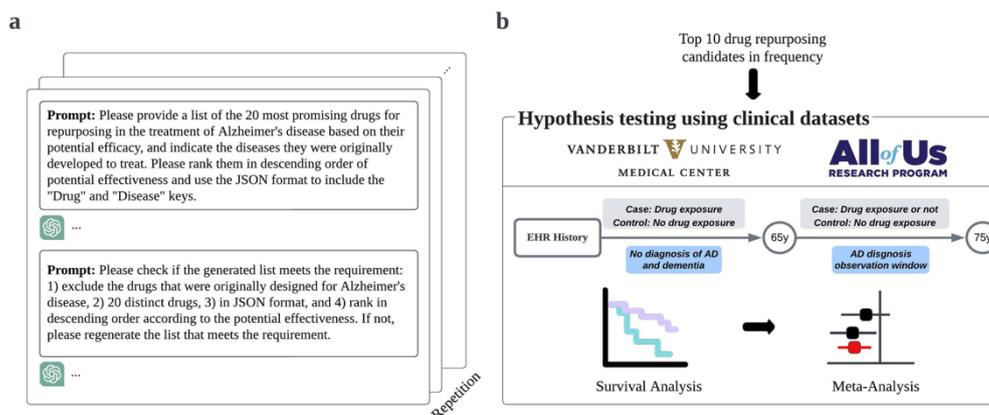


Figure 36 : Illustration de la conception de l'étude utilisant l'IA pour réutiliser / réaffecter des médicaments (104)

a Requête itérative sur ChatGPT pour fournir une liste de 20 médicaments prometteurs pour la maladie d'Alzheimer.
b Évaluation de l'efficacité potentielle des dix médicaments les plus fréquemment suggérés en utilisant les données des dossiers médicaux électroniques de deux grandes bases de données cliniques.

Si l'IA semble être une ressource intéressante pour **prioriser** des candidats à la réaffectation des médicaments et concevoir de nouveaux médicaments **en accélérant l'analyse et la synthèse des connaissances scientifiques**, il est important de souligner que ces découvertes marquent le début d'un voyage scientifique complexe. Bien que les résultats initiaux soient encourageants, une **exploration approfondie** et une **validation rigoureuse** seront nécessaires pour évaluer et exploiter pleinement le potentiel de ces méthodes.

2. IA et essais cliniques

Discipline essentielle du développement de médicaments, la **recherche clinique** vise à évaluer **l'efficacité et la sécurité** des médicaments chez l'être humain. Les essais cliniques se composent de trois phases distinctes. La **phase I** implique des tests sur un petit groupe de volontaires sains pour évaluer la sécurité et la tolérance du médicament. La **phase II** étend les tests à un groupe de patients atteints de la maladie cible afin d'évaluer l'efficacité du médicament à des concentrations différentes et les effets secondaires associés. La **phase III** consiste en des essais sur un plus grand nombre de patients pour confirmer l'efficacité, notamment en comparant avec les traitements de référence et surveiller les EI.

Ces essais cliniques sont souvent coûteux, longs et confrontés à des défis complexes. En effet, **moins d'un tiers** des composés de phase II parviennent à passer en phase III, et **plus d'un tiers** des candidats de phase III ne parviennent pas à obtenir une autorisation de mise sur le marché (AMM). Ces échecs ont un impact économique important, la phase III représentant environ **60 % des coûts** totaux des essais puisqu'elle nécessite de plus grandes cohortes de patients. (105)

L'IA apporte des solutions innovantes à ces défis en utilisant ses algorithmes sophistiqués et ses capacités d'analyses de données pour renforcer **l'efficacité, la sécurité et la précision des essais cliniques**. Dans cette partie, nous aborderons les applications concrètes de l'IA dans les essais cliniques, en commençant par son rôle dans la **sélection et le recrutement des participants**, visant ainsi à optimiser la composition de la cohorte. Ensuite, nous explorerons comment l'IA peut soutenir les chercheurs et les coordinateurs d'essais cliniques dans **le suivi** des patients tout au long de ces essais. Enfin, nous aborderons la manière dont l'IA améliore la **valorisation des données** issues des essais cliniques.

2.1. Optimisation de la sélection et du recrutement des patients participants

Les essais cliniques sont stratégiquement conçus pour cibler des **sous-groupes spécifiques** de patients, afin de **maximiser les chances de démontrer l'efficacité** du traitement testé. L'inscription de **patients inappropriés** peut automatiquement **réduire l'efficacité observée** du médicament testé, augmentant ainsi le risque d'échec de l'essai clinique. Par conséquent, il est impératif de sélectionner scrupuleusement les participants pour **garantir l'efficacité et la validité des résultats**.

Toutefois, la composition de la cohorte peut poser des défis majeurs, l'un des principaux écueils rencontrés dans les essais cliniques résidant dans la **sélection** et le **recrutement** des patients participants. En effet, le recrutement des patients, qui occupe **un tiers de la durée totale** d'un essai, constitue un défi majeur, avec **86 %** des essais ne respectant pas les délais d'inscription et **un tiers** des essais de phase III échouant en raison de problèmes de recrutement. (105)

Ces échecs résultent de deux facteurs principaux. D'une part, les **critères** sont souvent très **restrictifs** : les antécédents médicaux d'un patient peuvent le rendre inéligible, tout comme le fait qu'il ne corresponde pas au profil spécifique requis pour le médicament testé, tel qu'un sous-phénotype particulier ou un stade spécifique de la maladie. D'autre part, la **complexité** des critères d'admissibilité aux essais cliniques, en termes de nombre et de terminologie médicale, constitue souvent un obstacle pour les patients qui ont du mal à évaluer leur propre éligibilité.

L'utilisation de **profils diagnostiques spécifiques**, qui combinent des données provenant de diverses sources comme les **dossiers médicaux électroniques (DME)** et les données **génétiques** permet de sélectionner des patients adéquats pour un essai. L'extraction manuelle d'informations pertinentes à partir de ces données **non structurées** représente une tâche lourde et laborieuse pour les médecins et favorise les biais. Ainsi, pour analyser ces données et prendre en compte de nombreux autres facteurs tels que les antécédents médicaux, les comorbidités et les symptômes de la maladie, des méthodes d'analyse sophistiquées sont indispensables. L'IA, via ses composantes telles que le **ML**, le **DL** et le **NLP**, se révèle comme un outil novateur permettant d'optimiser la sélection des patients en automatisant l'analyse de ces données. (105)

IBM Watson for Clinical Trial Matching (CTM) illustre cette approche en exploitant le **NLP** pour associer les patients aux essais cliniques appropriés. Ce système analyse les données structurées et non structurées des **DME** des patients par rapport aux **critères d'éligibilité du protocole**.

Un projet pilote, mené conjointement par la **Mayo Clinic, IBM Watson Health et Novartis**, a été mis au point pour permettre une évaluation exhaustive des situations cliniques des patients, afin de déterminer leur aptitude à participer à des essais cliniques. Pour ce faire, l'équipe a identifié sept protocoles Novartis en cours ou terminés (quatre pour le cancer du sein et trois pour le cancer du poumon) et recueilli les données des **DME de 376 patients** atteints de cancer du sein et du poumon traités à la Mayo Clinic en 2017.

En tenant compte des critères d'éligibilité des essais, du contexte spécifique du patient (métastatique vs non métastatique), du stade de la tumeur, des marqueurs génétiques et de l'histologie, les infirmières ont passé en revue manuellement les dossiers pour déterminer si les patients étaient potentiellement éligibles ou non pour chaque protocole. Le temps requis pour effectuer cet examen par patient a été chronométré, et la précision prédictive globale, la sensibilité et la spécificité ont été évalués par rapports aux résultats des infirmières. Les résultats ont montré que le logiciel avait une précision globale de **87,6 %** pour l'admissibilité aux protocoles de cancer du sein et de **74,9 %** pour le cancer du poumon. Le temps total nécessaire pour examiner manuellement le cas de 376 patients s'élevait à 441 minutes (7,4 heures). En conclusion, l'utilisation de Watson CTM s'est révélée être une **méthode rapide et efficace** pour **évaluer l'éligibilité des patients** aux essais cliniques, permettant de réduire de **78%** le temps nécessaire pour effectuer le screening des patients et excluant automatiquement **94%** des patients qui ne correspondent pas. (106) (107)

Deep 6 AI est un autre exemple d'utilisation de l'IA pour simplifier la sélection des participants aux essais cliniques, principalement en oncologie. L'approche de Deep 6 AI repose sur une analyse avancée des données médicales non structurées issues des DME, en utilisant des techniques de **ML** et de **NLP**. **92 %** des critères d'inclusion/exclusion reposant sur des **données non structurées**, cette analyse revêt une importance capitale, d'autant plus que **15 à 20 %** des patients éligibles sont uniquement identifiés grâce à des données non structurées. (108) Pour identifier les patients répondant aux critères spécifiques de l'étude, divers paramètres tels que le type de cancer, le stade de la maladie, les antécédents médicaux, les comorbidités et les traitements antérieurs sont pris en compte. En analysant ces données en profondeur, Deep 6 AI peut **générer une liste de participants potentiels** répondant aux exigences de l'essai clinique.

Il est à noter que l'amélioration des performances de ces systèmes d'IA dépendra en grande partie de la **qualité et de la quantité des données disponibles**. De manière plus générale, les techniques d'IA, notamment le NLP, sont envisagées pour extraire de manière proactive des **contenus accessibles au public**, comme les bases de données d'essais numériques, les **annonces d'essais cliniques** ou encore les **réseaux sociaux**. Cela pourrait **automatiser l'identification des correspondances potentielles** entre les essais et les patients, accélérant ainsi la **sensibilisation** des patients aux essais appropriés et facilitant leur **interaction avec les cliniciens** pour évaluer leur admissibilité. L'intégration de l'IA devrait donc considérablement étendre la portée et l'efficacité de ces plans de recrutement numérique.

2.2. Suivi des patients

Pour s'assurer du succès d'un essai clinique, il est impératif non seulement de recruter les bons patients mais également de s'assurer que ces patients **restent engagés, suivent les procédures et respectent les règles** de l'essai. De plus, il est impératif de collecter de manière exhaustive, efficace et fiable toutes les données nécessaires pour **évaluer l'impact** du médicament ou du DM testé. Les désistements dus à un non-respect des protocoles d'essai exigent des **efforts de recrutement supplémentaires**, ce qui entraîne des retards dans le développement des essais. Il est donc crucial de développer des **systèmes de suivi** des patients personnalisés. (105)

Les patients engagés dans un essais doivent tenir des **registres détaillés** de leur prise de médicament et d'autres données liées à leurs fonctions corporelles, à la réponse de leurs organismes aux médicaments, etc. Cette tâche peut s'avérer difficile, entraînant une **non-adhésion moyenne de 40 %** après **150 jours** d'essai clinique. (109). L'utilisation de **capteurs portables** et de la **surveillance vidéo** permet de pallier ce problème en automatisant la collecte de données, réduisant ainsi la charge pour les patients. Les modèles d'IA peuvent ensuite analyser en temps réel ces données pour détecter des événements pertinents. Les données récoltées peuvent également servir de **preuve d'adhésion thérapeutique**, offrant une mesure plus fiable des critères d'évaluation (*endpoints*) que les méthodes d'autosurveillance actuelles.

L'IA a également le potentiel de **prédire** voire **prévenir** le **risque d'abandon** chez un patient donné, en détectant les signes précurseurs du comportement d'abandon du protocole d'étude. En 2018, des algorithmes d'IA ont été utilisés pour **optimiser les schémas posologiques de chimiothérapie** dans le cadre d'essais cliniques sur des patients atteints de tumeurs cérébrales. Ce système de ML a analysé les schémas de traitement en cours et ajusté les doses de manière itérative, aboutissant finalement à un plan thérapeutique optimal visant à **minimiser la puissance et la fréquence des doses** tout en maintenant une **efficacité comparable**. Au cours des essais portant sur **cinquante patients**, les schémas thérapeutiques ont été ajustés avec des doses diminuées **jusqu'à la moitié voire le quart** de leur intensité habituelle, tout en maintenant leur efficacité sur la réduction de la tumeur. Cette approche laisse entrevoir une amélioration potentielle de **l'adhérence des patients** et une **diminution des taux d'abandon**, car les traitements mieux tolérés étant souvent associés à une meilleure observance. L'utilisation de l'IA favoriserait donc une démarche proactive qui adresserait les problèmes sous-jacents avant qu'ils ne conduisent à un abandon de l'essai. (105) (110)

Certaines entreprises ont pleinement saisi le fort potentiel de l'IA pour améliorer le suivi des patients dans les essais cliniques et ont développé des solutions à cet effet. Par exemple, **AiCure** propose une solution basée sur l'IA et la **vision par ordinateur** pour **surveiller à distance** le comportement des patients participant aux essais. Cette approche vise à combler le manque de moyens de vérification de l'observance des patient, les méthodes actuelles se limitant souvent aux retours de blisters et aux dosages biologiques. AiCure utilise des **caméras**, notamment celles intégrées aux smartphones ou aux tablettes, ainsi que des algorithmes avancés pour **suivre en temps réel** les interactions des patients avec les médicaments et les DM utilisés dans les essais cliniques (Figure 37).



Figure 37 : Illustration de l'utilisation d'Aicure (111)

Les caméras peuvent aussi récolter des données importantes, telles que **l'expression faciale** lors de **troubles dépressifs majeurs** ou encore les **tremblements** chez des patients atteints de la **maladie de Parkinson**. Ces données sont instantanément transmises aux entreprises pharmaceutiques conduisant les essais. De plus, AiCure peut alerter les patients sur le point d'oublier leur médicament et ainsi **prévenir la non-observance** non intentionnelle, par le biais de **rappels** et de **notifications automatiques**, qui peuvent être adaptés au mode de vie et aux préférences de communication de chaque participant. En cas de problème, la plateforme peut mettre les patients en contact avec leur prescripteur ou encore informer les investigateurs en cas de non-observance intentionnelle. La plateforme AiCure a déjà démontré des résultats prometteurs dans le suivi du traitement lors d'essais cliniques, notamment en **améliorant l'observance** au traitement pour lutter contre les déficits cognitifs chez les patients atteints de schizophrénie lors d'une **étude de phase II**. Dans cet essai clinique (NCT01655680) sponsorisé par **Abbvie**, l'observance moyenne sur 24 semaines était de **89,7%** pour les sujets suivis à l'aide de la plateforme d'IA **contre 71,9%** pour les sujets suivis de manière plus traditionnelle. (111) A l'avenir, cette plateforme pourrait également suivre l'observance des traitements pour des pathologies chroniques dans la pratique courante. (112)

2.3. Une meilleure valorisation des données d'essais cliniques

Actuellement, la mise sur le marché d'un médicament nécessite en moyenne 12 ans et 2 milliards de dollars, mais l'exploitation efficace des données des essais cliniques pourrait considérablement réduire ces délais et ces coûts.

Les nombreux projets d'IA dans les essais cliniques lancés par les géants pharmaceutiques sont le reflet du potentiel énorme de cette technologie pour accélérer et améliorer ces essais. Par exemple, Novartis, à travers son projet **Data42** (Figure 38), exploite l'IA pour analyser de manière approfondie les données d'essais cliniques accumulées au cours des deux dernières décennies afin d'identifier de nouveaux biomarqueurs dans le cancer du sein ou l'insuffisance cardiaque. (112)



Figure 38 : Configuration de la plateforme Data42 de Novartis (113)

En somme, l'intégration de l'IA dans les essais cliniques présente plusieurs avantages. Tout d'abord, elle optimise la sélection et le recrutement des patients participants en identifiant ceux qui répondent aux critères d'éligibilité spécifiques des protocoles d'étude, améliorant ainsi significativement la pertinence de la cohorte. De plus, l'IA propose des solutions innovantes pour le suivi des patients tout au long des essais cliniques, ce qui favorise une meilleure adhérence au traitement et une réduction des taux d'abandon. En outre, en analysant en profondeur les données d'essais cliniques, l'IA permet une meilleure valorisation de ces données, ouvrant la voie à de nouvelles découvertes et à une compréhension plus approfondie des résultats. En combinant ces différents aspects, l'intégration de l'IA dans les essais cliniques offre des perspectives prometteuses pour l'avancement de la recherche médicale.

3. IA et production pharmaceutique

La production pharmaceutique, qui englobe l'ensemble du processus de fabrication des médicaments, depuis la **synthèse des principes actifs** jusqu'à leur **conditionnement** final, est en pleine mutation grâce à l'intégration de l'IA. Cette transformation s'appuie notamment sur l'utilisation de **capteurs** pour surveiller en **temps réel** la qualité de la fabrication et de conditionnement des médicaments. De plus, des **outils de visualisation** sont instaurés pour simplifier la supervision de la fabrication, contribuant ainsi à une production pharmaceutique plus efficace et plus sûre. On observe une tendance vers l'avènement d'une **production quasiment entièrement automatisée** : une machine pourrait être capable d'analyser un bon de commande numérique, de planifier la production, de commander les matériaux requis, de superviser et de gérer l'ensemble du processus de fabrication, de conditionnement et de stockage, puis d'expédier le produit au client. Dans cette partie, nous examinerons comment l'IA peut contribuer à la surveillance des paramètres de production en temps réel, à l'optimisation des processus de production, et à l'amélioration de la planification de la production.

3.1. Surveillance des paramètres de production en temps réel

L'émergence d'outils innovants tels que les processeurs ultrarapides, combinés aux solutions de stockage de données abordables, positionnent l'IA comme une alliée précieuse pour surveiller en temps réel les processus de fabrication pharmaceutique. Grâce à l'intégration de **capteurs** aux équipements de production, l'IA recueille une multitude de données à chaque étape du processus. Ces données comprennent des informations telles que la **température**, la **pression**, le **débit**, le **pH**, la **qualité de l'eau** (ressource essentielle du processus de production, et notamment lors du nettoyage des équipements) et d'autres variables critiques. Ces données sont ensuite **analysées en permanence** par des algorithmes afin de détecter rapidement toute déviation par rapport aux normes établies, signalant les anomalies potentielles avant qu'elles ne compromettent la qualité du produit final. Cette capacité de **surveillance proactive** offerte par l'IA aide les entreprises pharmaceutiques à maintenir des normes de qualité élevées et à garantir la sécurité des produits pharmaceutiques destinés aux patients.

Par exemple, sous la direction de Vasant Narasimhan, dont l'intérêt pour les technologies est manifeste, l'IA est devenue une priorité majeure chez **Novartis**, touchant toutes les facettes de l'entreprise y compris la production. Sous la supervision de Bertrand Bodson, directeur numérique de Novartis, l'entreprise a instauré un **contrôle en temps réel des paramètres critiques** de la bioproduction, comme la température et le pH sur **trois sites industriels pilote**. Grâce à des logiciels dotés d'IA, ce dispositif détecte et de corrige toute dérive avant qu'elle n'impacte le lot de production. (112)

Sanofi, également engagé dans une transformation numérique profonde, a intégré le numérique et l'IA dans ses usines, comme en témoigne son **site de production de Framingham**. Entièrement **connecté** et **modulable**, ce site n'a aucune ligne de production fixe. Ses installations sont conçues pour s'adapter à chaque type de produits manufacturés, offrant ainsi une **flexibilité opérationnelle maximale**. Sanofi a également mis en place un "**jumeau numérique**" de ce site, réplique exacte de l'usine de 10 000 mètres carrés en trois dimensions. Capitalisant sur les **données collectées** à chaque étape du processus, depuis les lots cliniques jusqu'à la production en masse, le jumeau numérique a deux utilisations principales : la formation des employés et la **prédiction** grâce au monitoring des données en temps réel, qui permet notamment de prévenir les ruptures d'approvisionnement (maintenance prédictive). Ainsi, ce jumeau numérique sert in fine à optimiser la production de biomédicaments. (112)

L'intégration de l'IA améliore donc l'assurance qualité des produits pharmaceutiques en surveillant en temps réel les paramètres de production. Elle permet d'identifier rapidement et précisément les défauts d'emballage ou d'étiquetage, dépassant ainsi le processus traditionnel fondé sur l'inspection visuelle humaine. (114)

3.2. Optimisation des processus de fabrication

L'IA est un outil intéressant pour **optimiser les processus de fabrication** pharmaceutique en identifiant les **conditions opérationnelles optimales** afin de maximiser le rendement et d'assurer une qualité constante. Elle permet d'optimiser divers processus tels que la compression, la granulation ou la synthèse chimique, en éliminant les étapes redondantes et en consolidant les opérations dans des équipements polyvalents. Ces améliorations se traduisent par une réduction des coûts et une augmentation de la qualité des produits.

Amgen, par exemple, intègre des systèmes basés sur l'IA pour améliorer la production de médicaments biologiques. Ainsi, en novembre 2023, l'entreprise a élargi son partenariat avec le géant Amazon dans le but d'améliorer non seulement la découverte mais aussi la production de médicaments. Plus spécifiquement, Amgen intensifie sa collaboration décennale avec **Amazon Web Services (AWS)**, une **plateforme cloud** largement adoptée, pour développer une **IA générative** visant à **accroître le débit** de fabrication des produits pharmaceutiques. Les deux entreprises envisagent d'exploiter cette technologie pour optimiser les opérations et la durabilité de l'usine d'assemblage et d'emballage des produits finis d'Amgen dans le centre de l'Ohio dont l'ouverture des portes a été annoncée en février 2024. Cette nouvelle installation, conçue pour intégrer les dernières technologies numériques et robotiques, inclut une plateforme connectée à AWS via **Amazon SageMaker**, un outil utilisé pour développer, entraîner et déployer des modèles de ML. L'objectif de cette plateforme est, entre autres, de **réduire la dépendance à l'intervention humaine** pendant le processus de fabrication, tout en permettant la **production de rapports** sur les mesures de performance en temps réel. Ces données seront intégrées aux modèles de ML pour prédire et prévenir les défaillances de l'équipement. (115)

3.3. Amélioration de la planification de la production

L'IA joue également un rôle essentiel dans la **planification** de la production pharmaceutique en prenant en compte divers paramètres tels que les délais de livraison, les ressources disponibles, les exigences réglementaires, et les coûts. Un exemple concret de cette utilisation de l'IA est **Rombio**, une solution logicielle avancée basée sur l'IA générant des **plannings** (Figure 39) optimisés et fiables à court, moyen et long terme, en tenant compte des ressources disponibles et des contraintes ainsi que de l'incertitude inhérente.

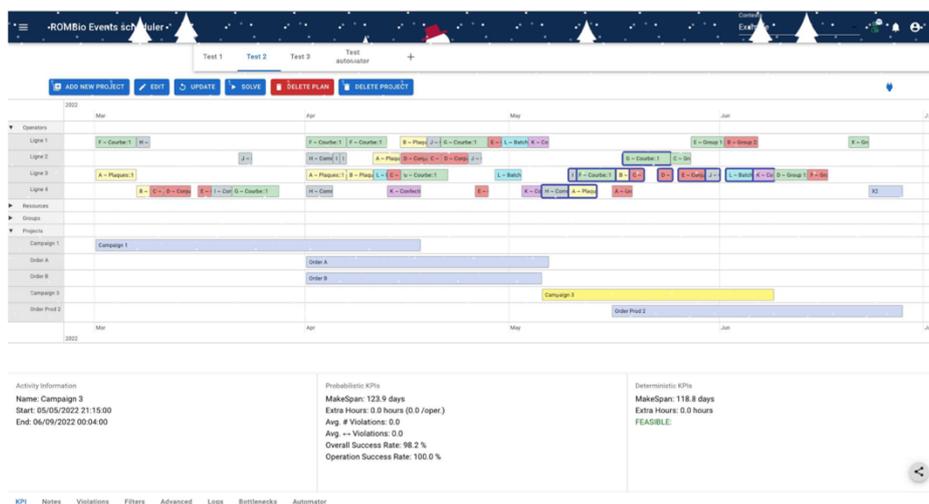


Figure 39 : Interface de planification de Rombio (116)

Fondée sur une technologie initialement développée pour les missions spatiales, Rombio a été étendu à l'industrie pharmaceutique par son créateur, le docteur Michael Saint-Guillain, en raison des similitudes entre ces deux domaines impliquant des processus complexes soumis à des contraintes strictes et coûteuses, avec des ressources souvent limitées et une incertitude omniprésente. La jeune entreprise a rapidement convaincu plusieurs laboratoires pharmaceutiques de l'efficacité de son approche. Rombio est notamment déjà intégré chez le géant **Takeda**. Après seulement cinq mois d'utilisation, Takeda a observé une **augmentation moyenne d'un à deux lots supplémentaires par mois**, représentant une avancée significative, pour Takeda, dont le système de production avait déjà été optimisé. L'objectif de Rombio est d'aider Takeda à accroître leur production de **5 %** avec les mêmes ressources, ce qui peut se traduire par **plusieurs dizaines de millions de bénéfices nets par an**. D'autres clients ont par la suite également fait confiance à Rombio, dont **Baxter** et **Exothera**. (116)

4. IA et logistique pharmaceutique

La logistique pharmaceutique vise à assurer un **approvisionnement sûr et efficace** en médicaments, depuis leur fabrication jusqu'à leur distribution aux patients. Cette chaîne logistique complexe rencontre divers défis, tels que la **gestion des stocks**, la **traçabilité** des produits et la **conformité** aux normes de sécurité. De plus, elle doit faire face aux **fluctuations** de la demande, aux **contraintes réglementaires** et aux défis liés à la **conservation** des médicaments. L'IA offre une solution prometteuse pour remédier à ces défis, en améliorant la chaîne logistique, en prévenant les ruptures d'approvisionnement et en luttant contre médicaments contrefaits.

4.1. Une chaîne logistique plus performante

De nombreuses entreprises pharmaceutiques explorent des solutions basées sur l'IA pour optimiser leur logistique et prévoir les perturbations potentielles dans leur chaîne d'approvisionnement. C'est notamment le cas du géant pharmaceutique **Merck**, qui, depuis 2015, a intégré l'IA dans ses opérations pour optimiser **l'analyse de la circulation de ses produits à travers son réseau logistique complexe**. Merck a fait appel aux logiciels développés par **Aera Technology** et **Tracelink** pour analyser attentivement les multiples flux de données afin d'anticiper les éventuels incidents et d'assister les équipes dans la mise en œuvre d'actions correctives. (117)

Offrant une vision plus claire du parcours des produits à travers **un vaste réseau de fournisseurs et de partenaires**, l'intégration de l'IA dans la chaîne logistique de Merck représente une étape cruciale dans sa transformation digitale, passant d'une approche réactive à une **approche proactive** basée sur la **prévision** et la prescription de **mesures correctives**. Cette évolution offre ainsi des opportunités significatives en termes de **rentabilité**. À titre indicatif, pour des entreprises telles que Merck, une amélioration de seulement un pour cent de l'efficacité de la chaîne logistique peut se traduire par des bénéfices nets supplémentaires se chiffrant en **millions de dollars**.

Dans le processus d'intégration de l'IA chez Merck, les logiciels d'Aera Technology jouent un rôle central. Ils analysent régulièrement plus de **42 systèmes ERP**¹⁶ (Enterprise Resource Planning) agrégeant ainsi les données issues de différentes sources. En identifiant les **goulots d'étranglement** (*bottlenecks*), le logiciel génère des recommandations basées sur des indicateurs impossibles à suivre manuellement en raison du volume de données. Ces recommandations aident l'entreprise à prendre des décisions éclairées sur l'ajustement des niveaux de stock ou la modulation de la production, grâce à un tableau de bord intégré. Ainsi, les logiciels d'Aera technology ont **réduit excédents de stocks**, favorisant une **livraison plus ponctuelle et plus complète** des produits pharmaceutiques.

De plus, Merck tire parti du logiciel d'IA de TraceLink pour optimiser le **dernier kilomètre de la distribution pharmaceutique**. Le concept de dernier kilomètre dans l'industrie pharmaceutique fait référence à la **phase finale de la distribution des médicaments**, couvrant leur livraison depuis les centres de distribution jusqu'aux points de vente, tels que les pharmacies, les hôpitaux et les patients. Cette utilisation innovante de l'IA renforce la capacité de Merck à maintenir une chaîne logistique fluide et efficace, tout en garantissant la qualité et la fiabilité de ses produits pharmaceutiques.

L'utilisation de l'IA dans la logistique pharmaceutique par les géants de l'industrie ainsi que cette transition vers une *supply chain* intelligente et proactive illustrent la montée en puissance de l'IA dans le secteur pharmaceutique et ses impacts tangibles sur la **performance opérationnelle** et la **compétitivité** des entreprises du domaine. (118)

¹⁶ Système informatisé qui intègre et gère l'ensemble des processus opérationnels et fonctionnels d'une entreprise, tels que la production, les achats, la logistique, la comptabilité et les ressources humaines.

4.2. Lutte contre les ruptures d'approvisionnement

Les récentes **ruptures de stocks** de médicaments comme le **Doliprane** ont mis en évidence l'importance de **l'optimisation des niveaux de stock** afin d'assurer un approvisionnement continu et de répondre efficacement aux besoins des patients. Selon l'ANSM (Agence Nationale de Sécurité du Médicament et des produits de santé), les ruptures d'approvisionnement ont atteint un niveau sans précédent. En 2018, pas moins de **868** médicaments ont été signalés en **tension d'approvisionnement** ou en **rupture de stock**, un chiffre **multiplié par vingt en vingt ans**, qui concerne uniquement les médicaments d'intérêt thérapeutique majeur (notamment anti-infectieux ou relatifs au système nerveux). Cette situation soulève des inquiétudes encore plus importantes pour l'ensemble des produits commercialisés, d'autant plus que la situation n'a pas évolué favorablement depuis lors (*Figure 40*). Ces pénuries ont un impact significatif sur le traitement des patients concernés ; **45%** des français ayant vu leur traitement modifié en raison de cette situation, dont **14%** ayant présenté une augmentation des symptômes. (119)

Un enjeu de santé publique majeur en France

Source : ANSM

1 français sur 4

s'est déjà vu refuser un médicament pour cause de rupture de stocks dans les pharmacies

3 000

pénuries ou risques de pénuries de médicaments enregistrés en 2022

+39%

d'augmentation par rapport à 2021 avec 1100 types de médicaments concernés

Figure 40 : Infographie sur les pénuries de médicaments en 2022 (120)

Le LEEM (Les Entreprises du Médicament) identifie deux causes principales à l'augmentation des ruptures de médicaments : la **forte demande mondiale** qui met sous pression les approvisionnements en matières premières (principes actifs et excipients) et les capacités de production ainsi que la **délocalisation significative de la production** de l'Europe vers des pays d'Asie tels la Chine et l'Inde. A titre illustratif, environ **80 %** des fabricants de substances actives sont situés en dehors de l'UE d'après un rapport du Sénat sur la pénurie de médicaments et les choix de l'industrie pharmaceutique française. (121)

Cependant, une partie significative des ruptures pourrait être évitée grâce à une **gestion prévisionnelle adéquate** des besoins et des ordres de production et/ou de distribution. C'est sur cette proportion que la maîtrise des données et l'IA pourraient offrir des avantages significatifs. L'IA peut optimiser les niveaux de stock en créant des itinéraires de livraison efficaces, et en réduisant les risques de détérioration des médicaments dus à des conditions de transport inadéquates. En exploitant les historiques de ventes, les tendances saisonnières, les cycles de production, et les délais de livraison des fournisseurs, ainsi que les facteurs externes tels que les crises sanitaires ou les événements climatiques, l'IA élabore des **modèles de prévision de la demande**. Ces modèles sont continuellement ajustés en fonction des données en temps réel, assurant ainsi une **planification de stock agile et adaptative** qui minimise les coûts de stockage tout en garantissant une disponibilité adéquate des médicaments.

Des entreprises telles qu'**Apodis Pharma** ont déjà amorcé l'exploration de solutions innovantes et illustrent la réalité de ces avancées. En effet, Apodis a mis au point une plateforme numérique qui permet la **synchronisation en temps réel des données provenant des officines françaises**. Les algorithmes détectent les problèmes dans les minutes qui suivent leur apparition, offrant ainsi une **réactivité** sans précédent. Grâce à une connexion sécurisée aux bases de données des pharmacies, la plateforme analyse les données et identifie en temps réel les ruptures d'approvisionnement d'un médicament. L'information est instantanément transmise au laboratoire concerné pour que le grossiste-répartiteur nécessitant un réapprovisionnement soit livré en priorité afin de réassortir rapidement les pharmacies de sa zone géographique. En cas d'extrême urgence, Apodis envoie directement une proposition de livraison sur le smartphone du titulaire de l'officine directement depuis le laboratoire pour répondre aux besoins immédiats de réassort. Apodis Pharma affirme que sa plateforme permet une **diminution de 30% des ruptures logistiques** enregistrées dès les premières semaines d'utilisation. (120)

4.3. Lutte contre les médicaments contrefaits

Les Centers for Disease Control and Prevention (CDC) des États-Unis ont signalé une augmentation des décès liés à des surdoses causées par des médicaments contrefaits depuis 2019. Cette tendance est alarmante et met en lumière un problème croissant de **contrefaçon et d'abus de médicaments**. (122)

Selon l'Organisation mondiale de la Santé (OMS), les produits médicaux "**falsifiés**" sont ceux *“dont l'identité, la composition ou la source sont représentées de manière trompeuse, que ce soit délibérément ou frauduleusement”*. (123) Le fléau des médicaments contrefaits constitue un défi contemporain qui pèse sur la santé publique à l'échelle mondiale, impactant tous les groupes d'âge et tous les domaines thérapeutiques, y compris les médicaments humains, les vaccins et les médicaments vétérinaires. Les **pays à faible et moyen revenu** sont encore plus à risque puisque **un produit médical sur dix** y serait de qualité inférieure ou falsifié. Ces chiffres soulignent l'urgence d'une action coordonnée à l'échelle mondiale pour lutter contre ce fléau. Cette activité, encore peu sanctionnée, génère d'énormes profits pour les organisations criminelles impliquées. En effet, le trafic de produits de santé falsifiés serait dix à vingt fois plus rentable que le trafic d'héroïne. (123)

Les produits médicaux falsifiés peuvent compromettre l'efficacité d'un traitement, engendrer des effets indésirables, voire mettre la vie des patients en danger et représentent donc une menace sérieuse pour la santé publique. Ainsi, une **gestion logistique efficace** et une **traçabilité précise** sont essentielles pour lutter contre ces médicaments contrefaits. En suivant le cheminement des médicaments depuis leur fabrication jusqu'à leur distribution, on réduit les risques de contrefaçon et d'altération. Une parfaite traçabilité permet également d'identifier rapidement les produits falsifiés et de les retirer du marché, préservant ainsi la santé des patients. Dans cette lutte, l'IA et d'autres technologies récentes telles que les **datamatrix** et la **blockchain** peuvent renforcer la traçabilité des médicaments et détecter les contrefaçons plus efficacement.

Le **datamatrix 2D** est un code-barre bidimensionnel obligatoire qui contient des informations essentielles sur le médicament : **code CIP, numéro de lot, numéro d'AMM et date de péremption** (Figure 41). Ce système de codage permet de **tracer** et d'**authentifier** les médicaments tout au long de leur parcours dans la chaîne d'approvisionnement, depuis leur fabrication jusqu'à leur délivrance aux patients. Ainsi, en cas de rappel de produits ou de suspicion de contrefaçon, le datamatrix facilite l'identification et le retrait des médicaments défectueux du marché. Depuis la directive européenne entrée en application le 9 février 2019, plus aucune boîte de médicament délivrée sur ordonnance ne devrait circuler sur le territoire européen sans être sérialisée. (124)



Figure 41 : Exemple de boîte de médicaments avec un datamatrix (124)

L'utilisation conjointe de l'IA et de la **blockchain** peut protéger les entreprises et les consommateurs contre la perte ou la divulgation de données tout au long du trajet des produits. L'intégration de ces outils numériques, et notamment le **ML** vient ainsi compléter les réglementations en vigueur dans le secteur pharmaceutique et contribue à préserver la sécurité des patients en luttant contre l'introduction de médicaments contrefaits sur le marché. Les entreprises pharmaceutiques ont donc un rôle proactif à jouer en adoptant des technologies anti-contrefaçon adaptées et novatrice pour protéger leurs marques et la santé publique.

Les pays du Sahel (tels que le Burkina Faso, le Tchad, le Mali, la Mauritanie et le Niger) dépendent largement des produits médicaux importés, car leurs industries pharmaceutiques en sont encore aux premiers stades de développement. De plus, la forte prévalence de maladies infectieuses, y compris le paludisme, combiné aux problèmes d'accessibilité et de disponibilité des traitements crée un environnement où la demande de médicaments n'est pas entièrement satisfaite par les canaux officiels.

L'écart entre l'offre et la demande favorisant le trafic, l'Office des Nations Unies contre la Drogue et le Crime (ONUDC) estime que **50 % des médicaments en Afrique de l'Ouest sont soit contrefaits, soit de mauvaise qualité**. Ainsi, les startups africaines spécialisées dans les technologies de la santé jouent un rôle actif dans la lutte contre ce fléau.

Par exemple, l'entreprise nigériane **RxAll** exploite l'IA pour détecter les contrefaçons pharmaceutiques. Son dispositif, le **RxScanner**, est un authenticateur portable (*Figure 42*) permettant aux patients de vérifier la fiabilité de leurs médicaments. En seulement **20 secondes**, le RxScanner évalue la qualité des médicaments sur ordonnance et affiche les résultats (authentique ou non) en temps réel via des applications mobiles, en les comparant à une vaste base de données de médicaments authentiques. (125) En tant que premier appareil au monde à utiliser l'IA pour la vérification de la qualité des médicaments, le RxScanner est déployé dans **15 pays** à travers le monde et a testé avec succès plus de **300 types de médicaments**, avec une précision remarquable de **99,9%**. Les **grossistes** peuvent exploiter le potentiel du RxScanner pour confirmer la qualité des nouveaux lots de médicaments reçus de leurs fournisseurs. De même, les **hôpitaux** peuvent l'utiliser pour détecter les contrefaçons et garantir la délivrance de médicaments authentiques. Ainsi, en garantissant l'authenticité et la sécurité des médicaments, RxScanner assure une chaîne d'approvisionnement fiable.



Figure 42 : Photographie du RxScanner (126)

Fruit de plus de huit ans de recherche à Yale aux États-Unis, le RxScanner peut identifier les médicaments à travers leur emballage, évitant ainsi **tout risque d'exposition, de manipulation ou de destruction de l'échantillon**. Les résultats des scans d'échantillons de médicaments sont enregistrés et stockés sur la **blockchain**, offrant ainsi une traçabilité totale et immuable des médicaments tout au long de la chaîne d'approvisionnement. Cette technologie met en lumière la manière dont l'IA ouvre de nouvelles perspectives pour garantir la sécurité et la fiabilité des médicaments à travers le monde. (126)

5. IA et pharmacovigilance

Les efforts pour garantir l'innocuité des médicaments débutés dès les essais précliniques et cliniques se prolongent par une surveillance post-commercialisation des effets indésirables au sein des populations réelles. L'IA joue un rôle croissant dans cette phase, en permettant une analyse plus rapide et approfondie des données pour identifier les tendances et les signaux potentiellement préoccupants. (Figure 43)

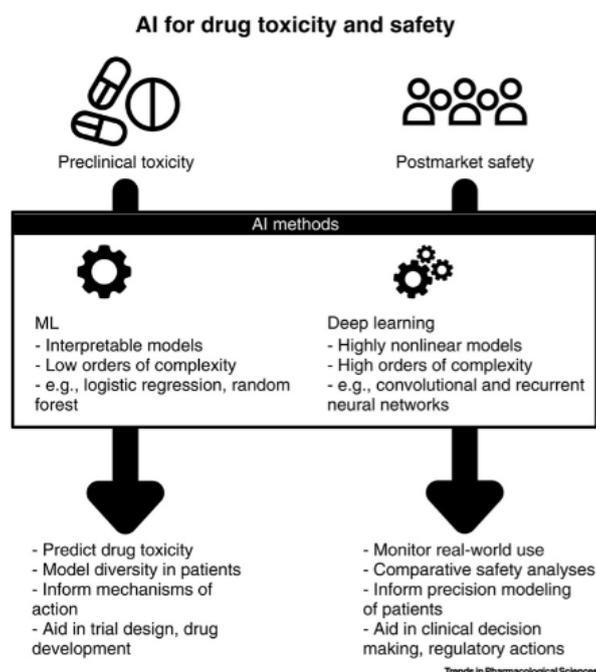


Figure 43 : Exemples d'utilisation de l'IA pour évaluer la toxicité en préclinique et la sécurité post AMM (91)

Les avancées en IA offrent des opportunités prometteuses pour renforcer la sécurité des médicaments, avant et après leur mise sur le marché. Ces algorithmes permettent une approche basée sur les données pour évaluer l'innocuité, identifiant des schémas qui pourraient passer inaperçus. Les méthodes de ML, telles que la régression logistique, les forêts aléatoires et les machines à vecteurs de support, permettent de déterminer l'impact de certaines variables sur la fréquence ou le risque d'un événement indésirable. D'autre part, le DL et ses réseaux neuronaux profonds, bien que nécessitant une quantité substantielle de données, peuvent détecter automatiquement des schémas non linéaires sans intervention humaine, ce qui les rendent particulièrement adaptés pour les applications post-AMM, où un plus grand nombre de patients est impliqué, fournissant ainsi des données plus massives.

Cette surveillance post-commercialisation, connue sous le nom de **pharmacovigilance (PV)**, implique la collecte et l'évaluation scientifique des EI afin de prévenir les risques pour la santé publique. Actuellement, les programmes de PV reposent principalement sur les **signalements spontanés** d'EI émanant des **professionnels de santé** ou des **patients**, suivis d'enquêtes "manuelles" menées par les organismes de pharmacovigilance. Cependant, ces rapports spontanés sont sujets à des biais tels que la **sous-déclaration**, notamment pour les événements rares et les interactions médicamenteuses. (127)

Le mésusage des médicaments, qui entraîne chaque année 130 000 hospitalisations évitables et 30 000 décès en France, ainsi que l'augmentation de la **polypharmacie** et de l'**automédication** viennent accentuer les limites de ces méthodes traditionnelles de surveillance. Parallèlement, l'abondance croissante de données offre une opportunité d'intégrer l'IA pour renforcer la PV en automatisant certaines parties du processus.

Medication Shield, une solution d'IA dédiée à la **gestion des déclarations d'EI** illustre parfaitement l'apport des technologies dans le domaine de la PV. Développée par la start-up française **Synapse Medicine** en collaboration avec l'**ANSM** ainsi que les centres régionaux de pharmacovigilance (**CRPV**) de Bordeaux et Marseille, cette technologie a fait l'objet de plusieurs études de validation. Medication Shield gère quasiment en **temps réel** les déclarations d'EI faites par le grand public sur le portail de signalement en ligne. Grâce à ces algorithmes de **NLP**, Medication Shield extrait automatiquement les termes utilisés lors de la déclaration d'un EI et les code selon la classification internationale **MedDRA** (Figure 44). Cette classification standardisée est utilisée pour encoder les EI dans les bases de données de PV, l'uniformisation de la terminologie facilitant l'échange et l'analyse de données. Ensuite, Medication Shield **classe et priorise** les EI en fonction de leur **gravité** ainsi que leur **fréquence d'apparition** telle que déterminée dans la monographie du médicament concerné afin de déterminer si ces effets sont **attendus ou non**. Grâce à ces avancées, Medication Shield a permis de réduire le temps passé par les experts de PV sur chaque rapport d'EI de **40 à 2 minutes**, marquant ainsi une amélioration considérable de l'efficacité dans leur travail. (128)

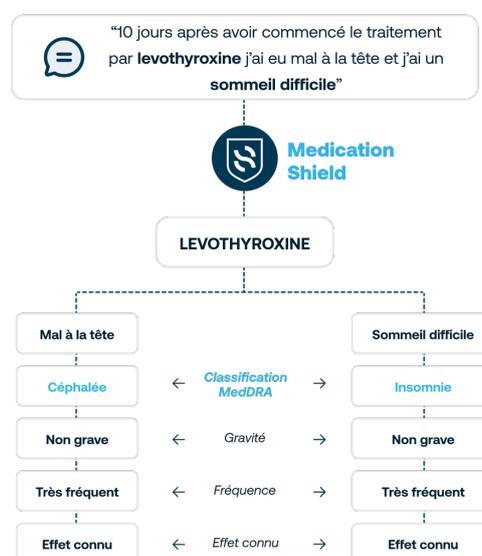


Figure 44 : Illustration du fonctionnement de Medication Shield (129)

En somme, l'application de l'IA à la PV offre une approche **plus proactive** et **efficace** pour la détection, l'évaluation et la gestion des EI des médicaments. Bien que l'évaluation de la toxicité et de l'innocuité des médicaments assistée par l'IA soit un domaine en pleine expansion présentant un potentiel clinique important, des recherches approfondies sont encore nécessaires pour évaluer pleinement son potentiel et ses impacts. (91)

6. IA et promotion pharmaceutique

La promotion pharmaceutique désigne les activités de communication et de marketing par les entreprises pharmaceutiques pour **promouvoir leurs produits médicaux** auprès des professionnels de la santé, des patients et du grand public. Cela comprend la publicité directe aux consommateurs, les visites des représentants médicaux aux professionnels de la santé, les campagnes de sensibilisation, les programmes éducatifs et autres initiatives visant à informer et à influencer les décisions de prescription et d'achat de médicaments.

À l'instar des cookies Internet, le secteur du marketing bénéficie déjà en partie des outils de l'IA. Ces technologies permettent de **profiler** chaque utilisateur, **adaptant** ainsi les publicités selon les préférences du public pour maximiser les ventes. En exploitant de grandes quantités de données, il devient possible d'analyser le parcours du consommateur et les facteurs influençant ses choix d'achat.

Dans le cas des industries de santé, l'IA s'avère particulièrement précieuse dans les secteurs des **médicaments en vente libre (ou *Over-The-Counter* OTC)** et des **produits de santé grand public** en raison de la nécessité de toucher un large public de consommateurs sans l'intermédiaire direct d'un professionnel de santé pour prescrire ces produits. Cependant, son utilité ne se limite pas à ces domaines spécifiques ; elle englobe également tous les produits de santé, qu'ils soient destinés à un usage courant ou prescrits par un professionnel de santé. Ainsi, l'IA occupe une place de plus en plus importante dans la commercialisation des médicaments, notamment dans la recherche de marché et le marketing ciblé grâce à ses capacités d'analyse de données.

En utilisant l'IA, les entreprises pharmaceutiques peuvent **anticiper la demande, optimiser les campagnes publicitaires et personnaliser les offres** en fonction des besoins individuels des consommateurs. De plus, l'IA surveille activement les **réseaux sociaux** et les **forums en ligne** pour détecter rapidement les **tendances émergentes** et répondre aux préoccupations des consommateurs, contribuant ainsi à accroître la **satisfaction** des clients et à maintenir la **compétitivité** sur le marché des produits de santé.

Aktana, une entreprise américaine, utilise l'IA pour renforcer la collaboration entre les entreprises pharmaceutiques et les prestataires de soins de santé tels que les hôpitaux (*Figure 45*). Parmi les clients d'Aktana figure le géant Novartis, qui a eu recours à l'IA d'Aktana pour **soutenir ses équipes commerciales**. En exploitant les bases de données, cette IA permet à Novartis de **mieux comprendre les habitudes de prescription des médecins**, afin d'optimiser l'approche de leurs visiteurs médicaux et leurs interactions avec ces professionnels de santé. (112)

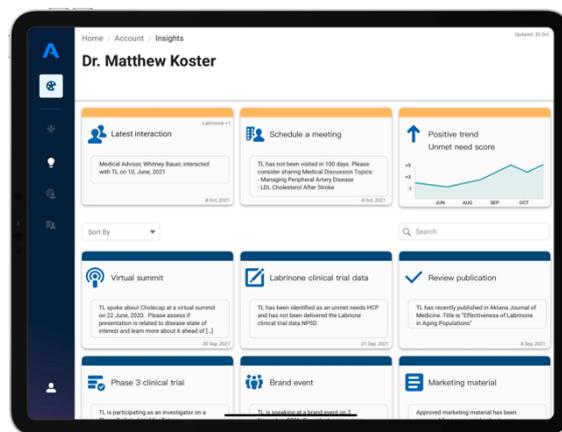


Figure 45 : Exemple de tableau de bord de la plateforme Aktana (130)

En mars 2024, Aktana a lancé « **Aktana for MedTech** », une suite d'outils destinés à optimiser **les activités des représentants sur le terrain, la coordination des comptes et les programmes de formation médicale** pour les DM grâce à des informations et des conseils enrichis par l'IA. Cette initiative répond à la demande croissante des prestataires de soins de santé qui expriment clairement leur intérêt à exploiter divers canaux numériques pour mener leurs propres recherches, interagir avec leurs pairs et échanger avec les représentants des entreprises de technologie médicale.

Aktana for MedTech permet ainsi aux entreprises de DM de planifier et de gérer efficacement leurs interactions avec les prestataires de soins grâce à des informations optimisées par l'IA. Aktana for MedTech offre une solution complète, comprenant **Aktana Copilot Mobile**, un assistant numérique interactif qui optimise les stratégies de vente et de marketing des équipes commerciales sur le terrain via des appareils mobiles. Aktana for MedTech facilite la **planification des appels**, la **gestion des stocks** et **prédit les seuils** de produit en identifiant les changements critiques de volume, ce qui contribue à augmenter les ventes et à réduire les coûts. Enfin, en fournissant des suggestions dynamiques pour une **formation médicale efficace**, Aktana for MedTech garantit que les prestataires de soins disposent des informations nécessaires au bon moment pour mieux répondre aux besoins des patients. (130)

Cette approche démontre comment l'IA peut optimiser les stratégies de commercialisation dans l'industrie des produits de santé en identifiant de nouvelles opportunités de marché, personnalisant les approches marketing et anticipant les tendances émergentes.

Les divers exemples abordés dans cette section de la thèse sur les applications de l'IA dans le cycle du médicament démontrent clairement que l'IA révolutionne ce cycle à tous les niveaux, de la recherche préclinique à la pharmacovigilance, garantissant une qualité, une sécurité et une efficacité inégalées à chaque étape. Alors que l'IA joue un rôle de plus en plus important dans le développement pharmaceutique, elle se positionne également comme un pilier essentiel de la médecine du futur. La suite de notre exploration se concentrera sur les impacts concrets de l'IA sur la façon dont patients et professionnels de santé appréhendent la médecine d'aujourd'hui et de demain.

II. L'IA, pilier de la médecine du futur

La médecine du futur, connue sous le nom de **médecine 4P** (Prédictive, Préventive, Personnalisée et Participative), transforme fondamentalement **la prestation de soins de santé** en intégrant des approches innovantes pour améliorer la santé individuelle et collective. Cette approche holistique représente une véritable révolution dans le paysage médical actuel et se déploie sur plusieurs fronts. Elle **anticipe** les problèmes de santé grâce à une approche prédictive qui permet des interventions précoces et ciblées. Son aspect préventif vise à identifier les **facteurs de risque individuels** et encourager **des modes de vie sains**. De plus, cette médecine du futur, personnalisée, **adapte** les traitements médicamenteux et chirurgicaux en fonction des **caractéristiques individuelles** de chaque patient, optimisant ainsi l'efficacité thérapeutique tout en limitant les EI. Enfin, elle **engage activement les patients** dans leur propre prise en charge grâce à des applications de santé mobiles, des dispositifs de surveillance personnelle et une communication renforcée avec les professionnels de la santé.

Cette section met en lumière le rôle fondamental de l'IA en tant que pilier de la médecine de demain. Nous examinerons dans un premier temps les apports de l'IA pour permettre des diagnostics plus précis et plus précoces. Puis, dans un second temps, nous nous pencherons sur la façon dont l'IA favorise une médecine personnalisée et des traitements sur mesure. Nous analyserons ensuite comment l'IA permet une approche anticipative en prédisant les risques pour mieux prévenir les maladies. Enfin, nous verrons la manière dont l'IA rend la médecine plus participative, plaçant le patient au centre des décisions médicales le concernant.

1. Des diagnostics plus précis et plus précoces

En analysant de vastes ensembles de données médicales, génétiques et comportementales avec **une rapidité et une précision inégalée**, l'IA permet d'obtenir des diagnostics plus précis et plus précoces. Les systèmes d'IA peuvent détecter des schémas subtils et des corrélations dans les données, facilitant ainsi la prédiction précoce de maladies. Cette capacité à identifier des signes avant-coureurs avant même qu'ils ne se manifestent cliniquement offre aux professionnels de santé la possibilité d'intervenir de manière proactive et d'instaurer des traitements plus ciblés.

En 2019, **DeepMind** a développé un modèle d'IA capable de **détecter les signes précoces** d'une **insuffisance rénale aiguë (IRA)** jusqu'à 48 heures avant son apparition. La capacité à prédire précocement les IRA pourrait jouer un rôle important dans l'assistance au personnel soignant, étant donné que **11 %** des décès à l'hôpital sont imputables à l'incapacité de reconnaître et de traiter rapidement les patients dont l'état se détériore. (131) **Basée sur le DL**, la solution développée par DeepMind s'appuie sur une modélisation intégrant les événements indésirables extraits des DME. Le modèle a été entraîné à partir d'une vaste base de données incluant les DME de 703 782 patients dans 172 sites hospitaliers et 1 062 sites ambulatoires. Les résultats des recherches menées au Royaume-Uni et aux États-Unis sont extrêmement prometteurs, le modèle de Deepmind ayant prédit **90,2% des cas d'IRA nécessitant une dialyse ultérieure**. Cette avancée a considérablement amélioré la prise en charge des patients (*Figure 46*), permettant une **intervention dans les quinze minutes**, contre plusieurs heures sans l'application. L'IA a ainsi conduit à un affinement des diagnostics, réduisant le taux d'erreurs (cas critiques non détectés) de **12,4 % à 3,3 %**. (132)

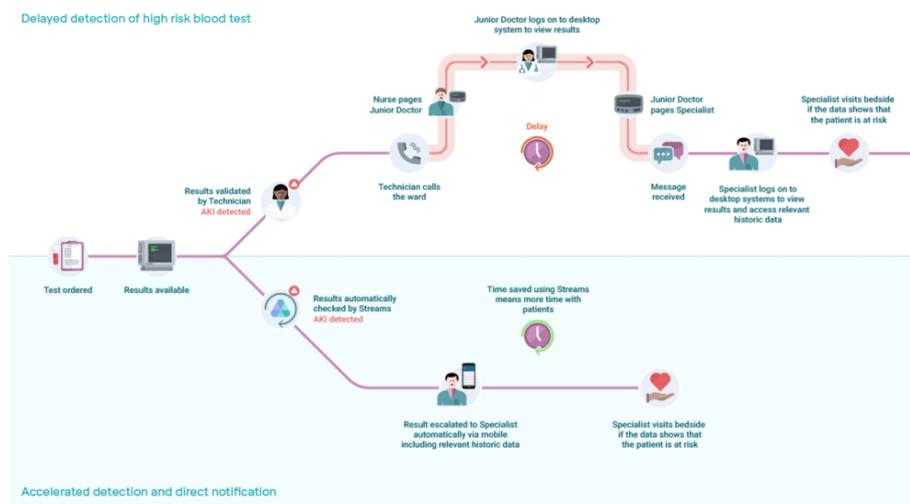


Figure 46 : Parcours de soin avec et sans l'IA de Deepmind. AKI : Acute Kidney Injury (IRA) (133)

L'IA joue également un rôle important dans l'amélioration des diagnostics **en oncologie** à partir des **données d'imagerie**. Par exemple, un algorithme de prédiction du **cancer du sein** combinant des approches de ML et de DL a été développé et évalué. L'algorithme a été entraîné sur un ensemble de données comprenant 38 444 images de mammographies provenant de 9611 femmes. Cet algorithme intègre pour la première fois des **données d'imagerie** avec des informations de **dossiers médicaux** associées. Dans une étude rétrospective, 52 936 images ont été collectées chez 13 234 femmes ayant subi au moins une mammographie entre 2013 et 2017.

L'algorithme a démontré son aptitude à prédire la **malignité des biopsies** et à différencier les résultats de dépistage normaux et anormaux. Son application permet une **évaluation du cancer du sein comparable à celle des radiologues** (avec une spécificité de 77,3% et une sensibilité de 87%), tout en ayant le potentiel de considérablement **réduire les diagnostics manqués de cancer du sein**, 48% des faux négatifs ayant été identifiés par l'algorithme. Ce modèle pourrait donc servir de **support aux radiologues** dans la détection du cancer du sein et représente ainsi une aide précieuse en tant que deuxième lecteur. (134) Ainsi, l'IA constitue un outil précieux pour les professionnels de santé, comme l'a souligné le docteur Curtis Langlotz, professeur de radiologie à Stanford, qui a affirmé que "les radiologues qui utilisent l'IA remplaceront ceux qui ne le font pas". (135)

2. Une prise en charge thérapeutique personnalisée

La médecine du futur est une **médecine personnalisée**, également connue sous le nom de **médecine de précision**, qui vise à permettre de donner le « bon médicament à la bonne dose pour le bon patient ». (136)

Cette approche aspire à individualiser les soins de santé en tenant compte des **particularités individuelles** de chaque patient. Plutôt que de suivre une approche standardisée, elle s'appuie sur des données précises telles que le **patrimoine génétique, les biomarqueurs biologiques et le mode de vie**. La médecine personnalisée permet par exemple **d'identifier les phénotypes** des patients qui réagissent de façon moins fréquente aux traitements conventionnels. Grâce aux progrès technologiques tels que la génomique et l'IA, cette approche permet une compréhension plus approfondie des mécanismes des maladies, favorisant ainsi des diagnostics plus précis et une sélection de traitements mieux adaptés.

L'IA joue un rôle crucial dans cette démarche (*Figure 47*) en permettant une analyse exhaustive des caractéristiques génétiques, en facilitant la détection précoce des risques de maladies et en optimisant les traitements. Grâce à des approches telles que la pharmacogénomique et la modélisation des réponses individuelles, elle offre une médecine personnalisée où chaque décision thérapeutique est éclairée par une compréhension précise des mécanismes biologiques propres à chaque individu. Ainsi, cette approche axée sur la prise en charge personnalisée présente des avantages significatifs en termes **d'efficacité thérapeutique**, de **prévention** des maladies et **d'amélioration de la qualité de vie** des patients. (137)

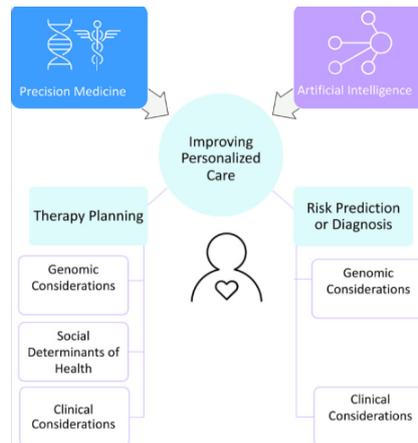


Figure 47 : Synergies entre l'IA et la médecine de précision (137)

La convergence de la médecine de précision et des techniques d'IA révolutionne la personnalisation des soins de santé. D'une part, elle permet la planification thérapeutique en se basant sur des facteurs cliniques, génomiques, sociaux et comportementaux. D'autre part, elle facilite la prédiction des risques et le diagnostic en exploitant des données génomiques et cliniques.

La médecine de précision requiert un accès étendu à des données essentielles collectées par des initiatives telles que les **biobanques** ou encore le **projet All of Us** aux Etats-Unis. Les biobanques, instaurées en 2001 par l'OCDE (Organisation de Coopération et de Développement Économiques), sont des structures clés dans la gestion des ressources biologiques à des fins de recherche. Organisées en **réseaux nationaux et internationaux**, elles se subdivisent en biobanques **populationnelles et thématiques** par maladie. Devenues des piliers de la recherche biomédicale, elles sont indispensables pour l'analyse des maladies et le développement de la médecine personnalisée. Ainsi, des initiatives telles que **UK Biobank** au Royaume-Uni et **CONSTANCES** en France facilitent la **collecte, le stockage et le partage d'échantillons et de données** pour des études épidémiologiques et cliniques.

UK Biobank est une étude de grande envergure menée au Royaume-Uni depuis 2006 qui vise à examiner les **interactions entre la prédisposition génétique et l'exposition environnementale** dans le développement des maladies (*Figure 48*). CONSTANCES, quant à elle, se positionne comme un outil de recherche épidémiologique majeur en France, suivant de près **220 000 volontaires** par le biais de questionnaires annuels, d'examens de santé tous les quatre ans, et en les reliant régulièrement aux bases de données administratives. En outre, elle établit une biobanque de 80 000 volontaires, offrant une représentation précise d'une population de 400 000 personnes. (138) Ces structures jouent un rôle crucial dans **l'identification de biomarqueurs et la prédiction de la réponse aux traitements.**

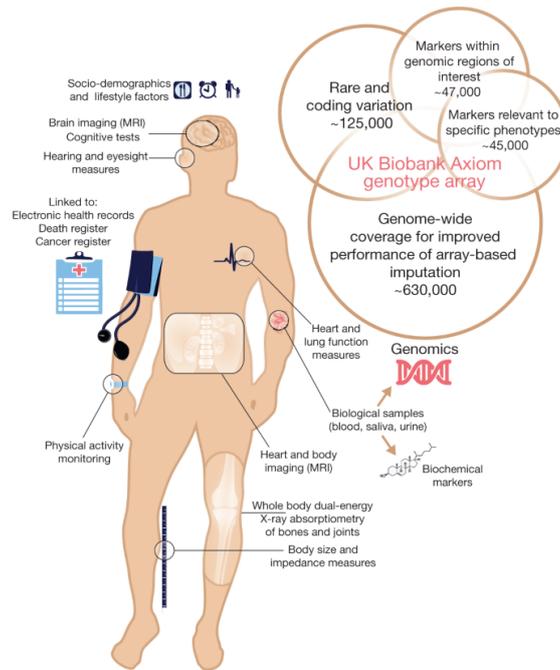


Figure 48 : Résumé des principaux éléments de la biobanque britannique (139)

A l'échelle européenne, des programmes tels que **BBMRI (Biobanking and Biomolecular Resources Research Infrastructure)** encouragent l'émergence de centres d'expertise garantissant des échantillons de qualité supérieure pour la recherche. BBMRI est une infrastructure mise en place pour faciliter la recherche biomédicale en Europe en promouvant la **collaboration entre les biobanques**, en améliorant l'accès aux ressources biologiques et aux données associées et en harmonisant les pratiques de *biobanking* à travers le continent. (140)

Les Etats-Unis prennent également part à l'essor de la collecte de données de santé, notamment à travers le programme de recherche **All of Us**, une initiative ambitieuse visant à rassembler des données de santé auprès **d'un million de personnes** ou plus vivant aux États-Unis.

Une vaste réserve de données inexploitées demeure en dehors des systèmes médicaux conventionnels, malgré le potentiel significatif qu'elles offrent. Cette réserve englobe notamment de données sur les **modes de vie, la nutrition, l'environnement et l'accès aux soins** qui peuvent fortement impacter le niveau de santé d'un individu. Grâce aux progrès technologiques, il est désormais possible de surveiller et d'évaluer efficacement ces aspects comportementaux et sociaux, notamment à l'aide de **dispositifs portables**.

Avec l'essor croissant des **montres connectées**, des **dispositifs médicaux connectés**, des **implants** et autres **capteurs** qui recueillent des données en temps réel, chaque individu génère désormais plus **d'un million de gigaoctets** de données liées à sa santé au cours de sa vie, ce qui équivaut à plus de plus de **300 millions de livres**. (141) Ces variables socio-économiques et comportementales jouent un rôle prépondérant, représentant environ **60% des déterminants de santé** tandis que les gènes ne contribuent que pour environ 30%. En comparaison, l'historique médical concret quant à lui ne représente qu'une proportion de 10%. (142)

Ainsi, la médecine personnalisée repose sur une **approche intégrée**, combinant l'analyse de données multi-omiques, les antécédents médicaux, les facteurs sociaux et comportementaux ainsi que les contextes environnementaux. Cette approche vise à définir de manière précise les états de santé, les conditions pathologiques et les options thérapeutiques adaptées à chaque individu. L'IA s'inscrit parfaitement dans cette approche, en permettant une analyse détaillée de ces données complexes afin de fournir des recommandations thérapeutiques personnalisées.

Dans cette perspective, l'exploitation de l'IA représente un véritable tournant dans la **lutte contre le cancer**, comme en témoignent plusieurs études récentes de premier plan. Par exemple, dans le cas du **médulloblastome**, l'analyse assistée par l'IA de centaines d'**exomes** a permis de définir des **sous-groupes moléculaires** distincts de la maladie, facilitant ainsi la **prescription précise du traitement, à la dose adéquate, pour la bonne cohorte de patients**. Alors que le traitement conventionnel de cette maladie nécessitait une **approche multimodale** incluant chirurgie, chimiothérapie et irradiation cérébrale totale, la médecine de précision basée sur la génomique a permis de traiter un sous-groupe de tumeur particulier « *wingless* », plus fréquent chez les enfants, uniquement avec de la chimiothérapie. (143) Cette éviction de la radiothérapie revêt une importance particulière pour réduire les risques de séquelles neurocognitives et de cancers secondaires chez les survivants de la maladie. (144)

D'autre part, la **radiogénomique**, issue des avancées de l'IA dans la reconnaissance d'images, établit des **liens entre les caractéristiques d'imagerie du cancer et l'expression génique** (Figure 49). Par exemple, des **réseaux de neurones convolutifs** ont permis de prédire de manière non invasive le génotype de l'isocitrate déshydrogénase, marqueur important dans le **gliome** lié au **pronostic** et à la **réponse aux traitements** à partir de données d'Imagerie par Résonance Magnétique (IRM). (145) (146)

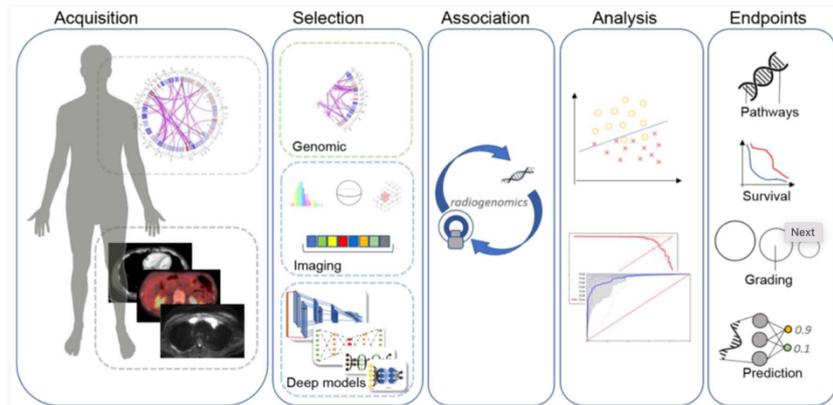


Figure 49 : Illustration du processus de radiogénomique (145)

Le processus inclut la collecte de données génomiques et d'imageries, l'extraction de caractéristiques à partir de ces données, notamment grâce au DL et la sélection des caractéristiques les plus importantes via des analyses statistiques. L'association et la modélisation de ces biomarqueurs hautement discriminants permettent d'améliorer la fiabilité des outils d'aide à la décision.

En outre, le ML est un outil précieux pour établir des **modèles prédictifs de réponse aux médicaments** en se basant sur les profils génomiques des cellules ciblées. Par exemple, une étude a utilisé un algorithme basé sur une **SVM** pour prédire les réponses de patients atteints de cancer à **diverses chimiothérapies** standards à partir des profils d'expression génomique de leurs tumeurs, obtenant une précision de prédiction de plus de **80%** (Figure 50). (147)

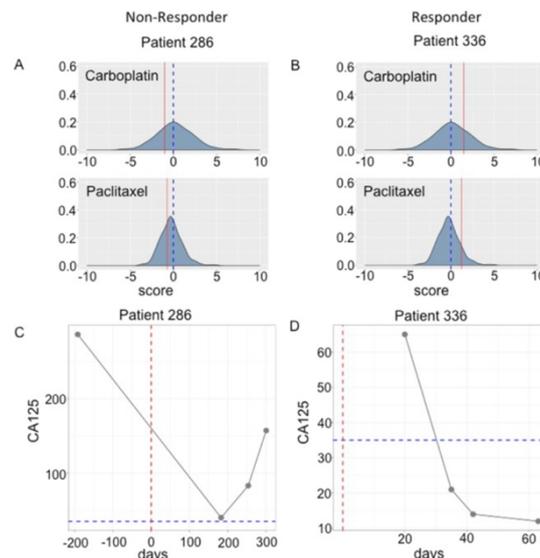


Figure 50 : Comparaison des réponses prédites par l'algorithme et observées de deux patientes atteintes d'un cancer de l'ovaire aux traitements par carboplatine et paclitaxel (147)

Les scores de réponse prédits pour chaque patiente sont indiqués par la ligne rouge, superposée à la distribution des scores prédits pour 273 patientes atteintes d'un cancer de l'ovaire. Par exemple, la patiente 286 (A) devrait ne pas répondre aux deux médicaments (scores négatifs), tandis que la patiente 336 (B) devrait répondre aux deux. Pour déterminer la réponse au traitement, les valeurs de CA-125 pour chaque patiente sont comparées aux valeurs normales (<35, ligne bleue en pointillés). Ainsi, la patiente 286 (C) ne répond pas au traitement, tandis que la patiente 336 (D) y répond.

Ces avancées démontrent le potentiel de l'IA dans la personnalisation des traitements contre le cancer, ouvrant ainsi la voie à des stratégies thérapeutiques plus efficaces et ciblées.

3. Une médecine plus proactive : Prédire pour mieux prévenir

La médecine évolue progressivement vers une approche proactive, cherchant à **anticiper** et à **prévenir** les maladies avant qu'elles ne se manifestent pleinement. L'intégration de l'IA dans les industries des produits de santé ouvre la voie à cette transition. Grâce à des algorithmes sophistiqués et à l'analyse de vastes ensembles de données médicales, l'IA permet aux professionnels de la santé **d'identifier les facteurs de risque**, de **prédire les éventuelles complications** et de **recommander des interventions précoces** pour prévenir les maladies, offrant ainsi de nouvelles opportunités pour améliorer la santé publique et individuelle.

Les pathologies complexes telles que les **maladies cardiovasculaires** résultent souvent de l'interaction entre divers facteurs tels que le **genre**, la **génétique**, le **mode de vie** et l'**environnement**, nécessitant la prise en compte de la diversité et de la complexité des données. Les méthodes d'IA, qui excellent dans la découverte de relations complexes parmi un grand nombre de facteurs, ouvrent ainsi des perspectives prometteuses.

Une étude américaine (148) a inclus des patients de l'essai clinique randomisé NCT00005487 ou Multi-Ethnic Study of Atherosclerosis (MESA) pour étudier la progression des maladies subcliniques vers des événements cardiovasculaires sur des patients initialement exempts de maladie cardiovasculaire. L'objectif principal de cette étude était d'évaluer la capacité des techniques de ML à **prédire des événements cardiovasculaires sur des périodes relativement longues**. En comparant l'efficacité des approches de ML aux méthodes traditionnelles telles que le modèle de Cox-PHM et aux scores de risque standards, cette étude visait également à identifier les **facteurs prédictifs significatifs** associés à six résultats cliniques cardiovasculaires : décès, accident vasculaire cérébral (AVC), maladie coronarienne, maladies cardiovasculaires, insuffisance cardiaque et fibrillation auriculaire.

MESA est une étude de **cohorte prospective** qui a observé une population de 6 814 individus, âgés de 45 à 84 ans, représentant quatre groupes ethniques différents et ne présentant pas de maladie cardiovasculaire lors de leur inclusion dans l'étude entre 2000 et 2002. En utilisant une variété de mesures de base comprenant des tests d'imagerie, des questionnaires et des biomarqueurs, ils ont appliqué la technique des **forêts aléatoires** pour identifier les 20 meilleurs prédicteurs de chaque résultat parmi les 735 variables recueillies. Ces mesures de base ont été utilisées pour prédire les résultats cardiovasculaires sur une période de suivi de **12 ans**.

Les résultats de cette étude ont clairement démontré que l'approche basée sur les forêts aléatoires **surpasse** les scores de risque standard en termes de **précision de prédiction** (Figure 51), mettant en évidence le potentiel de l'IA pour améliorer la prévision des événements cardiovasculaires chez les populations initialement asymptomatiques.

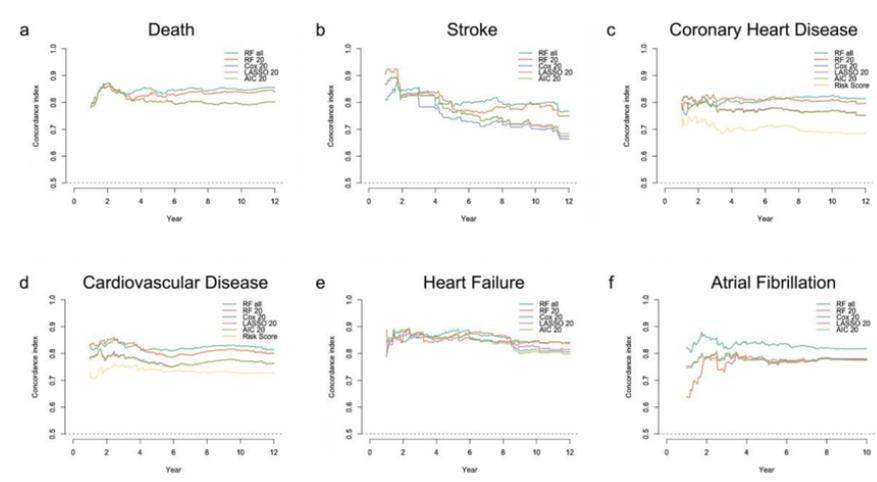


Figure 51 : Indice de concordance pour chacun des modèles testés au cours des années (148)

On observe que les méthodes de forêts aléatoires démontrent une capacité de prédiction supérieure à celle des méthodes traditionnelles telles que les modèles de Cox standard, Lasso-Cox et AIC-Cox et scores de risque standard, pour prédire les six résultats cliniques cardiovasculaires étudiés

Les analyses ont également révélé que **l'imagerie, l'électrocardiographie** et les **biomarqueurs sériques** étaient parmi les facteurs les plus significatifs, surpassant les facteurs de risque cardiovasculaire traditionnels. L'âge est apparu comme le premier prédicteur pour la mortalité toutes causes confondues, tandis que les niveaux de glucose à jeun et les mesures d'échographie carotidienne se sont avérés être des prédicteurs importants pour prédire les AVC.

Dans la même veine, une étude américaine menée en 2019 a utilisé des modèles de ML et de DL pour prédire les événements cardiovasculaires sur une période de **10 ans** en utilisant des DME et des données **génétiques**, sur une cohorte de 109 490 individus. Après avoir exploré l'efficacité de différents modèles, y compris les **forêts aléatoires**, les CNN et les **réseaux neuronaux récurrents** avec des unités à mémoire à court terme (LSTM)¹⁷, les chercheurs ont comparé les performances avec les approches actuellement utilisées dans la pratique clinique courante : l'équation de risque de la cohorte combinée de l'American College of Cardiology et de l'American Heart Association (*ACC/AHA pooled cohort risk equation*).

¹⁷ Type de réseau de neurones récurrents spécialisés dans le traitement de séquences de données, capables de conserver et de manipuler des informations sur de longues périodes de temps.

En incorporant des caractéristiques génétiques, les chercheurs ont constaté une **amélioration significative de la prédiction des événements cardiovasculaires** par rapport aux approches conventionnelles, soulignant l'importance de l'intégration de ce type de données dans la prédiction des risques cardiovasculaires. (149)

Dans un contexte plus large, l'IA se révèle être une méthode prometteuse pour **stratifier les risques** au sein de vastes populations. Les méthodes d'IA offrent une approche statistique puissante pour analyser les données biomédicales en constante augmentation en termes de volume, de vitesse et de variété, surmontant ainsi les défis liés à la dimensionnalité des données. Contrairement aux approches traditionnelles, ces méthodes ne requièrent **pas nécessairement d'hypothèses préalables** sur la causalité, ce qui les rend adaptées à l'évaluation du rôle des **nouveaux biomarqueurs** dans la prédiction des maladies. (148)

4. Une médecine plus participative

L'avènement des **dispositifs connectés** et la **facilité croissante d'accès aux données** médicales via internet, les groupes de soutien de patients, les forums ou encore la littérature médicale accessible au grand public accordent à l'individu une position centrale dans le système de santé.

Ces avancées favorisent le développement d'une **médecine participative** et encouragent les patients à prendre un **rôle plus actif** dans la gestion de leur santé. Les informations générées par ces nouveaux outils de santé numérique, complémentaires à celles provenant du secteur médical traditionnel, peuvent jouer un rôle essentiel. Dans les années à venir, les patients seront probablement amenés à participer plus activement à leur propre prise en charge, aux côtés des médecins et autres professionnels de santé. (150) Cette **autonomisation** et cette **responsabilisation** croissante des patients leur permettant de surveiller leur état de santé et d'engager des interactions avec le système de santé est accompagnée de profondes mutations dans les méthodes de soins.

D'une part, les patients deviennent de plus en plus actifs dans la gestion de leur santé, bénéficiant **d'une meilleure compréhension** de leur condition médicale et des options thérapeutiques disponibles. Cette évolution remet en question la dynamique traditionnelle de la **relation médecin-patient**, avec une volonté croissante des patients de voir leur opinion prise en considération. (151)

D'autre part, l'introduction de thérapies novatrices et de DM connectés entraîne une **décentralisation** de certaines formes de prise en charge vers **le domicile** des patients, reflétant ainsi une **tendance ambulatoire** exacerbée par le vieillissement de la population et la prévalence croissante des maladies chroniques. (151)

Cette autonomisation du patient repose sur **l'éducation thérapeutique**, dont l'objectif, tel que défini par l'OMS en 1996, est *“d' aider les patients à acquérir ou maintenir les compétences dont ils ont besoin pour gérer au mieux leur vie avec une maladie chronique”*(152). L'éducation thérapeutique, qui implique le médecin et le pharmacien, rend les patients conscients de leur maladie, des soins, des procédures hospitalières et des comportements liés à la santé, afin d'assumer leur propre prise en charge et améliorer leur qualité de vie.

La prolifération d'**applications** et de **dispositifs connectés** sur le marché ces dernières années accompagne les patients dans leur quête d'autonomie. Agissant telles des **conseillers personnalisés**, ces solutions fournissent en **temps réel** des données et des recommandations adaptées, anticipant les fluctuations des paramètres biologiques.

Parmi les maladies chroniques en expansion, le **diabète** se distingue particulièrement, avec une prévalence en constante augmentation (*Figure 52*). Selon les projections de l'Assurance Maladie, si les tendances observées entre 2015 et 2021 se poursuivent, **520 000** personnes supplémentaires pourraient être atteintes par le diabète en 2027 par rapport à 2021, dont 500 000 par un diabète de type 2 et 20 000 par un diabète de type 1. (153)

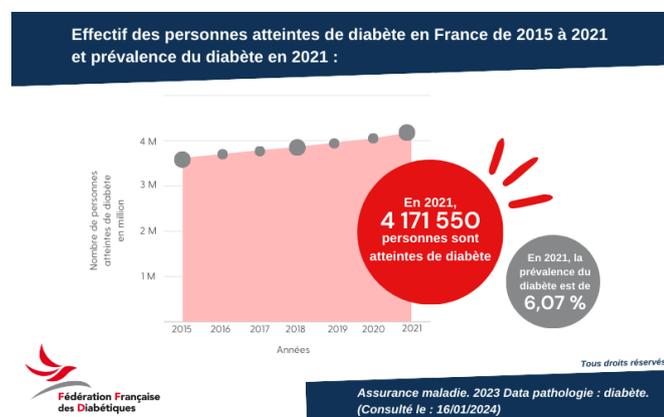


Figure 52 : Augmentation des patients atteints de diabètes en France de 2015 à 2021 (153)

Devant cette réalité, un grand nombre d'acteurs ont émergé sur le marché de **la gestion du diabète**, témoignant d'un développement fulgurant de l'écosystème entourant cette pathologie. Une multitude **d'applications** ont été conçues pour surveiller et anticiper les fluctuations de la glycémie. À titre d'exemple, voici trois entreprises proposant des solutions d'accompagnement basées sur l'IA pour les patients diabétiques.

L'entreprise **PK Vitality**, spécialisée dans les biotechnologies portables, a développé la **K'Watch Glucose**, une **montre connectée** qui permet de **surveiller la glycémie** en temps réel sans prélèvement sanguin. Dotée de micro-capteurs et biocapteurs, elle mesure la glycémie à travers la peau sans douleur, offrant ainsi un suivi discret et continu. Les données enregistrées sont analysées par un algorithme de **ML** qui **détecte les variations anormales** et **envoie des alertes** discrètes en cas de glycémie trop élevée ou trop basse. Cette technologie permet aux patients de reprendre le contrôle de leur glycémie et de leur santé, avec la possibilité de partager les alertes avec leurs proches. La K'Watch agit comme un véritable coach de vie en suivant le nombre de pas, la distance parcourue et les calories brûlées tout au long de la journée. Elle permet ainsi de surveiller les progrès du patient et de l'encourager à atteindre ses objectifs. Avec une interface simple et claire (*Figure 53*), elle facilite la gestion d'une pathologie complexe en réduisant le nombre de gestes quotidiens nécessaires. Les alertes intégrées guident l'utilisateur dans ses décisions et lui fournissent un soutien lors des crises, ce qui **réduit le sentiment d'isolement**. De plus, elle assure la tranquillité d'esprit des proches en les informant en cas de situation à risque non résolue, notamment pour les personnes âgées ou les étudiants vivant seuls.

(154)



Figure 53 : Illustration de la K'watch Glucose et de son application (155)

La start-up **Hillo** se concentre également sur l'**autonomisation du patient diabétique** via une **application mobile** qui collecte les données liées à la gestion du diabète, telles que la glycémie, les repas, les activités physiques et les injections d'insuline pour permettre au patient un suivi continu. Grâce à l'IA et après seulement deux semaines d'utilisation, l'application peut **prédire les variations glycémiques futures** du patient avec une grande précision (*Figure 54*), permettant ainsi au patient d'anticiper les situations à risque (hypo et hyperglycémies) et d'adapter son traitement en conséquence. (156)

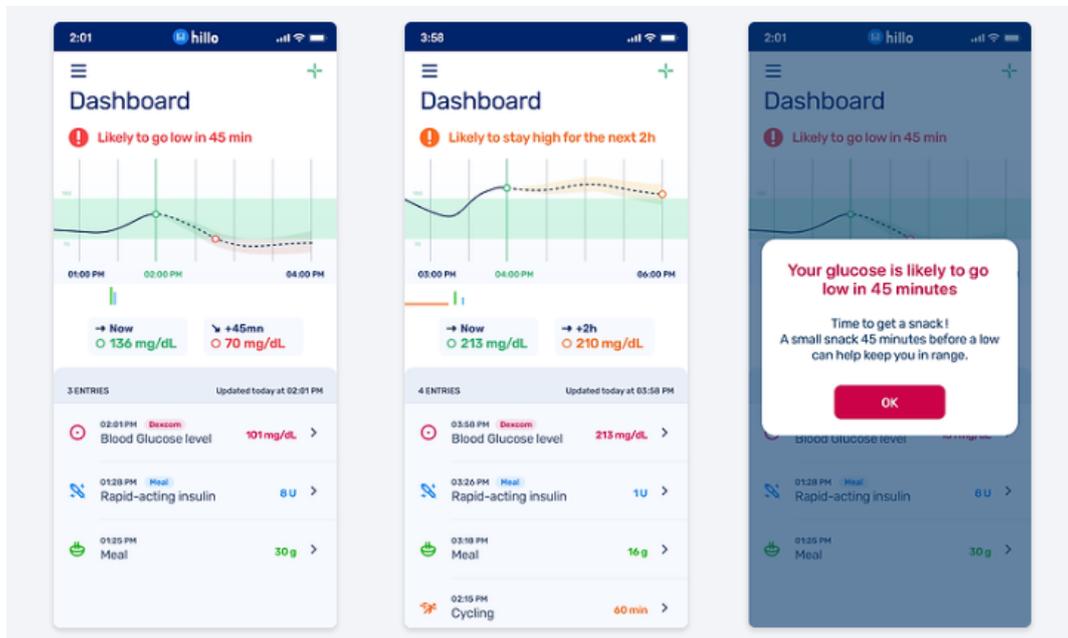


Figure 54 : Exemples de capture d'écran de l'application Hillo (156)

En parallèle, le développement des **stylos à insuline connectés** prend une importance croissante, à l'image de **Mallya**, un produit phare de la biotech auvergnate **Biocorp**, récemment rachetée par Novo Nordisk. En tant que DM intelligent, Mallya est actuellement le seul de sa catégorie marqué **classe IIb** conformément à la norme CE. Amovible et compatible avec la plupart des stylos à insuline (Flexpen, Solostar, Kwickpen, Flextouch...), Mallya rend ces stylos connectés (*Figure 55*).



Figure 55 : Photographie de Mallya installé sur un stylo Solostar (157)

Mallya collecte automatiquement les données sur chaque injection (date, heure, dose) et les transmet en temps réel à l'application **Gluci-check** de Roche. Mallya est simple à utiliser et ne modifie pas la manière dont le stylo est utilisé. Avec une recharge mensuelle nécessaire pour une utilisation quotidienne de 3 à 4 injections, il offre une facilité d'utilisation remarquable. L'application associée permet un accès facile aux informations de chaque injection, illustrant ainsi les avantages de l'IA pour une médecine plus participative, dans le diabète mais aussi dans d'autres aires thérapeutiques telles que les **troubles de la croissance**. (157)

Toutes ces solutions convergent vers l'**autonomisation** du patient, promouvant l'idée que le patient doit être le moteur de sa propre santé, en prenant une part active dans les décisions qui le concernent. Cette autonomisation et cette responsabilisation renforcées par l'IA replace le patient au cœur du processus de soins, mettant ainsi l'accent sur sa personne plutôt que sur le problème de santé en lui-même en améliorant sa qualité de vie.

En somme, l'intégration de l'IA dans le domaine médical éclaire un horizon prometteur où les diagnostics deviennent plus précis et plus précoces, les traitements personnalisés, les soins proactifs et les patients pleinement impliqués dans leur prise en charge. Cette évolution vers une médecine 4P, propulsée par l'IA, redéfinit les standards de santé et promet des avancées significatives pour la santé humaine. Cependant, ces avancées sont accompagnées de défis complexes et de questions éthiques soulevés par l'IA, notamment autour de confidentialité des données, la responsabilité en cas d'erreurs et l'équité dans l'accès aux soins. Dans la prochaine partie de cette thèse, nous explorerons ces défis et proposerons des pistes de réflexion et des stratégies pour les surmonter, assurant ainsi que l'IA continue de bénéficier à l'ensemble de la société à long terme.

PARTIE 3 : Les principaux enjeux et défis liés à l'intelligence artificielle dans l'industrie des produits de santé

Comme illustré tout au long de la partie précédente, l'IA offre un potentiel considérable dans l'industrie des produits de santé, avec une large gamme d'applications extrêmement prometteuses. Cependant, son adoption soulève également de nombreux **défis et enjeux**. Sans prétendre à l'exhaustivité, cette section vise à offrir une vue d'ensemble des diverses questions, enjeux et défis soulevés par l'évolution rapide de l'IA dans le domaine de la santé. Nous débuterons en examinant les aspects éthiques, suivi des aspects réglementaires et juridiques, puis nous aborderons les considérations techniques, économiques, organisationnelles, sociétales, et enfin environnementales.

I. Les enjeux éthiques

Bien que l'IA offre des avantages indéniables pour la santé publique en fournissant des données améliorées, une modélisation et une surveillance en temps réel, ses **implications éthiques** demeurent cruciales. Ces préoccupations, liées aux principes éthiques fondamentaux et à la nécessité d'une supervision humaine dans les interactions entre l'IA et les individus, soulèvent des questions essentielles en matière de confiance et nécessitent une réflexion approfondie. Dans ce volet portant sur les enjeux éthiques entourant l'utilisation de l'IA dans l'industrie des produits de santé, nous examinerons plus précisément les questions de la protection des données personnelles, l'opacité des décisions de l'IA, les biais algorithmiques potentiels et leur impact sur l'équité, l'utilisation non éthique des systèmes d'IA, ainsi que les enjeux éthiques liés au développement d'agents moraux artificiels et au pouvoir des entreprises commercialisant des systèmes d'IA.

1. Protection des données personnelles et de la vie privée des patients

L'intégration de l'IA dans le secteur de la santé soulève des préoccupations majeures concernant la **confidentialité des données médicales** des patients. Ces données, d'une extrême sensibilité, requièrent une protection rigoureuse contre tout accès non autorisé ou toute utilisation abusive. Or, comme précédemment discuté, l'efficacité des technologies d'IA repose sur l'analyse de vastes ensembles de données, ce qui en fait des ressources **hautement convoitées** par divers acteurs, tant publics que privés, poursuivant des objectifs divers.

Les défis éthiques associés à l'utilisation de l'IA dans le domaine de la santé sont étroitement liés à la nature même de ces données, ainsi qu'à leur **collecte**, leur **stockage**, leur **traitement**, leur **utilisation**, leur **partage** et leur **archivage** (Figure 56). Dans cette section, nous aborderons les principaux enjeux éthiques liés à l'exploitation des données massives et de l'IA dans le domaine de la santé.

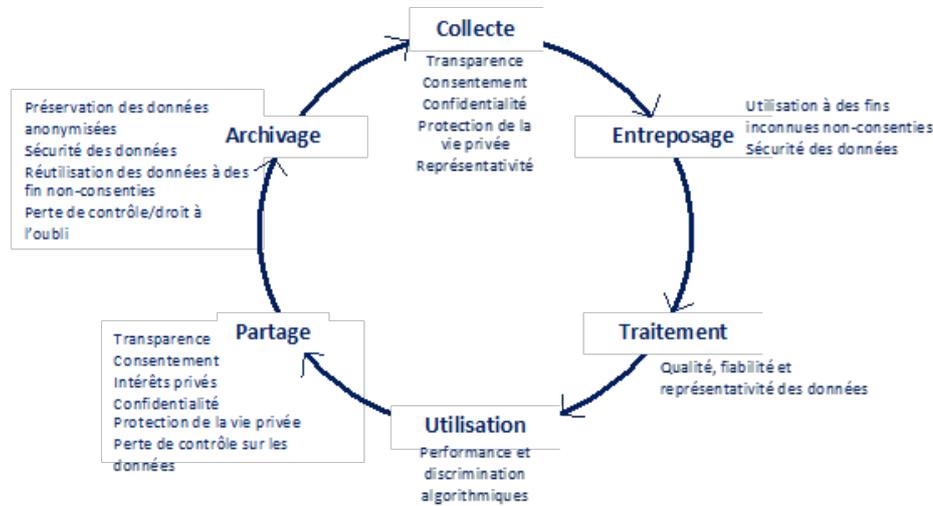


Figure 56 : Enjeux éthiques et cycle de vie des données (158)

1.1. Enjeux liés au consentement à la collecte et au partage des données

Les enjeux éthiques de l'IA en santé découlent largement de la **collecte des données personnelles**, qui vont au-delà des données biomédicales et cliniques pour inclure un large éventail d'informations sur le mode de vie quotidien des individus telles que les données de nutrition, de sommeil et d'exercice physique. Ces données, provenant d'objets connectés comme les smartphones et les montres connectées, sont converties en données "médicales" et analysées pour obtenir des informations sur la santé globale des personnes (159).

La multiplication des sources de données implique une **diversification des acteurs** impliqués dans leur collecte, y compris des organisations publiques de santé et des entreprises privées. Cependant, les individus sont souvent **peu informés** sur la **quantité** et **l'utilisation réelle** de leurs données, ce qui suscite des inquiétudes concernant la confidentialité et la possibilité de **vente de données**. En effet, les diverses entités impliquées dans le développement des technologies d'IA en santé ont tendance à recueillir un volume de données souvent plus important que nécessaire, parfois avec l'intention de les exploiter à des **fins commerciales ultérieures**, sans que les individus dont les données sont collectées ne soient pleinement conscients de ces pratiques. (160)

Malgré les progrès réalisés dans la réglementation entourant l'obtention du consentement pour la collecte et le partage des données personnelles, le processus de consentement associé aux applications de santé et aux objets connectés est souvent **rapide** voire, dans certains cas, **trompeur**. En pratique, les utilisateurs se retrouvent souvent à cocher une case dans une application ou sur un site web sans réellement comprendre les implications de leur consentement, ni avoir eu une explication claire et compréhensible des termes du contrat auquel ils adhèrent en cochant cette case. Lorsqu'ils tentent de lire les formulaires de consentement et les contrats de confidentialité des données personnelles des entreprises privées proposant des objets connectés ou des applications d'IA, ils se retrouvent souvent confrontés à un **langage juridique complexe**, rendant ces documents longs et difficiles à comprendre pour la plupart des utilisateurs. (160) (161) Ces pratiques opaques de collecte et de partage des données remettent en question le **consentement libre et éclairé** et **restreignent l'autonomie individuelle** (162)

1.2. Enjeux liés à l'archivage et à la conservation des données

Les données personnelles, une fois **anonymisées**, échappent souvent aux règles juridiques de protection des données personnelles (puisque plus considérées comme telles) et peuvent être utilisées à d'autres fins que celles initialement consenties. (163) Cependant, ce processus soulève des préoccupations éthiques, notamment en raison du **risque de réidentification** des données médicales, ce qui devient problématique lorsque ces informations sensibles se retrouvent entre les mains de personnes non autorisées. (164) En effet, malgré les méthodes de protection mises en place par les organisations telles que la **désidentification**, la **pseudonymisation** et l'**anonymisation**, les risques de réidentification demeurent, mettant en péril la confidentialité des informations. Par exemple, malgré l'anonymisation, il a été démontré qu'il suffit de **quatre données anonymisées** provenant de différentes sources pour **réidentifier** un individu avec une grande précision. (159)

Nos données personnelles constituent ce qui est parfois appelé notre "**profil digital**" (165) représentant ce que les acteurs du milieu de la santé savent de nous. Pour les personnes vivant avec des conditions de santé spécifiques, parfois **stigmatisantes**, l'utilisation de leurs données personnelles peut devenir particulièrement contraignante. En effet, le fait que leurs données personnelles continuent d'être utilisées et stockées peut entraver leur **liberté** et leur **autonomie**.

Par exemple, une personne confrontée à des problèmes de dépendance pourrait être gênée de recevoir des publicités pour des centres de désintoxication, tandis qu'une femme ayant avorté pourrait trouver perturbant de voir des publicités pour des articles de maternité. Ainsi, certains patients ou utilisateurs d'objets connectés et d'applications de santé peuvent souhaiter retrouver un contrôle sur leurs données et les faire effacer, même après les avoir partagées initialement de manière consentie, on parle alors de "**droit à l'oubli**". (158)

2. Opacité des décisions de l'IA

Les systèmes d'IA peuvent parfois prendre des décisions complexes et difficiles à expliquer, ce qui pose des défis en termes de responsabilité en cas d'erreur ou de préjudice.

Actuellement, les systèmes de décision algorithmique en santé agissent principalement comme des **outils de soutien à la décision** humaine et ne disposent pas d'une pleine autonomie décisionnelle. (166) Bien que le jugement final reste entre les mains des professionnels de la santé, la complexité de certains systèmes d'IA rend parfois difficile, voire impossible, pour ces derniers de comprendre le raisonnement derrière les recommandations formulées. C'est ce qu'on appelle le phénomène de la "**boîte noire**", symbole de l'opacité du fonctionnement de certains système d'IA, en particulier ceux utilisant le ML en raison de la complexité de leurs algorithmes et de leur nature adaptative. (166)

Cette opacité peut entraîner une diminution de **l'interprétabilité**, de **l'acceptabilité** et de **l'adoption** de la technologie par les professionnels et les patients, et soulève plusieurs enjeux. Par exemple, un clinicien pourrait être en désaccord avec une recommandation d'un système d'IA mais ne pas pouvoir comprendre le raisonnement derrière celle-ci, ce qui entrave sa confiance dans la technologie. De même, des recommandations potentiellement erronées des algorithmes pourraient causer des préjudices aux patients, surtout s'ils ne sont pas informés que des décisions concernant leur santé ont été prises à l'aide de l'IA. (159)

De plus, bien que développés pour améliorer la santé, les systèmes de santé basés sur l'IA sont sujets **aux erreurs humaines de leurs développeurs** et aux **problèmes techniques** et peuvent donc fournir des recommandations erronées. Le **manque de transparence** de ces systèmes rend ces recommandations difficilement contestables et peut avoir un impact significatif sur la qualité des soins et la sécurité des patients.

Il est donc souhaitable que les entreprises augmentent l'interprétabilité de ces systèmes et mettent en place des **mécanismes de transparence** pour garantir que les décisions prises par les systèmes d'IA en santé sont justes, équitables et responsables. En effet, il est essentiel que les algorithmes soient conçus de manière à être **compréhensibles et vérifiables** par le personnel soignant. Les algorithmes utilisés pour l'aide à la prise de décisions médicales doivent être transparents, offrant ainsi la possibilité de comprendre les **facteurs déterminants** qui ont conduit à une recommandation diagnostique ou de traitement parmi la multitude de données analysées par l'IA. Cette transparence des modèles permettra à toutes les parties concernées, qu'il s'agisse des professionnels de santé, des patients ou des autorités publiques, de s'assurer du respect des réglementations et de minimiser les risques de biais éthiques ou moraux.

3. Biais algorithmiques et équité dans les décisions de santé

Les algorithmes des systèmes informatiques de santé utilisant le ML doivent s'appuyer sur des ensembles de données de qualité, fiables et **représentatifs**. Si les données utilisées pour entraîner les algorithmes d'IA sont **biaisées**, cela peut entraîner des préjudices considérables aux individus et communautés concernés. Les décisions prises par ces algorithmes peuvent **reproduire et amplifier les préjugés** présents dans les données, conduisant ainsi à des résultats **discriminatoires et injustes**. (159)

La **diversité des sources** de données, la **gestion du cycle de vie** des données, la **quantité** et la **représentativité** des données d'entraînement sont donc essentielles pour garantir la performance des algorithmes et la qualité de leurs prédictions et recommandations. L'étape d'évaluation de l'algorithme pour mesurer sa performance et sa capacité à généraliser doit utiliser un **ensemble de données distinct** de celui utilisé pour l'entraînement afin d'assurer une évaluation impartiale sur des données réelles.

Les algorithmes de ML sont influencés par les **décisions humaines** prises lors de leur développement, ce qui oriente leur conception, leurs paramètres et leur déploiement. Ces décisions reflètent souvent les **perspectives, croyances et biais individuels** des concepteurs, ce qui peut conduire à la **reproduction** et à l'**amplification de biais** dans les systèmes, par exemple en discriminant les individus issus de communautés marginalisées (167).

Par exemple, si un algorithme est entraîné sur des données principalement masculines, il pourrait donner des conclusions erronées lorsqu'il est confronté à des données féminines en contexte clinique. Respecter les **critères de validation** permet de réduire les risques que les systèmes informatiques de santé fonctionnent de manière sous optimale auprès de certaines populations et conduisent à **des traitements discriminatoires**. (158)

Ainsi, dans une perspective plus large, les décisions prises par les systèmes d'IA peuvent avoir un **impact sur l'équité en santé**, en favorisant certains groupes de population au détriment d'autres, notamment en raison de **biais raciaux, ethniques ou socio-économiques**. Il est donc crucial de mettre en place des **mesures d'atténuation** des biais algorithmiques, telles que l'utilisation de données diversifiées, la surveillance continue des performances des algorithmes et la sensibilisation à l'équité en santé parmi les développeurs et les utilisateurs d'IA en santé.

4. Utilisation non éthique des systèmes d'IA

Si l'IA peut améliorer divers aspects du système de santé, les décideurs **doivent évaluer l'éthique des applications** envisageables, car toutes ne sont pas souhaitables. Celles qui enfreignent les principes éthiques établis et causent de graves préjudices ne devraient pas être tolérées et devraient être modifiées ou retirées du marché. (158)

Cependant, le **caractère non éthique** de certains systèmes n'est pas toujours évident à détecter. Les **intentions** des développeurs ou propriétaires peuvent être dissimulées dans les conditions d'utilisation et les politiques de confidentialité, causant parfois un préjudice insoupçonné aux individus. (160) (161).

Par exemple, une montre intelligente enregistre la localisation des entraînements physiques mais peut être utilisée pour surveiller les déplacements et orienter les publicités. Ce type de pratique porte atteinte à l'**autonomie** et à la **liberté** des individus qui devraient avoir le choix de leur mode de vie.

La validité scientifique et l'efficacité d'un système d'IA ne suffisent pas à en justifier le déploiement et à le qualifier de "**souhaitable**". Malgré les avantages potentiels, les effets négatifs doivent être évalués et pourraient décourager ou exclure l'utilisation de la technologie (158).

De même, si la balance entre les coûts et les bénéfices est primordiale dans l'évaluation des systèmes d'IA en santé, l'évaluation des aspects éthiques doit en faire partie intégrante. Par exemple, un système d'IA promettant des économies de ressources importantes qui pourrait compromettre la qualité des soins ou entraîner des risques de discrimination ou de violation du secret médical ne devrait évidemment pas être approuvé. Les organisations doivent proactivement mettre en place des **processus transparents** pour explorer les enjeux éthiques de leurs technologies, atténuer les risques et réagir rapidement en cas de tort imprévu.

5. Enjeux éthiques liés au développement d'agents moraux artificiels

Certains défis éthiques liés à l'IA en santé concernent le développement d'agents moraux artificiels, capables de **raisonnement**, de **jugement** et de **prise de décision autonome** (168). Bien que ces systèmes ne soient pas encore largement utilisés dans le domaine de la santé, leur émergence dans d'autres secteurs (ChatGPT, voitures autonomes...) laisse envisager leur futur déploiement.

Les agents moraux artificiels pourraient avoir diverses applications dans le domaine de la santé, telles que le **diagnostic autonome** des patients (159). Cependant, la **délégation totale de l'expertise médicale** à de tels agents pourrait compromettre la participation du patient dans la prise de décision concernant sa santé. De plus, la complexité de ces systèmes pourrait **affecter la qualité du consentement** du patient et **perturber la relation** entre le professionnel de la santé et son patient (159)

Une autre conséquence de la délégation de responsabilité morale aux systèmes d'IA concerne les **valeurs morales** et **visions du monde** qu'ils véhiculent, ainsi que les **intérêts des développeurs**. Ces derniers sont susceptibles de biaiser le développement des systèmes en fonction de leurs propres perspectives et objectifs, ce qui soulève des questions sur la **neutralité des valeurs promues** par ces systèmes. Sans réglementation et encadrement éthique, il n'y a aucune garantie que ces valeurs soient alignées avec celles du patient ou de la société.

6. Enjeux éthiques liés au pouvoir des entreprises commercialisant les systèmes d'IA

Les systèmes d'IA utilisés dans le domaine de la santé proviennent principalement de la recherche universitaire et des entreprises privées. Cependant, certaines de ces entreprises opèrent avec un **manque de transparence** concernant la gestion et la protection des données personnelles de leurs utilisateurs. Cette opacité, couplée à une **asymétrie dans la collecte et la divulgation d'informations** (les entreprises accumulent un volume croissant de données sur leurs utilisateurs tout en divulguant le moins d'informations possibles), **renforce le pouvoir des entreprises** tout en **limitant celui des utilisateurs** et des autorités. (169) (170).

De plus, les grandes entreprises technologiques peuvent **exercer une pression sur les autorités** et les gouvernements pour **influencer les politiques et les lois** en fonction de leurs intérêts commerciaux, qui ne correspondent pas toujours aux intérêts de la santé publique et de la société. (158)

Pour conclure, il est indispensable de résoudre les défis éthiques associés à l'intégration de l'IA dans le domaine de la santé afin de consolider sa crédibilité et de favoriser son développement dans le respect des normes morales et éthiques. Ces enjeux jouent un rôle crucial dans la construction d'un cadre éthique solide qui garantira la confiance des patients et des professionnels de santé ainsi que l'acceptation sociale de l'IA en santé, ouvrant la voie à un progrès technologique éthiquement responsable.

II. Les défis réglementaires et juridiques

Dans un secteur déjà soumis à une régulation rigoureuse, l'émergence de l'IA dans l'industrie des produits de santé engendre des enjeux réglementaires et juridiques cruciaux. Dans cette section, nous aborderons cinq de ces défis. Tout d'abord, nous examinerons le cadre réglementaire pour la protection des données, suivi du cadre réglementaire pour l'approbation des DM basés sur l'IA. Ensuite, nous nous pencherons sur les enjeux juridiques de la responsabilité décisionnelle, ainsi que sur les défis juridiques des collaborations pharmaceutiques impliquant l'IA. Enfin, nous discuterons de l'importance de l'harmonisation des normes et des pratiques réglementaires au niveau national et international pour relever ces défis de manière efficace et cohérente.

1. Cadre réglementaire pour la protection des données

Pour assurer la confidentialité des informations des patients, les entreprises sont tenues de se conformer à des normes rigoureuses de protection des données telles que le **RGPD** (Règlement Général sur la Protection des Données) en Europe et les lois **HIPAA** (Health Insurance Portability and Accountability Act) aux États-Unis.

Le RGPD établit des directives strictes sur la **collecte**, le **stockage**, le **traitement** et la **protection** des données personnelles, y compris les données de santé. Les entreprises opérant sur le territoire de l'Union Européenne (UE) sont tenues de se conformer à ces règles, sous peine de **lourdes sanctions** en cas de non-respect. Pour les manquements les plus graves, ces sanctions peuvent atteindre jusqu'à **20 millions d'euros** ou **4% du chiffre d'affaires** annuel mondial de l'entreprise. (171)

De même, aux États-Unis, la loi HIPAA exige des mesures de sécurité strictes pour protéger la confidentialité des informations de santé, notamment en obligeant les entreprises de santé à mettre en place des protocoles de sécurité appropriés, à garantir l'accès restreint aux données et à notifier les violations de données dans les délais prescrits. En se conformant à ces normes, les entreprises peuvent renforcer la confiance des patients dans la protection de leurs données personnelles et médicales.

Pour garantir la conformité avec ces normes réglementaires, les entreprises pharmaceutiques doivent développer des **mécanismes d'audit robustes**. Ces audits sont essentiels pour évaluer non seulement les **algorithmes d'IA** eux-mêmes, mais aussi leur **intégration dans les processus de fabrication** et leur utilisation dans les **opérations quotidiennes**.

2. Cadre réglementaire pour l'approbation des dispositifs médicaux basés sur l'IA

La **loi sur l'IA de l'UE** est la première réglementation globale de l'IA au monde, visant à garantir la **sécurité**, la **transparence** et la **non-discrimination** des systèmes d'IA utilisés dans l'UE. Elle classe les systèmes d'IA en fonction du **niveau de risque** et impose des obligations spécifiques aux fournisseurs et utilisateurs en fonction de ce niveau de risque. (172)

Les systèmes d'IA à **risque inacceptable** c'est-à-dire considérés comme une menace pour les personnes, seront **interdits**. Cette catégorie regroupe les IA recourant à la manipulation cognitive et comportementale tels que des jouets qui encourageraient des comportements dangereux chez les enfants, l'établissement de scores sociaux basés sur diverses données individuelles telles que le statut socio-économique, la catégorisation et l'identification biométriques des personnes, ainsi que les systèmes de reconnaissance faciale en temps réel et à distance.

Les systèmes d'IA dont le niveau de risque est jugé comme **élevé** car ils pourraient compromettre la sécurité des individus ou leurs droits fondamentaux seront catégorisés en deux groupes distincts : les systèmes d'IA **intégrés dans des produits soumis à la réglementation de l'UE sur la sécurité**, comme les **DM**, et ceux spécifiques à certains domaines nécessitant un enregistrement dans une base de données de l'UE. Ce texte, qui devrait être pleinement applicable d'ici avril 2027 prévoit l'évaluation de tous les systèmes d'IA à haut risque **avant leur mise sur le marché et tout au long de leur cycle de vie** et permet aux citoyens de **déposer des plaintes** concernant les systèmes d'IA auprès des autorités nationales compétentes. (172)

Les exigences réglementaires concernant les DM sont définies par les règlements **UE 2017/745** et **2017/746**, tandis que le règlement **UE 2016/679** établit les normes en matière de protection des données. Le règlement 2017/745 englobe la **commercialisation** des systèmes d'IA utilisés pour le **diagnostic** ou le **traitement**, dès lors qu'ils entrent dans la catégorie des DM. Ces solutions d'IA sont là encore classées selon leur niveau de risque, allant de la **classe IIa** (risque potentiel modéré/mesuré) à la **classe III** (risque élevé) de la classification des DM. Les DM incluant l'IA doivent également démontrer un rapport bénéfice/risque favorable, qui doit être maintenu dans la durée. (173) (174)

	NIVEAU DE RISQUE	EXEMPLES
Classe I	Faible	<ul style="list-style-type: none"> • Compresses • Lunettes correctrices • Béquilles
Classe IIa	Modéré	<ul style="list-style-type: none"> • Lentilles de contact • Couronnes dentaires • Aides auditives
Classe IIb	Elevé	<ul style="list-style-type: none"> • Préservatifs • Pomes à perfusion • Produits de désinfection pour lentilles
Classe III	Très élevé	<ul style="list-style-type: none"> • Implants mammaires • Stents • Prothèses de hanche

Figure 57 : Classification des DM

Le **marquage CE** (Conformité Européenne) **médical** est la norme unifiée pour tous les DM (*Figure 58*), imposant un ensemble d'exigences spécifiques qui prennent en compte les particularités de chaque DM et notamment des spécificités propres aux logiciels, telles que la gestion des interactions avec les systèmes informatiques, la protection contre les accès non autorisés aux données, ainsi que des critères ergonomiques spécifiques à l'utilisation d'un DM sur un smartphone (173).

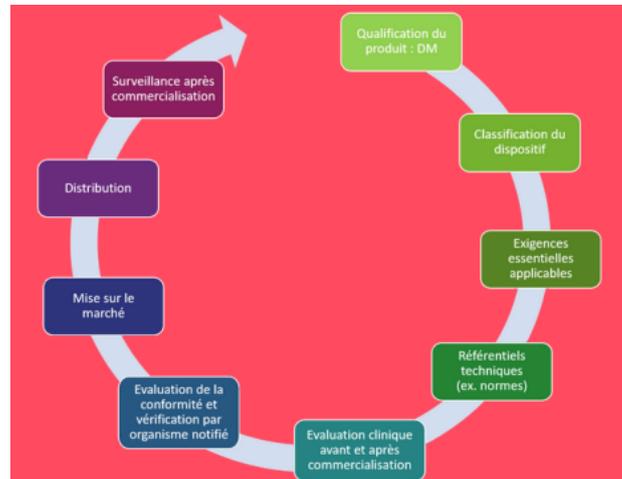


Figure 58 : Étendue du marquage CE médical (173)

Avant leur commercialisation, tous les DM, y compris ceux utilisant l'IA, doivent subir une **évaluation clinique** rigoureuse, fondée sur plusieurs éléments. Cela comprend une **analyse critique de la littérature pertinente**, l'examen des **résultats des études cliniques** disponibles sur le DM, ainsi que la **considération des alternatives thérapeutiques**. De plus, un plan de **surveillance post-commercialisation** doit être établi avant la mise sur le marché. Pendant la certification du DM, l'organisme notifié vérifie et valide ces aspects. L'objectif de cette évaluation pré commercialisation est de déterminer l'équilibre entre les bénéfices et les risques par rapport aux autres solutions thérapeutiques disponibles sur le marché, ainsi que de définir les bénéfices cliniques significatifs et mesurables pour le patient. (173)

La **Haute Autorité de Santé (HAS)** a récemment publié **un guide** portant sur les critères spécifiques pour **l'évaluation clinique** des **DM connectés**. De plus, elle requiert désormais des informations détaillées sur les fonctionnalités des DM exploitant des techniques de ML. Cependant, le système d'évaluation demeure encore limité en ce qui concerne les outils intégrant l'IA, et certains équipements disponibles pour les professionnels de la santé ne sont pas encore couverts par la Commission Nationale d'Évaluation des Dispositifs Médicaux et des Technologies de Santé (CNEDiMTS).

Ces évaluations cliniques sont réalisées et vérifiées par des professionnels dotés d'une **expertise clinique**, tant du côté du fabricant que de l'organisme notifié. Du côté du fabricant, l'évaluation clinique doit être validée et signée par des individus ayant une compétence documentée, incluant leurs CV et les formations suivies. De leur côté, les organismes notifiés doivent disposer d'un personnel possédant une expertise clinique, dont la compétence est évaluée par les autorités compétentes telles que l'ANSM en France, lors des procédures de notification et des inspections de suivi. L'organisme notifié effectue également des vérifications chez le fabricant concernant les compétences des personnes ayant réalisé et signé les évaluations cliniques du dispositif médical, ainsi que les résultats cliniques présentés dans le cadre de la certification du marquage CE médical. (173)

Après la mise sur le marché, lors des **évaluations de surveillance et de recertification**, l'organisme notifié vérifie la conformité au plan de surveillance post-commercialisation établi par le fabricant et le maintien d'un rapport bénéfice/risque favorable. L'objectif est de continuellement mettre à jour l'évaluation clinique initiale du DM en se basant sur les données recueillies lors du suivi clinique post-commercialisation pendant **l'utilisation réelle du dispositif**. En **identifiant tout risque émergent potentiel**, ce processus vise à confirmer la sécurité et les performances du DM. (173)

L'un des défis majeurs posés par les DM intégrant l'IA réside dans la **nature évolutive** de cette technologie, alimentée par un processus d'apprentissage continu. Dans ce contexte, le **COCIR** (coordination européenne des industries radiologiques, électromédicales et de technologies de l'information pour soins de santé), a émis des recommandations pour promouvoir l'innovation tout en garantissant des normes élevées de qualité et de sécurité pour les patients. Le COCIR propose notamment de **définir les modalités** de surveillance d'un dispositif basé sur l'IA par un opérateur humain, ainsi que les aspects et les mécanismes de son évolution. Cela comprend la spécification de l'état **verrouillé** ou **adaptatif** de l'IA, la description du **processus d'apprentissage** et ses limites, les procédures en cas de modification en cours d'exécution, les mesures pour garantir la performance et la sécurité pendant l'adaptation, et les mécanismes pour contrôler la qualité des nouvelles données. Il est essentiel de définir clairement **les conditions déclenchant le changement d'algorithme**, ainsi que les **niveaux de performance** et de **confiance** attendus. (174)

En outre, le COCIR préconise que les protocoles de rejet ou de retour à un état antérieur en cas de changement soient spécifiés. Le COCIR encourage également à **informer l'utilisateur** lorsqu'il interagit avec une IA plutôt qu'avec un agent humain, surtout si cette distinction n'est pas évidente. (174)

Du point de vue réglementaire, le principal défi consiste donc à **valider les modifications issues de l'apprentissage continu des algorithmes**, une pratique actuellement non autorisée selon la réglementation en vigueur pour les DM. Cette réglementation **limite les modifications algorithmiques** à des **seuils prédéfinis**, établis lors de l'évaluation de la conformité. Pour permettre le déploiement d'IA capables de modifications au-delà de ces limites, une **révision réglementaire** est nécessaire. Par ailleurs, le RGPD pose des obstacles aux IA médicales autonomes en Europe, notamment par le biais de son article 22 concernant les "Décisions individuelles automatisées", garantissant à l'utilisateur "le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé de ses données à caractère personnel". (174) Des directives spécifiques semblent donc nécessaires pour évaluer la sécurité et l'efficacité de ces dispositifs, en tenant compte de leur **évolutivité**. De plus, une mise à jour réglementaire doit prendre en compte la rapidité de l'évolution des technologies d'IA et mettre en place des **processus d'approbation accélérée** pour garantir leur conformité **aux dernières avancées technologiques** tout en maintenant des normes élevées de sécurité et d'efficacité.

En parallèle, il est crucial de **poursuivre le dialogue** entre les autorités réglementaires, les professionnels de santé et les développeurs technologiques pour établir des **cadres réglementaires appropriés** favorisant l'innovation tout en assurant la sécurité et le respect de la vie privée des individus.

3. Enjeux juridiques de la responsabilité décisionnelle

L'utilisation des systèmes d'IA pour l'aide à la prise de décision clinique complique la question de la **responsabilité**. En effet, il est souvent difficile de déterminer qui est responsable en cas d'erreur ou de dommage causé par un système d'IA, que ce soit l'entreprise commercialisant l'IA, l'utilisateur final (le clinicien) ou les concepteurs des algorithmes. Si le clinicien n'a ni compréhension ni contrôle sur le système, peut-il être tenu responsable, légalement et/ou moralement de l'erreur ? La performance supérieure des systèmes d'IA peut justifier leur utilisation en clinique, mais qui est responsable en cas d'erreur ?

Le **partage de la responsabilité** entre les acteurs impliqués dans le développement, l'approbation et le déploiement des systèmes d'IA est complexe et peut aboutir à une **dilution de la responsabilité** qui finit parfois par n'être portée par personne, ce qui est préjudiciable aux patients en cas d'erreur médicale.

Cette situation souligne la nécessité urgente d'un **encadrement juridique clair et précis de la responsabilité décisionnelle** liée aux systèmes d'IA en médecine. Une approche concertée entre les législateurs, les professionnels de santé, les développeurs d'IA et les experts en éthique est indispensable pour définir des directives claires sur la répartition de la responsabilité et les mécanismes de reddition de comptes en cas d'erreurs ou de préjudices. Il est également crucial d'instaurer des **procédures de recours efficaces** pour les patients victimes d'erreurs médicales impliquant des systèmes d'IA, afin de leur garantir un accès équitable à la justice et à une indemnisation adéquate en cas de préjudice.

4. Enjeux juridiques des collaborations pharmaceutiques - IA : questions de la propriété intellectuelle

Les collaborations entre les entreprises pharmaceutiques et les startups spécialisées en IA suscitent des interrogations juridiques complexes. Les entreprises pharmaceutiques, disposant de vastes données et d'une expertise en développement de médicaments, s'associent aux spécialistes de l'IA, dont le savoir-faire leur permet d'analyser de grandes quantités de données. Cette convergence vise à déployer des technologies d'IA de pointe dans la R&D pharmaceutique et peut prendre différentes formes. En effet, la structure de la collaboration dépend de l'approche et de la perspective des parties prenantes ; D'un côté, le partenaire IA peut se limiter à offrir des services de conseil et une plateforme technologique à utiliser avec les données du partenaire pharmaceutique. D'un autre côté, le partenaire IA peut souhaiter mener sa propre R&D pharmaceutique et pourrait vouloir tirer des enseignements des résultats de la collaboration et réserver les droits d'utilisation. Ainsi, ce type de collaboration soulève des interrogations juridiques complexes quant à **la propriété intellectuelle** et à **la répartition des droits** entre les partenaires.

L'évolution continue des modèles générés par les technologies d'IA complexifient la propriété intellectuelle des résultats ainsi que **l'utilisation des données sous-jacentes**. En effet, le modèle résultant de l'entraînement intègre les données qui lui ont servi de base, créant ainsi une tension entre la **propriété intellectuelle principale** (associée à l'IA elle-même) et **celle de fond** (les données fournies par le partenaire pharmaceutique pour former les modèles d'IA). Par conséquent, la question de la détention de la propriété intellectuelle résultante et les modalités d'utilisation des données dérivées demeure un sujet de débat entre les parties, les accords pouvant varier en fonction des circonstances.

Par ailleurs, l'implication de l'IA dans la découverte de médicaments soulève des préoccupations en ce qui concerne la **qualification et l'éligibilité** des inventions résultantes pour l'obtention de **brevets**. Actuellement, dans de nombreux pays, les lois sur les brevets ne reconnaissent pas les machines comme des inventeurs. Dans l'ensemble, le domaine de la propriété intellectuelle liée à l'IA est en constante évolution, et de nouvelles lois peuvent être adoptées pour relever ces défis. (175)

5. Harmonisation des normes réglementaires au niveau national et international

La dimension **politique et géostratégique** de l'IA suscite des questionnements quant **au cadre de régulation** le plus approprié. Actuellement, des initiatives de réglementation émergent à différents niveaux, allant des chartes locales aux directives internationales. Cependant, cette diversité peut **complexifier la lisibilité** pour les parties prenantes et ainsi **diminuer la confiance** portée aux régulations autour de l'IA, et par extension à l'IA elle-même.

Bien que de nombreux pays européens aient développé leur propre stratégie nationale sur l'IA comme le gouvernement français qui a lancé en **2018** sa **Stratégie Nationale pour l'Intelligence Artificielle (SNIA)** pour positionner la France en tant que leader européen et mondial de l'IA, il est devenu évident que l'échelon national seul ne suffit pas à établir des normes internationalement reconnues, en particulier par les grandes entreprises. Ainsi, pour garantir une réglementation efficace et cohérente des technologies d'IA en santé, il semble nécessaire de favoriser la **coopération** et **l'harmonisation des normes** et des pratiques réglementaires entre les pays.

L'UE, avec son marché de taille critique et son leadership en matière d'exigences éthiques, pourrait jouer un rôle clé dans l'établissement de standards mondiaux comme l'a démontré le RGPD. Cette dynamique est cruciale pour garantir une **souveraineté européenne** dans le domaine de l'IA, notamment en santé. Le **projet de réglementation européen sur l'IA** reflète cette ambition en visant à établir des normes élevées de protection des droits et de la sécurité des citoyens. Cependant, le projet actuel **manque de clarté et de cohérence**, présentant des **lacunes** et des **conflits réglementaires** non résolus, ce qui soulève des questions sur sa mise en œuvre effective et son impact potentiel sur l'innovation et la compétitivité.

En somme, il est impératif d'envisager une **évolution réglementaire** pour répondre aux défis posés par l'utilisation de l'IA dans le domaine de la santé afin de favoriser l'innovation tout en protégeant la sécurité des patients.

III. Les enjeux techniques

1. Les défis liés à l'accès aux données

Comme évoqué dans les parties précédentes, les techniques d'IA nécessitent des données massives, fiables et accessibles. Bien que la France dispose de bases de données fournies, leur sous-utilisation persiste en raison de leur **dispersion**, de leur **accès complexes** et d'une **faible interopérabilité**. Pour remédier à cette situation, le **Health Data Hub** a été lancé en **2019** pour regrouper ces bases de données et permettre une exploitation sécurisée à des fins de recherche (*Figure 59*). Il rassemble actuellement des bases telles que le **SNIIRAM** (base associée à la carte vitale qui enregistre les données de remboursement), le **PMSI** (base qui suit le parcours hospitalier des patients) et le **CépiDC** (base recensant les causes médicales de décès). Le Health Data Hub est enrichi par un catalogue de données évolutif, adapté aux besoins de l'écosystème dont la première composition a été établie par un arrêté en date du 12 mai 2022. Cependant, malgré les efforts mis en place, les processus d'accès aux données de santé demeurent longs et complexes, entravant leur utilisation pour la recherche et l'innovation. (176)

Health Data Hub, mode d'emploi

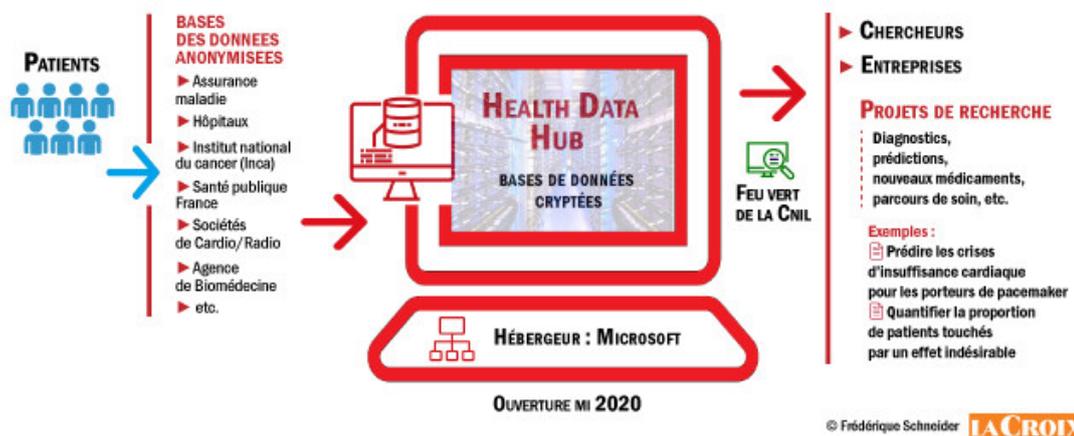


Figure 59 : Fonctionnement du Health Data Hub (177)

La problématique du stockage des données de santé présente différents modèles, allant du stockage **centralisé** au stockage **décentralisé**. En France, une approche centralisée est favorisée à travers le Health Data Hub, bien que certains acteurs maintiennent des entrepôts de données locaux, tels que l'APHP (Assistance Publique - Hôpitaux de Paris).

2. Les défis liés à la standardisation des données et à l'interopérabilité des systèmes

En intégrant les systèmes d'IA en santé avec les infrastructures existantes, telles que les DME et les systèmes de gestion hospitalière, plusieurs défis techniques se posent. Tout d'abord, les infrastructures médicales traditionnelles peuvent être **fragmentées** et **hétérogènes**, avec des **normes** et des **formats de données** différents d'un établissement à l'autre, ce qui complique l'adaptation des systèmes d'IA pour fonctionner de manière cohérente et efficace dans ces environnements variés. De plus, les systèmes d'IA nécessitent souvent un accès à de vastes ensembles de données pour fonctionner efficacement. Cependant, les données médicales et personnelles des patients sont intrinsèquement complexes, hétérogènes et souvent dispersées dans différents systèmes, ce qui rend leur **intégration ardue et chronophage**.

Ainsi, la **standardisation des données** est essentielle pour faciliter l'échange de données entre les différents acteurs du domaine de la santé et **garantir l'interopérabilité** des systèmes d'IA. **L'harmonisation de la nomenclature**, qui constitue un élément clé de cette standardisation, inclut la normalisation des langages médicaux, biologiques, génétiques, formels ou informels, ainsi que la grammaire variée des textes ou des images.

Pour une compréhension cohérente et une communication efficace entre différents langages, l'IA doit impérativement s'appuyer sur un vocabulaire standardisé et contrôlé. Des standards de nomenclature tels que **Snomed** (www.snomed.org/snomed-ct) ou **Human Phenotype Ontology** (<http://human-phenotype-ontology.github.io/about.html>) représentent des exemples de ces normes que l'IA doit être en mesure de gérer. De plus, des normes d'interopérabilité, telles que la **norme HL7** (Health Level Seven ; <http://www.hl7.org/>), facilitent l'intégration de données provenant de différentes sources dans le système d'IA, en structurant et codifiant les données et le vocabulaire à des fins d'interprétation et d'interopérabilité (*Figure 60*). (150)

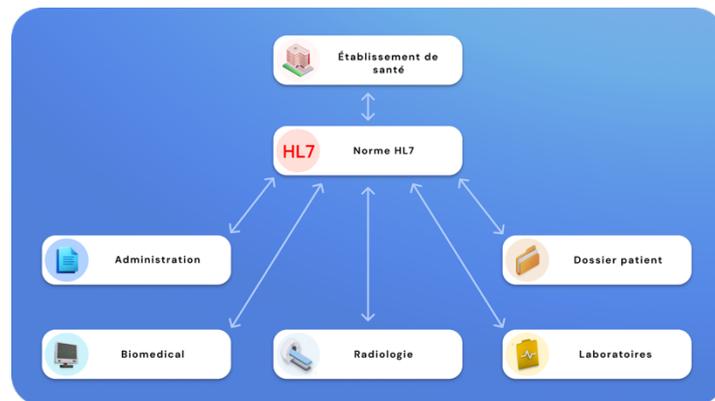


Figure 60 : HL7, Pilier de l'interopérabilité dans le secteur de la santé (178)

HL7 établit un langage partagé entre les établissements de santé et offre ainsi une interopérabilité fluide

La standardisation des données n'est pas le seul défi à relever pour assurer une intégration réussie des systèmes d'IA en santé avec les infrastructures existantes. En effet, les processus de **mise en œuvre et d'intégration** des systèmes d'IA peuvent être coûteux et exiger des ressources considérables en termes de temps et de ressources humaines. Les établissements de santé doivent souvent investir dans de **nouvelles infrastructures**, **former leur personnel** à l'utilisation des nouvelles technologies et mettre en œuvre des **protocoles de sécurité** robustes pour protéger les données des patients.

Enfin, il est essentiel de s'assurer que les systèmes d'IA intégrés fonctionnent de manière harmonieuse avec les **processus cliniques existants** et qu'ils améliorent véritablement l'efficacité des soins de santé. Cela nécessite une collaboration étroite entre les prestataires de soin de santé, les développeurs de technologie et les régulateurs pour s'assurer que les systèmes d'IA répondent aux besoins réels des professionnels de la santé et des patients, tout en respectant les normes éthiques et réglementaires en vigueur.

3. Sécurité et protection contre les cyberattaques

Les systèmes d'IA en santé sont susceptibles d'être la cible de **cyberattaques** visant à compromettre l'intégrité et la confidentialité des données médicales des patients, ce qui soulève des préoccupations majeures en matière de sécurité. Ces attaques menacent l'intégrité des systèmes informatiques de santé, compromettant non seulement la **confidentialité des données** mais également la **performance des algorithmes**. Ces incidents peuvent entraîner des conséquences graves telles que le **ciblage indésirable des individus** ou des **tentatives d'hameçonnage** crédibles sur la base de leurs données médicales, le **vol d'identité**, et même **paralyser des établissements de santé** voire tout un système de santé.

Par exemple, plus de **33 millions** de français ont été affectés par une **fuite de données** lors d'une cyberattaque contre **deux gestionnaires du tiers payant** pour des complémentaires santé en février 2024. Les sociétés visées sont **Viamedis** et **Almerys**, qui sont des intermédiaires entre les professionnels de santé et les assureurs complémentaires. Les données compromises comprennent **l'état civil**, la **date de naissance**, le **numéro de sécurité sociale** et les **garanties du contrat**. Ce vol massif de données personnelles expose potentiellement les assurés à des risques tels que le vol d'identité ou la fraude. (179)

De même, les **hôpitaux** sont devenus des cibles privilégiées pour les cybercriminels en raison des données sensibles qu'ils détiennent. A titre illustratif, en 2020, en pleine crise du Covid-19, **27 cyberattaques** d'hôpitaux ont été recensées en France. Ces cyberattaques, souvent accompagnées de **demandes de rançons**, entraînent des conséquences graves telles que **l'interruption des soins** et la **perte de données médicales critiques**. En outre, la paralysie d'un établissement de santé en raison d'une cyberattaque a des répercussions majeures sur l'accessibilité, la communication et la sécurité des soins de santé, soulignant ainsi l'urgence de renforcer la sécurité des systèmes d'IA en santé. (180)

Pour contrer ces menaces et renforcer la sécurité, plusieurs **mesures de prévention** sont indispensables. Tout d'abord, la **segmentation réseau**, qui consiste à diviser le réseau informatique en segments distincts pour **limiter la propagation** des cyberattaques et **isoler les données sensibles**. Ensuite, la **sensibilisation du personnel** est cruciale ; les professionnels de santé doivent être formés aux **bonnes pratiques de sécurité informatique**, être conscients des risques et des procédures à suivre en cas d'incident. Enfin, la **collaboration avec l'Agence Numérique en Santé** est également essentielle.

Cette agence peut fournir des conseils et des ressources pour améliorer la sécurité des systèmes d'information de santé, ainsi que des lignes directrices pour une meilleure gestion des risques et des incidents. En parallèle, pour protéger les données traitées par les systèmes d'IA, l'utilisation de techniques de **cryptage** et **d'authentification avancées** est nécessaire. En combinant ces différentes mesures, les établissements de santé peuvent renforcer leur **cyber-résilience** et mieux protéger les données médicales de leurs patients. Une gestion efficace des risques implique une **surveillance continue**, une **détection précoce des menaces** et une **réponse rapide** en cas d'incident pour minimiser les impacts sur les patients et les opérations cliniques. (180)

IV. Les enjeux économiques

1. Tarification et modèle économique

Sur le plan économique, l'introduction de l'IA dans les industries de santé soulève des questions sur **sa tarification**. En effet, les modèles de tarification actuels ne semblent **pas adaptés** aux nouvelles dynamiques portées par les systèmes d'IA en santé. Les entreprises doivent donc développer des modèles de tarification qui reflètent la valeur ajoutée apportée par les solutions d'IA tout en restant compétitives sur le marché. Dans le cadre de l'intégration de ces technologies dans les structures hospitalières, il semblerait judicieux de considérer un investissement initial amorti dans le temps avec des **modèles de tarification dynamiques** (on peut par exemple imaginer des modèles basés sur des abonnements).

L'IA ne peut pas toujours être évaluée de la même manière que les autres technologies de santé, notamment en raison de leur **grande diversité** et de **leurs usages variés**. Ceci explique pourquoi les études d'efficacité actuelles se concentrent principalement sur l'imagerie, où l'utilisation de l'IA est plus standardisée. De plus, une **évaluation économique dynamique** des outils d'IA reste à élaborer puisque leur efficacité s'améliore avec les données, rendant les évaluations médico-économiques statiques actuelles obsolètes et inadaptées pour ce type de technologies. La HAS examine cette problématique à travers la CNEDiMTS afin de moderniser ses méthodes et procédures d'évaluation. (151)

L'efficience, qui consiste à utiliser des ressources limitées de manière optimale pour obtenir les meilleurs résultats, revêt une importance cruciale dans l'évaluation économique en santé. En évaluant l'efficience, les décideurs peuvent **comparer** diverses interventions médicales ou technologies de santé afin d'identifier celles offrant le **meilleur rapport qualité-prix**. Cependant, dans le contexte d'essai impliquant l'IA, il est difficile de comparer un bras expérimental à un bras contrôle comme dans un essai clinique plus traditionnel. Par conséquent, la revue systématique de la littérature révèle un **niveau de preuve** actuellement **faible** sur l'efficience de l'IA. Les études disponibles se concentrent principalement sur la minimisation des coûts, avec peu d'analyses sur la qualité de vie ou l'efficacité des coûts indirects. De plus, les données de vie réelle pour évaluer l'efficience dans la population générale sont **limitées**, et les informations sur l'impact à moyen/long terme de l'IA sont encore rares. Bien que certains avantages de l'IA soient mis en avant dans la littérature, des effets néfastes ayant des impacts économiques indirects sont également signalés : augmentation de la productivité des cliniciens mais aussi de leur **charge de travail**, amélioration de la santé de certaines populations mais **exacerbation des inégalités**, efficacité de l'IA mais **risque de surutilisation** entraînant des **conséquences budgétaires** défavorables, gains de temps grâce à l'automatisation des décisions mais **perte de savoir-faire**. Dans l'ensemble, il semble nécessaire de disposer de **davantage de preuves d'efficience** étant donné la prolifération croissante des applications d'IA. (151)

En outre, les **modèles économiques des entreprises** concevant des solutions d'IA en santé sont remis en question. En effet, les startups proposant des solutions d'IA rencontrent des **défis de financement** puisqu'actuellement, ces startups ne bénéficient pas de financements spécifiques au-delà des voies dérogatoires. Ainsi, elles sont souvent contraintes de **s'associer** à de plus grandes entreprises ou de **se faire racheter**, ce qui soulève des préoccupations de **souveraineté** dans un contexte de **concurrence internationale**. Il est donc nécessaire de créer des **voies d'accès au marché spécifiques et durables** pour l'IA, ce qui implique d'anticiper et de définir les modalités de sortie des voies dérogatoires et expérimentales, d'ajuster les modalités de financement aux particularités des solutions numériques, et de mener des recherches pour évaluer les valeurs économiques revendiquées par ces solutions. (151)

2. Remboursement des soins impliquant l'IA par les assurances santé

Alors que les avancées technologiques permettent l'intégration croissante de l'IA dans les pratiques médicales, les assureurs doivent s'adapter pour déterminer les **modalités de remboursement** appropriées. Une interrogation surgit alors : une **assurance obligatoire** (sécurité sociale) ou **complémentaire** (mutuelle) doit-elle payer la même somme pour les actes médicaux réalisés par les professionnels de santé et ceux réalisés par de systèmes d'IA ? Cette interrogation soulève des défis quant à **l'équité des remboursements** ainsi que l'évolution du **rôle des assureurs** dans la santé de leurs assurés.

L'intégration de l'IA dans le domaine de la santé risque de changer la manière dont les organismes d'assurances, publics et privés, appréhendent **la gestion du risque**. En effet, **l'individualisation** et la **personnalisation** potentielle de l'offre assurantielle permise par l'IA entraînera une **responsabilisation** accrue des assurés avec la possibilité de pénalités et des primes individualisées. Cette évolution remet donc en question les modèles classiques de **mutualisation des risques** et les **principes traditionnels d'assurabilité**. De plus, elle pourrait transformer les assureurs en acteurs majeurs de la **promotion de la santé préventive**, contribuant ainsi à la **réduction des risques et des coûts associés**. (151)

V. Les défis organisationnels et sociétaux

1. Impact sur les professionnels de santé

L'intégration de l'IA dans les pratiques médicales promet une automatisation de certaines tâches, libérant ainsi du temps pour les professionnels de santé afin qu'ils se concentrent davantage sur **l'écoute** et **l'accompagnement** des patients. Cette évolution pose la question de la **redéfinition du rôle** des professionnels de santé, nécessitant **l'acquisition de nouvelles compétences** en matière de traitement des données, **d'interprétation des résultats** générés par les systèmes d'IA et de collaboration avec ces technologies.

Bien que les algorithmes facilitent le processus de diagnostic, leur validation par un professionnel de santé reste indispensable, impliquant une **compréhension approfondie des principes sous-jacents** de fonctionnement des algorithmes, de leur **performance**, de leurs **limites** ainsi que les implications cliniques de leurs résultats.

En effet, si certains postes risquent d'être rendus obsolètes en raison de l'intégration de l'IA, engendrant ainsi une **redistribution des responsabilités** professionnelles et potentiellement des **pertes d'emploi**, de nombreux travailleurs devront s'engager dans une **collaboration** avec ces systèmes d'IA. Cela soulève la nécessité d'une **formation en IA** pour le personnel médical mais également administratif et de soutien, ainsi que pour les patients et leurs proches (159). Actuellement, la compréhension de l'IA et de ses applications reste limitée parmi les professionnels de la santé et les patients (181), ce qui peut engendrer une **dépendance excessive** à la technologie et **empêcher l'intégration** efficace des systèmes d'IA dans leur quotidien. De plus, l'IA pourrait remettre en question l'autonomie et la dignité des travailleurs de la santé en **dévalorisant leur expertise** (159). Il est donc important de souligner que le développement des outils nécessite **l'implication de spécialistes** pour constituer des bases de données adéquates (annotation, association aux bons diagnostics...), démontrant ainsi l'importance de la collaboration entre l'IA et les professionnels de la santé.

L'évolution du métier de médecin et des autres professionnels de santé semble donc inévitable, avec une **redéfinition de leurs contours** et la nécessité d'une **meilleure coordination** entre les différents acteurs. Ainsi, une intégration efficace de l'IA en santé nécessite une adaptation des institutions de formation et des organisations de santé. Des **programmes de formation continue** doivent être mis en place pour permettre aux professionnels de santé de rester à jour avec les dernières avancées en matière d'IA et de s'adapter aux changements dans leur pratique quotidienne. Malgré les progrès réalisés, il reste encore beaucoup à faire pour généraliser l'utilisation de ces outils. L'intégration de ces compétences dans les programmes universitaires est également un défi, étant donné la charge d'apprentissage déjà importante pour les étudiants en santé. Ainsi, une approche coordonnée est essentielle pour développer des programmes de formation adaptés et sensibiliser les professionnels de santé à l'utilisation optimale de l'IA.

2. Enjeux organisationnels et managériaux

L'intégration croissante de l'IA dans le secteur de la santé entraîne des enjeux organisationnels et managériaux majeurs. Les entreprises et les établissements de santé doivent repenser leurs processus de travail pour intégrer efficacement les technologies d'IA. Cela implique souvent des **changements organisationnels** significatifs, notamment dans la manière dont les tâches sont réparties, coordonnées et exécutées.

De plus, le déploiement de l'IA nécessite une adaptation des **structures managériales** pour accompagner efficacement ces transformations. Le **management transactionnel**, axé sur la supervision des tâches et des opérations quotidiennes, est particulièrement touché, car il doit s'adapter pour intégrer les nouveaux processus basés sur l'IA. Les gestionnaires doivent également se tourner vers le **management transformationnel**, axé sur la vision, l'innovation et le changement, pour conduire efficacement ces évolutions. Enfin, le **management de réseaux** qui vise à favoriser la transmission et le partage des connaissances au sein de l'organisation ainsi que le développement de nouvelles expertises devient crucial dans un contexte où les partenariats et la collaboration entre les acteurs de la santé et du médico-social sont essentiels pour exploiter pleinement le potentiel de l'IA et relever les défis complexes du secteur. (182)

En parallèle, il est impératif de promouvoir et mettre en œuvre le concept de "**RSE digitale**". Cette approche implique d'encourager activement les établissements de soins à **anticiper** les changements à venir en adoptant une approche proactive sur l'évolution des métiers dans le domaine de la santé. De plus, il est essentiel de mettre en place des **méthodes d'évaluation des compétences** des employés afin d'identifier les besoins en formation liés aux nouvelles technologies. Il est également crucial de **sensibiliser** les employés aux innovations à venir et de les former aux technologies actuelles, en mettant particulièrement l'accent sur les fonctions susceptibles d'être partiellement automatisées par l'IA. Ainsi, les enjeux organisationnels et managériaux de l'IA en santé nécessitent une **approche stratégique et proactive** pour garantir une intégration réussie et bénéfique de ces technologies dans le système de santé. (151)

3. Les inégalités d'accès à l'IA et leur implication dans les injustices sociales

Bien que l'IA et ses multiples applications semblent omniprésentes, leur accès n'est pas universel. Souvent, seuls les individus et les organisations issus **des pays riches** ou de **milieux socio-économiques élevés** ont les moyens d'y accéder. Ces disparités d'accès et d'utilisation contribuent à **aggraver les inégalités sociales** existantes et impactent la **représentativité** des données disponibles pour développer des systèmes d'IA conformes aux principes d'équité, de diversité et d'inclusion propres à une IA responsable. (159)

Ces écarts se manifestent également à l'échelle locale, entre différents groupes et communautés au sein d'un même pays. Les **groupes marginalisés**, historiquement précaires, ont souvent un accès limité à certains équipements de base et encore plus aux technologies d'IA, ce qui influence la qualité des technologies développées à l'échelle locale et mondiale. (159) L'accès limité à ces technologies signifie que certains groupes sont **sous-représentés voire exclus** des données utilisées pour les développer. Cette exclusion **renforce les disparités** sociales existantes et **marginalise** davantage ces groupes selon différents critères tels que le genre, l'âge, l'origine ethnique ou la condition médicale (167). De plus, lorsque ces systèmes sont déployés dans des contextes culturels différents de ceux pris en compte lors de leur développement, ils peuvent avoir des impacts néfastes significatifs sur la vie de ces individus. Des recommandations telles que la création d'un **comité d'éthique** des technologies numériques et de l'IA sont avancées pour aborder ces enjeux. (167)

4. Impact de l'IA en santé sur la société et ses valeurs

L'utilisation de l'IA soulève des questions sur **les choix fondamentaux de société** et sur la place des technologies d'IA dans nos vies. Cela renvoie à la définition d'une **vie satisfaisante et digne**, ainsi qu'aux **valeurs morales** qui peuvent varier selon les cultures et les époques. Par exemple, le déploiement d'un robot compagnon muni d'un agent conversationnel dans les maisons de retraite pour briser l'isolement des personnes âgées pose des questions éthiques sur notre relation avec les personnes âgées et sur le remplacement du contact humain par celui d'un robot. (167)

Certains auteurs mettent en garde contre les risques que des systèmes dotés d'une IA générale se détournent de leurs objectifs initiaux et adoptent des objectifs contraires à l'intérêt humain, voire échappent au contrôle humain (183). Si ces scénarios sont **controversés** et souvent **peu crédibles**, le développement de systèmes d'IA avancés nécessite de la prudence. De nombreux questionnements éthiques émergent des **changements du marché du travail** induits par l'IA en santé, suscitant des incertitudes sur la nature des transformations potentielles des tâches et des emplois, ainsi que sur leurs impacts sur le personnel et son identité professionnelle et donc sur la société. Cette évolution soulève des interrogations sur la pertinence et l'acceptabilité sociale des changements à venir. (184) (185)

VI. Les enjeux écologiques

L'adoption de l'IA dans le domaine de la santé soulève des enjeux écologiques importants. Bien que peu étudiés jusqu'à présent, les **impacts environnementaux** de l'IA sont significatifs et son déploiement massif peut entraîner une **augmentation de la consommation énergétique** et des **émissions de gaz à effet de serre**. En effet, les infrastructures nécessaires au fonctionnement de l'IA, telles que les centres de données et les serveurs informatiques, consomment d'importantes quantités d'énergie. De plus, le processus d'entraînement des modèles d'IA demande des ressources computationnelles considérables, contribuant ainsi à une empreinte carbone significative. (186)

Pour atténuer ces impacts écologiques, il est impératif de développer des stratégies visant à **optimiser l'efficacité énergétique** des infrastructures informatiques, à promouvoir l'utilisation de **sources d'énergie renouvelable** et à encourager des pratiques de développement logiciel **plus durables**. De plus, une **gestion responsable** des données de santé, en mettant l'accent sur la **minimisation** et le **stockage efficace** des données, peut contribuer à réduire l'empreinte environnementale de l'IA en santé. En intégrant des considérations écologiques dès la conception et l'implémentation des systèmes d'IA, il semble possible de concilier les progrès technologiques avec la préservation de l'environnement, garantissant ainsi un développement durable de la santé numérique.

Parallèlement, les entreprises doivent assumer leur responsabilité environnementale en intégrant des considérations écologiques dans leur stratégie d'entreprise et leurs décisions opérationnelles. Cela implique la mise en place de politiques de gestion des déchets électroniques pour une **élimination responsable des équipements obsolètes**, en favorisant le **recyclage** et la réutilisation des composants électroniques ainsi que l'adoption de normes environnementales strictes et la publication de rapports de durabilité pour rendre compte de leur performance environnementale.

Conclusion

Cette exploration approfondie met en évidence le rôle central de l'IA dans l'avenir de l'industrie des produits de santé. Elle révolutionne chaque étape du cycle du médicament, en accélérant la découverte et les essais pré-cliniques, en améliorant les essais cliniques, en optimisant la production, et en renforçant la pharmacovigilance. En plus de promettre une accélération significative de la recherche pharmaceutique, réduisant ainsi les délais de découverte de médicaments, augmentant les taux de succès et permettant des économies substantielles, l'IA offre une lueur d'espoir dans un contexte où les coûts augmentent pour les laboratoires pharmaceutiques tandis que les revenus ont tendance à diminuer, notamment avec l'expiration des brevets des blockbusters. L'IA offre des perspectives passionnantes pour l'avenir de la santé, en favorisant une médecine 4P (Prédictive, Préventive, Personnalisée et Participative), notamment en permettant des diagnostics plus précoces et plus précis, une personnalisation des traitements et une autonomisation des patients.

Cependant, l'adoption généralisée de l'IA dans le domaine de la santé n'est pas sans défis. Des questions éthiques, réglementaires, techniques, économiques, sociétales et environnementales émergent avec le déploiement de cette technologie dans l'industrie des produits de santé. Pour s'assurer que l'IA soit un moteur de progrès équitable et durable dans le domaine de la santé, il est impératif d'aborder ces défis de manière proactive et collaborative, en tenant compte des valeurs éthiques et des besoins sociaux. Une collaboration multidisciplinaire entre chercheurs, professionnels de santé, industriels et décideurs politiques s'impose donc pour surmonter ces défis et exploiter pleinement le potentiel de l'IA afin d'améliorer les soins de santé tout en veillant à ce que ses bénéfices soient accessibles à tous, dans le respect de nos principes fondamentaux et de notre environnement.

Le Doyen de l'UFR de Pharmacie,

Christiane FORESTIER

Le Président du Jury

David BALAYSSAC

Bibliographie

1. McCarthy J. What is Artificial Intelligence. 2007;
2. Georges B. Intelligence artificielle : de quoi parle-t-on ? : Constructif. 24 déc 2019;N° 54(3):5-10.
3. Larousse, éditeur. Intelligence Artificielle. In: Encyclopédie Larousse [Internet]. En ligne; [cité 27 oct 2022]. Disponible sur: https://www.larousse.fr/encyclopedie/divers/intelligence_artificielle/187257
4. Villani C. Rapport Villani « Donner un sens à l'intelligence artificielle ». 2018.
5. Maurice B. deeplylearning.fr. Cours théoriques de deep learning : fonctionnement du neurone artificiel. Disponible sur: <https://deeplylearning.fr/cours-theoriques-deep-learning/fonctionnement-du-neurone-artificiel/>
6. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*. déc 1943;5(4):115-33.
7. Turing AM. Computing Machinery and Intelligence. In: Epstein R, Roberts G, Beber G, éditeurs. *Parsing the Turing Test* [Internet]. Dordrecht: Springer Netherlands; 2009 [cité 12 déc 2022]. p. 23-65. Disponible sur: http://link.springer.com/10.1007/978-1-4020-6710-5_3
8. Jef A. *The Scientist Magazine*. Machine learning, 1951. Disponible sur: <https://www.the-scientist.com/machine--learning--1951-65792>
9. McCarthy J, Minsky M, Rochester N, Shannon C. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. 1955.
10. Perceptron – Tout savoir sur le plus vieil algorithme de Machine Learning [Internet]. Disponible sur: <https://www.lebigdata.fr/perceptron-machine-learning>
11. Nextpit [Internet]. #TBT - ELIZA : un des premiers chatbots de l'Histoire (1966). Disponible sur: <https://www.nextpit.fr/psychotherapeute-chatbot-eliza>
12. Richard R. 24pm academy. Eliza, le premier chatbot du monde en 1966. Disponible sur: <https://24pm.com/gpt/148-guide-de-l-intelligence-artificielle/chatbots/916-eliza-le-premier-chatbot-du-monde-en-1966>
13. Système expert. In: Larousse [Internet]. Disponible sur: https://www.larousse.fr/dictionnaires/francais/systeme_expert/76263
14. Shortliffe EH, Davis R, Axline SG, Buchanan BG, Green CC, Cohen SN. Computer-based consultations in clinical therapeutics: Explanation and rule acquisition capabilities of the MYCIN system. *Computers and Biomedical Research*. août 1975;8(4):303-20.
15. Dèbes F. *Les Echos*. 1973 : le premier « hiver » arrive sur l'IA. Disponible sur: <https://www.lesechos.fr/2017/08/1973-le-premier-hiver-arrive-sur-lia-181211>
16. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA*. avr 1982;79(8):2554-8.
17. Lee CS, Wang MH, Yen SJ, Wei TH, Wu IC, Chou PC, et al. Human vs. Computer Go: Review and Prospect [Discussion Forum]. *IEEE Comput Intell Mag*. août 2016;11(3):67-72.
18. Larousse [Internet]. Larousse - Définition d'un algorithme. Disponible sur: <https://www.larousse.fr/dictionnaires/francais/algorithme/2238>
19. Talevi A, Morales JF, Hather G, Podichetty JT, Kim S, Bloomingdale PC, et al. *Machine Learning in Drug Discovery and Development Part 1: A Primer*. *CPT Pharmacom & Syst Pharma*. mars 2020;9(3):129-42.
20. SAS institute. *Machine Learning - What it is and why it matters* [Internet]. Disponible sur: https://www.sas.com/en_us/insights/analytics/machine-learning.html
21. PraedictIA [Internet]. Qu'est-ce que l'apprentissage automatique (Machine Learning) ? Disponible sur: <https://praedictia.com/page/lapprentissage-machine/quest-ce.html>

22. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov.* juin 2019;18(6):463-77.
23. Niha A, Fahad U. Understanding deep learning — challenges and prospects. *Journal of the Pakistan Medical Association.*
24. Lo YC, Rensi SE, Tornig W, Altman RB. Machine learning in chemoinformatics and drug discovery. *Drug Discovery Today.* août 2018;23(8):1538-46.
25. Datatab.fr [Internet]. Analyse de régression. Disponible sur: <https://datatab.fr/tutorial/regression>
26. Schober P, Vetter TR. Linear Regression in Medical Research. *Anesthesia & Analgesia.* janv 2021;132(1):108-9.
27. Salzberg SL. C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. *Mach Learn.* sept 1994;16(3):235-40.
28. Bertrand S. Forêts aléatoires: De l'analyse des mécanismes de fonctionnement à la construction dynamique.
29. Freie Universität Berlin [Internet]. Random Forest.
30. Sarica A, Cerasa A, Quattrone A. Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. *Front Aging Neurosci.* 6 oct 2017;9:329.
31. Zhang Z. Introduction to machine learning: k-nearest neighbors. *Ann Transl Med.* juin 2016;4(11):218-218.
32. Zhang Z. Too much covariates in a multivariable model may cause the problem of overfitting. *J Thorac Dis.* sept 2014;6(9):E196-197.
33. Lantz B. Machine learning with R: expert techniques for predictive modeling. Third edition. Birmingham Mumbai: Packt Publishing; 2019. 437 p.
34. Gertrudes JC, Maltarollo VG, Silva RA, Oliveira PR, Honorio KM, Da Silva ABF. Machine Learning Techniques and Drug Design. *CMC.* 1 sept 2012;19(25):4289-97.
35. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *CGP* [Internet]. 2 janv 2018 [cité 16 févr 2024];15(1). Disponible sur: <http://cgp.iiarjournals.org/content/15/1/41.abstract>
36. Noble WS. What is a support vector machine? *Nat Biotechnol.* déc 2006;24(12):1565-7.
37. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol.* 15 août 2007;73(16):5261-7.
38. Domingos P, Pazzani M. On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. *Machine Learning.* 1997;29(2/3):103-30.
39. Carracedo-Reboredo P, Liñares-Blanco J, Rodríguez-Fernández N, Cedrón F, Novoa FJ, Carballedo A, et al. A review on machine learning approaches and trends in drug discovery. *Computational and Structural Biotechnology Journal.* 2021;19:4538-58.
40. Kondratovich E, Baskin II, Varnek A. Transductive Support Vector Machines: Promising Approach to Model Small and Unbalanced Datasets. *Molecular Informatics.* mars 2013;32(3):261-6.
41. Habehh H, Gohel S. Machine Learning in Healthcare. *CG.* 16 déc 2021;22(4):291-300.
42. MacCuish JD, MacCuish NE. Chemoinformatics applications of cluster analysis. *WIREs Comput Mol Sci.* janv 2014;4(1):34-48.
43. Goodswen SJ, Barratt JLN, Kennedy PJ, Kaufer A, Calarco L, Ellis JT. Machine learning and applications in microbiology. *FEMS Microbiology Reviews.* 8 sept 2021;45(5):fuab015.

44. Alloghani M, Al-Jumeily D, Mustafina J, Hussain A, Aljaaf AJ. A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. In: Berry MW, Mohamed A, Yap BW, éditeurs. *Supervised and Unsupervised Learning for Data Science* [Internet]. Cham: Springer International Publishing; 2020 [cité 16 févr 2024]. p. 3-21. (Unsupervised and Semi-Supervised Learning). Disponible sur: http://link.springer.com/10.1007/978-3-030-22475-2_1
45. Takeshima H. Deep Learning and Its Application to Function Approximation for MR in Medicine: An Overview. *MRMS*. 2022;21(4):553-68.
46. Montazeri M, Montazeri M, Montazeri M, Beigzadeh A. Machine learning models in breast cancer survival prediction. *THC*. 27 janv 2016;24(1):31-42.
47. Xiang F, Phuong T. An introduction to Convolutional Neural Network (CNN).
48. Kriegeskorte N, Golan T. Neural network models and deep learning. *Current Biology*. avr 2019;29(7):R231-6.
49. Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate. 2014 [cité 17 févr 2024]; Disponible sur: <https://arxiv.org/abs/1409.0473>
50. Gaudiaut T. Le Big Bang du Big Data [Image digitale] [Internet]. 2021. Disponible sur: <https://fr-statista-com.ezp.essec.fr/infographie/17800/big-data-evolution-volume-donnees-numeriques-genere-dans-le-monde/>
51. Parlement Européen. Mégadonnées : définition, avantages et défis (infographie). 17 févr 2021; Disponible sur: <https://www.europarl.europa.eu/news/fr/headlines/society/20210211STO97614/megadonnees-definition-avantages-et-defis-infographie>
52. Bourany T. Les 5V du big data: Regards croisés sur l'économie. 3 juin 2019;n° 23(2):27-31.
53. Gaudiaut T. Statista. Une minute sur Internet en 2020.
54. Gautier M. Statista. Les objets connectés - Faits et chiffres.
55. ANFR [Internet]. La 5G déployée en France - ANFR.
56. Zhang Y, Luo M, Wu P, Wu S, Lee TY, Bai C. Application of Computational Biology and Artificial Intelligence in Drug Design. *IJMS*. 5 nov 2022;23(21):13568.
57. Zhong F, Xing J, Li X, Liu X, Fu Z, Xiong Z, et al. Artificial intelligence in drug design. *Sci China Life Sci*. oct 2018;61(10):1191-204.
58. Batool M, Ahmad B, Choi S. A Structure-Based Drug Discovery Paradigm. *IJMS*. 6 juin 2019;20(11):2783.
59. Swinney DC, Anthony J. How were new medicines discovered? *Nat Rev Drug Discov*. juill 2011;10(7):507-19.
60. Jin S, Zeng X, Xia F, Huang W, Liu X. Application of deep learning methods in biological networks. *Briefings in Bioinformatics*. 22 mars 2021;22(2):1902-17.
61. Bravo À, Piñero J, Queralt-Rosinach N, Rautschka M, Furlong LI. Extraction of relations between genes and diseases from text and large-scale data analysis: implications for translational research. *BMC Bioinformatics*. déc 2015;16(1):55.
62. Sinkala M, Mulder N, Martin D. Machine Learning and Network Analyses Reveal Disease Subtypes of Pancreatic Cancer and their Molecular Characteristics. *Sci Rep*. 27 janv 2020;10(1):1212.
63. Wang T, Shao W, Huang Z, Tang H, Zhang J, Ding Z, et al. MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nat Commun*. 8 juin 2021;12(1):3445.
64. Xuan, Pan, Zhang, Liu, Sun. Graph Convolutional Network and Convolutional Neural Network Based Method for Predicting lncRNA-Disease Associations. *Cells*. 30 août 2019;8(9):1012.
65. You Y, Lai X, Pan Y, Zheng H, Vera J, Liu S, et al. Artificial intelligence in cancer

- target identification and drug discovery. *Sig Transduct Target Ther.* 10 mai 2022;7(1):156.
66. Jeon J, Nim S, Teyra J, Datti A, Wrana JL, Sidhu SS, et al. A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening. *Genome Med.* déc 2014;6(7):57.
67. Costa PR, Acencio ML, Lemke N. A machine learning approach for genome-wide prediction of morbid and druggable human genes based on systems-level data. *BMC Genomics.* déc 2010;11(S5):S9.
68. Koscielny G, An P, Carvalho-Silva D, Cham JA, Fumis L, Gasparyan R, et al. Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res.* 4 janv 2017;45(D1):D985-94.
69. Anderson AC. The Process of Structure-Based Drug Design. *Chemistry & Biology.* sept 2003;10(9):787-97.
70. Callaway E. 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. *Nature.* 10 déc 2020;588(7837):203-4.
71. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 26 août 2021;596(7873):583-9.
72. Pakhrin SC, Shrestha B, Adhikari B, Kc DB. Deep Learning-Based Advances in Protein Structure Prediction. *IJMS.* 24 mai 2021;22(11):5553.
73. Chen W, Liu X, Zhang S, Chen S. Artificial intelligence for drug discovery: Resources, methods, and applications. *Molecular Therapy - Nucleic Acids.* mars 2023;31:691-702.
74. Deepmind.google [Internet]. AlphaFold: a solution to a 50-year-old grand challenge in biology. Disponible sur: <https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>
75. INNEANCE [Internet]. AlphaFold a prédit la quasi-totalité des structures de protéines.
76. Wallach I, Dzamba M, Heifets A. AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery. 2015 [cité 18 mars 2024]; Disponible sur: <https://arxiv.org/abs/1510.02855>
77. Atomwise.com [Internet]. How AtomNet® Technology Improves Drug Design Using Convolutional Neural Networks. Disponible sur: https://blog.atomwise.com/introducing-atomnet-drug-design?__hstc=181257784.03dc3a4b99e4afb3e0924fc2e8c8bb5f.1604348830126.1634063012667.1634069536201.62&__hssc=181257784.6.1634069536201&__hsfp=1972113449
78. Atomwise Pipeline [Internet]. Disponible sur: <https://www.atomwise.com/pipeline/#rnd-pipeline>
79. Bayer Exercises Right to Enter Into Follow-On Research and Collaboration Option Agreement With Atomwise [Internet]. Disponible sur: <https://www.atomwise.com/2020/01/07/bayer-exercises-right-to-enter-into-follow-on-research-and-collaboration-option-agreement-with-atomwise/>
80. Atomwise.com [Internet]. GC Pharma Taps Small Molecule AI Expert Atomwise to Expand Hemophilia Franchise. Disponible sur: <https://www.atomwise.com/2020/09/08/gc-pharma-taps-atomwise-to-expand-hemophilia-franchise/>
81. Fierce Biotech [Internet]. Sanofi signs \$1.2B pact with Atomwise in latest high-value AI drug discovery deal. Disponible sur: <https://www.fiercebiotech.com/biotech/sanofi-signs-12b-pact-atomwise-latest-high-value-ai-drug-discovery-deal>
82. Segall MD, Barber C. Addressing toxicity risk when designing and selecting compounds in early drug discovery. *Drug Discovery Today.* mai 2014;19(5):688-93.
83. Onakpoya IJ, Heneghan CJ, Aronson JK. Worldwide withdrawal of medicinal products because of adverse drug reactions: a systematic review and analysis. *Critical Reviews in Toxicology.* 2 juill 2016;46(6):477-89.

84. Raies AB, Bajic VB. *In silico* toxicology: computational methods for the prediction of chemical toxicity. *WIREs Comput Mol Sci.* mars 2016;6(2):147-72.
85. Rim KT. In silico prediction of toxicity and its applications for chemicals at work. *Toxicol Environ Health Sci.* sept 2020;12(3):191-202.
86. Yang H, Sun L, Li W, Liu G, Tang Y. In Silico Prediction of Chemical Toxicity for Drug Design Using Machine Learning Methods and Structural Alerts. *Front Chem.* 20 févr 2018;6:30.
87. Mayr A, Klambauer G, Unterthiner T, Hochreiter S. DeepTox: Toxicity Prediction using Deep Learning. *Front Environ Sci [Internet].* 2 févr 2016 [cité 18 mars 2024];3. Disponible sur: <http://journal.frontiersin.org/Article/10.3389/fenvs.2015.00080/abstract>
88. Roy K, Kar S, Das RN. Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment. Amsterdam: Academic Press, an imprint of Elsevier; 2015.
89. Ma J, Sheridan RP, Liaw A, Dahl GE, Svetnik V. Deep Neural Nets as a Method for Quantitative Structure–Activity Relationships. *J Chem Inf Model.* 23 févr 2015;55(2):263-74.
90. Willighagen EL, Mayfield JW, Alvarsson J, Berg A, Carlsson L, Jeliaskova N, et al. The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching. *J Cheminform.* déc 2017;9(1):33.
91. Basile AO, Yahi A, Tatonetti NP. Artificial Intelligence for Drug Toxicity and Safety. *Trends in Pharmacological Sciences.* sept 2019;40(9):624-35.
92. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A Deep Learning Approach to Antibiotic Discovery. *Cell.* avr 2020;181(2):475-83.
93. Nag S, Baidya ATK, Mandal A, Mathew AT, Das B, Devi B, et al. Deep learning tools for advancing drug discovery and development. *3 Biotech.* mai 2022;12(5):110.
94. Lei T, Li Y, Song Y, Li D, Sun H, Hou T. ADMET evaluation in drug discovery: 15. Accurate prediction of rat oral acute toxicity using relevance vector machine and consensus modeling. *J Cheminform.* déc 2016;8(1):6.
95. Ye Z, Ouyang D. Prediction of small-molecule compound solubility in organic solvents by machine learning algorithms. *J Cheminform.* déc 2021;13(1):98.
96. Panapitiya G, Girard M, Hollas A, Sepulveda J, Murugesan V, Wang W, et al. Evaluation of Deep Learning Architectures for Aqueous Solubility Prediction. *ACS Omega.* 10 mai 2022;7(18):15695-710.
97. Korotcov A, Tkachenko V, Russo DP, Ekins S. Comparison of Deep Learning With Multiple Machine Learning Methods and Metrics Using Diverse Drug Discovery Data Sets. *Mol Pharmaceutics.* 4 déc 2017;14(12):4462-75.
98. Liu K, Sun X, Jia L, Ma J, Xing H, Wu J, et al. Chemi-Net: A Molecular Graph Convolutional Network for Accurate Drug Property Prediction. *IJMS.* 10 juill 2019;20(14):3389.
99. Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. *Sci Adv.* 6 juill 2018;4(7):eaap7885.
100. Mercado R, Rastemo T, Lindelöf E, Klambauer G, Engkvist O, Chen H, et al. Graph networks for molecular design. *Mach Learn: Sci Technol.* 1 juin 2021;2(2):025023.
101. Li Y, Zhang L, Liu Z. Multi-objective de novo drug design with conditional graph generative model. *J Cheminform.* déc 2018;10(1):33.
102. GlobeNewswire [Internet]. Insilico Medicine Successfully Discovered Potent, Selective, and Orally Bioavailable Small Molecule Inhibitor of CDK8 Using Generative AI. Disponible sur: <https://www.globenewswire.com/en/news-release/2023/04/24/2653269/31533/en/Insilico-Medicine-Successfully-Discovered-Potent-Selective-and-Orally-Bioavailable-Small-Molecule-Inhibitor-of-CDK8-Using-Generative-AI.html>

103. Ivanenkov YA, Polykovskiy D, Bezrukov D, Zagribelnyy B, Aladinskiy V, Kamyra P, et al. Chemistry42: An AI-Driven Platform for Molecular Design and Optimization. *J Chem Inf Model*. 13 févr 2023;63(3):695-701.
104. Yan C, Grabowska ME, Dickson AL, Li B, Wen Z, Roden DM, et al. Leveraging generative AI to prioritize drug repurposing candidates for Alzheimer's disease with real-world clinical validation. *npj Digit Med*. 26 févr 2024;7(1):46.
105. Harrer S, Shah P, Antony B, Hu J. Artificial Intelligence for Clinical Trial Design. *Trends in Pharmacological Sciences*. août 2019;40(8):577-91.
106. Helgeson J, Rammage M, Urman A, Roebuck MC, Coverdill S, Pomerleau K, et al. Clinical performance pilot using cognitive computing for clinical trial matching at Mayo Clinic. *JCO*. 20 mai 2018;36(15_suppl):e18598-e18598.
107. Weinberg AD, Chronic Disease Research Center Team. The EDICT Project: Policy Recommendations to Eliminate Disparities in Clinical Trials. 2008 [cité 25 avr 2024]; Disponible sur: <http://rgdoi.net/10.13140/RG.2.2.17044.40321>
108. Precision medicine requires a precision platform.
109. Blaschke TF, Osterberg L, Vrijens B, Urquhart J. Adherence to Medications: Insights Arising from Studies on the Unreliable Link Between Prescribed and Actual Drug Dosing Histories. *Annu Rev Pharmacol Toxicol*. 10 févr 2012;52(1):275-301.
110. Gregory Yauney, Pratik Shah. Reinforcement Learning with Action-Derived Rewards for Chemotherapy and Clinical Trial Dosing Regimen Selection [Internet]. Disponible sur: <http://proceedings.mlr.press/v85/yauney18a/yauney18a.pdf>
111. Bain EE, Shafner L, Walling DP, Othman AA, Chuang-Stein C, Hinkle J, et al. Use of a Novel Artificial Intelligence Platform on Mobile Devices to Assess Dosing Compliance in a Phase 2 Clinical Trial in Subjects With Schizophrenia. *JMIR Mhealth Uhealth*. 21 févr 2017;5(2):e18.
112. Ducruet C. Comment la pharmacie s'approprie le Big Data et l'intelligence artificielle. *Les Echos*. Disponible sur: <https://www.lesechos.fr/industrie-services/pharmacie-sante/comment-la-pharmacie-sapproprie-le-big-data-et-lintelligence-artificielle-1003161#:~:text=Le%20groupe%20français%20a%20par,optimiser%20la%20production%20de%20vaccins%20>.
113. Achim Plueckebaum. Medium. Unlocking the potential of data: lessons learned from leading data42. Disponible sur: <https://achimplueckebaum.medium.com/unlocking-the-potential-of-data-lessons-learned-from-leading-data42-9726c035d05a>
114. Pharmacos media [Internet]. L'industrie pharmaceutique en pleine mutation. Disponible sur: <https://www.pharmacos-media.fr/analyse/lindustrie-pharmaceutique-en-pleine-mutation-55553.html>
115. Fraiser K. Fierce Pharma. Amgen expands pact with Amazon to usher drug manufacturing into the AI era. Disponible sur: <https://www.fiercepharma.com/pharma/amgen-expands-pact-amazon-web-services-usher-manufacturing-ai-era>
116. Gosset O. Rombio, l'assistant intelligent qui dope la production pharmaceutique. *L'Echo*. Disponible sur: <https://www.lecho.be/entreprises/pharma-biotechnologie/rombio-l-assistant-intelligent-qui-dope-la-production-pharmaceutique/10482968.html>
117. Boulton C. Merck injecte de l'IA dans sa supply chain. *Le Monde Informatique*. Disponible sur: <https://www.lemondeinformatique.fr/actualites/lire-merck-injecte-de-l-ia-dans-sa-supply-chain-77538.html>
118. Boulton C. Réseaux-telecoms. Merck compte sur l'IA pour gérer la complexité de la supply chain.
119. France Assos santé. Pénuries de médicaments et de vaccins [Internet]. Disponible sur: <https://www.france-assos-sante.org/wp-content/uploads/2019/02/Penuries-medicaments-Resultats-BVA-dec2018.pdf>

120. Apodis-Sante [Internet]. Apodis Santé. Disponible sur: <https://www.apodis-sante.com/solution-ruptures-appvisionnement/>
121. Rapport du Sénat n°828 sur la pénurie de médicaments et les choix de l'industrie pharmaceutique française [Internet]. Disponible sur: <https://www.senat.fr/rap/r22-828-1/r22-828-11.pdf>
122. Espesson-Vergeat B. L'intelligence artificielle et la blockchain au service de la sécurisation logistique des produits dans le secteur pharmaceutique. Droit, Santé et Société 2021/2 (N° 2) [Internet]. Disponible sur: <https://www.cairn.info/revue-droit-sante-et-societe-2021-2-page-50.htm?contenu=article>
123. Sanofi [Internet]. La lutte contre la falsification de médicaments. Disponible sur: <https://www.sanofi.com/fr/magazine/impact-social/lutter-contre-les-medicaments-falsifies>
124. Leem les entreprises du médicament [Internet]. 100 Questions : Le médicament et sa mise au point. Disponible sur: <https://www.leem.org/sites/default/files/questionpdf/boite-de-medicament.pdf>
125. Techpoint Africa [Internet]. Les startups africaines des technologies de la santé luttent contre les médicaments contrefaits grâce à la blockchain et à l'IA. Disponible sur: <https://techpoint.africa/fr/2023/06/05/african-healthtech-startups-combat-counterfeit-drugs/#:~:text=Technologie%20mobile&text=mPedigree%2C%20une%20startup%20ghanéenne%2C%20utilise,de%20données%20centralisée%20pour%20vérification.>
126. Rxall.net [Internet]. Rx Scanner :Digital Solution to fight fake drugs. Disponible sur: <https://rxall.net/rxscanner/>
127. Basile AO, Yahi A, Tatonetti NP. Artificial Intelligence for Drug Toxicity and Safety. Trends in Pharmacological Sciences. sept 2019;40(9):624-35.
128. Martin GL, Jouganous J, Savidan R, Bellec A, Goehrs C, Benkebil M, et al. Validation of Artificial Intelligence to Support the Automatic Coding of Patient Adverse Drug Reaction Reports, Using Nationwide Pharmacovigilance Data. Drug Saf. mai 2022;45(5):535-48.
129. Prix Galien France [Internet]. Synapse Medicine, une start-up qui s'intéresse à la santé de tous les français. Disponible sur: <https://prixgalien.fr/laureate/synapse-medicine-une-start-up-qui-sinteresse-a-la-sante-de-tous-les-francais/>
130. Business wire [Internet]. Aktana annonce le lancement d'outils d'orchestration omnicanal de prochaine génération conçus pour les medtechs. Disponible sur: <https://www.businesswire.com/news/home/20240305164330/fr/>
131. Richard Thomson, Dagmar Luettel, Frances Healey, Sarah Scobie. Safer care for the acutely ill patient: learning from serious incidents [Internet]. (The fifth report from the Patient Safety Observatory). Disponible sur: https://www.baccn.org/static/uploads/resources/NPSA_Report.pdf
132. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature. 1 août 2019;572(7767):116-9.
133. Connell A, Black G, Montgomery H, Martin P, Nightingale C, King D, et al. Implementation of a Digitally Enabled Care Pathway (Part 2): Qualitative Analysis of Experiences of Health Care Professionals. J Med Internet Res. 15 juill 2019;21(7):e13143.
134. Akselrod-Ballin A, Chorev M, Shoshan Y, Spiro A, Hazan A, Melamed R, et al. Predicting Breast Cancer by Applying Deep Learning to Linked Health Records and Mammograms. Radiology. août 2019;292(2):331-42.
135. Center for Artificial Intelligence in Medicine & Imaging [Internet]. RSNA 2017: Rads who use AI will replace rads who don't. Disponible sur: <https://aimi.stanford.edu/news/rsna-2017-rads-who-use-ai-will-replace-rads-who-dont-0>
136. Collins FS, Varmus H. A New Initiative on Precision Medicine. N Engl J Med. 26 févr 2015;372(9):793-5.

137. Johnson KB, Wei W, Weeraratne D, Frisse ME, Misulis K, Rhee K, et al. Precision Medicine, AI, and the Future of Personalized Health Care. *Clinical Translational Sci.* janv 2021;14(1):86-93.
138. France cohortes [Internet]. CONSTANCES : Cohorte des consultants des centres d'examens de santé. Disponible sur: <https://francecohortes.org/cohortes-associees/constances>
139. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* oct 2018;562(7726):203-9.
140. Liney T, Mitov A, Laurent Onivogui G, Arrighi N. Les biobanques, des structures essentielles à la recherche médicale. *Med Sci (Paris).* mars 2020;36(3):274-6.
141. PR Newswire [Internet]. IBM and Partners to Transform Personal Health with Watson and Open Cloud. Disponible sur: <https://www.prnewswire.com/news-releases/ibm-and-partners-to-transform-personal-health-with-watson-and-open-cloud-300065025.html>
142. McGinnis JM, Williams-Russo P, Knickman JR. The Case For More Active Policy Attention To Health Promotion. *Health Affairs.* mars 2002;21(2):78-93.
143. Sengupta S, Pomeranz Krummel D, Pomeroy S. The evolution of medulloblastoma therapy to personalized medicine. *F1000Res.* 13 avr 2017;6:490.
144. Hoang DH, Pagnier A, Guichardet K, Dubois-Teklali F, Schiff I, Lyard G, et al. Cognitive disorders in pediatric medulloblastoma: what neuroimaging has to offer: A review. *PED.* août 2014;14(2):136-44.
145. Trivizakis E, Papadakis G, Souglakos I, Papanikolaou N, Koumakis L, Spandidos D, et al. Artificial intelligence radiogenomics for advancing precision and effectiveness in oncologic care (Review). *Int J Oncol.* 11 mai 2020;57(1):43-53.
146. Zhu Z, Albadawy E, Saha A, Zhang J, Harowicz MR, Mazurowski MA. Deep learning for identifying radiogenomic associations in breast cancer. *Computers in Biology and Medicine.* juin 2019;109:85-90.
147. Huang C, Clayton EA, Matyunina LV, McDonald LD, Benigno BB, Vannberg F, et al. Machine learning predicts individual cancer patient responses to therapeutic drugs with high accuracy. *Sci Rep.* 6 nov 2018;8(1):16444.
148. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, et al. Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. *Circ Res.* 13 oct 2017;121(9):1092-101.
149. Zhao J, Feng Q, Wu P, Lupu RA, Wilke RA, Wells QS, et al. Learning from Longitudinal Data in Electronic Health Record and Genetic Data to Improve Cardiovascular Event Prediction. *Sci Rep.* 24 janv 2019;9(1):717.
150. Beckmann E, Peyrou B, Gallay L, Vignaux JJ. L'intelligence artificielle à l'aide des myologues: Le point de vue d'un non robot. *Med Sci (Paris).* nov 2017;33:39-45.
151. SciencesPo chaire santé. Enjeux de l'intelligence artificielle en santé, Février 2023 [Internet]. Disponible sur: <https://doi.org/10.1051/medsci/201733s108>
152. HAS-sante.fr [Internet]. Qu'est-ce que l'éducation thérapeutique du patient ? Disponible sur: https://www.has-sante.fr/jcms/r_1496895/fr/education-therapeutique-du-patient-etp
153. Fédération Française des Diabétiques [Internet]. Les chiffres du diabète en France. Disponible sur: <https://www.federationdesdiabetiques.org/information/diabete/chiffres-france>
154. PKVitality [Internet]. K'Watch Glucose Discreet and Painless. Disponible sur: <https://www.pkvitality.com/ktrack-glucose/>
155. Ficca S. Clubic. K'Watch Glucose : la montre connectée qui surveille votre taux de glucose. Disponible sur: <https://www.clubic.com/mobilite-et-telephonie/objets-connectes/montre-connectee/actualite-847857-watch-glucose-connectee-surveille-continu-taux-glucose.html>

156. Unlocking precision diabetes care with AI and digital twin technology [Internet]. Disponible sur: <https://www.clubic.com/mobilite-et-telephonie/objets-connectes/montre-connectee/actualite-847857-watch-glucose-connectee-surveille-continu-taux-glucose.html>
157. Mallya [Internet]. Mallya : Liste des produits. Disponible sur: <https://my-mallya.com/product-list/aire-therapeutique/diabete/>
158. Martineau JT, Godin FR. Tour d’horizon des enjeux éthiques liés à l’IA en santé. *ethiquepublique* [Internet]. 27 sept 2023 [cité 7 avr 2024];(vol. 25, n° 1). Disponible sur: <http://journals.openedition.org/ethiquepublique/7978>
159. WHO guidance. Ethics and governance of artificial intelligence for health [Internet]. Disponible sur: <https://iris.who.int/bitstream/handle/10665/341996/9789240029200-eng.pdf?sequence=1>
160. Hand DJ. Aspects of Data Ethics in a Changing World: Where Are We Now? *Big Data*. sept 2018;6(3):176-90.
161. Andrejevic M. Big Data, Big Questions| The Big Data Divide.
162. Beauchamp TL. Informed Consent: Its History, Meaning, and Present Challenges. *Camb Q Healthc Ethics*. oct 2011;20(4):515-23.
163. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association*. 1 mars 2020;27(3):491-7.
164. Basso T, Matsunaga R, Moraes R, Antunes N. Challenges on Anonymity, Privacy, and Big Data. In: 2016 Seventh Latin-American Symposium on Dependable Computing (LADC) [Internet]. Cali, Colombia: IEEE; 2016 [cité 7 avr 2024]. p. 164-71. Disponible sur: <http://ieeexplore.ieee.org/document/7781852/>
165. Breidbach CF, Davern M, Shanks G, Asadi-Someh I. On the Ethical Implications of Big Data in Service Systems. In: Maglio PP, Kieliszewski CA, Spohrer JC, Lyons K, Patrício L, Sawatani Y, éditeurs. *Handbook of Service Science, Volume II* [Internet]. Cham: Springer International Publishing; 2019 [cité 7 avr 2024]. p. 661-74. (Service Science: Research and Innovations in the Service Economy). Disponible sur: https://link.springer.com/10.1007/978-3-319-98512-1_29
166. the Precise4Q consortium, Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak*. déc 2020;20(1):310.
167. Birhane A. Algorithmic injustice: a relational ethics approach. *Patterns*. févr 2021;2(2):100205.
168. Coeckelbergh M. Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Sci Eng Ethics*. août 2020;26(4):2051-68.
169. Mittelstadt BD, Floridi L. The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Sci Eng Ethics*. avr 2016;22(2):303-41.
170. Someh I, Davern M, Breidbach CF, Shanks G. Ethical Issues in Big Data Analytics: A Stakeholder Perspective. *CAIS*. 2019;718-47.
171. cnil.fr [Internet]. Sanctions : quelles sanctions peuvent être prononcées par la CNIL ? Disponible sur: <https://www.cnil.fr/fr/cnil-direct/question/sanctions-quelles-sanctions-peuvent-etre-prononcees-par-la-cnil#:~:text=Une%20suspension%20partielle%20ou%20totale,annuel%20mondial%20de%20la%20société>.
172. Parlement européen [Internet]. Loi sur l’IA de l’UE : première réglementation de l’intelligence artificielle. Disponible sur: <https://www.europarl.europa.eu/topics/fr/article/20230601STO93804/loi-sur-l-ia-de-l-ue-premiere-reglementation-de-l-intelligence-artificielle>
173. Snitem [Internet]. Logiciels et Intelligence Artificielle (IA) en santé : des dispositifs

- médicaux comme les autres ? Disponible sur: <https://www.snitem.fr/wp-content/uploads/2020/01/DM-IA-Logiciel-standard.pdf>
174. Promé G. Règlements et recommandations pour les IA médicales. Disponible sur: <https://www.qualitiso.com/reglementation-recommandations-ia-medicales/>
175. Walker-Osborn C, Asaad N. Key legal considerations in accelerating pharma using artificial intelligence. Pharmafile [Internet]. Disponible sur: <https://pharmafile.com/features/key-legal-considerations-accelerating-pharma-using-artificial-intelligence/>
176. Arrêté du 12 mai 2022 relatif aux données alimentant la base principale et aux bases de données du catalogue du système national des données de santé [Internet]. Disponible sur: <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000045796455>
177. Connected doctors [Internet]. Le Health Data Hub ou le risque d'une santé marchandisée. Disponible sur: <https://connectedoctors.fr/sante-connectee/connected-doctors/2020/04/health-data-hub-le-risque-dune-sante-marchandisee/>
178. Kheops Technologies. Comment HL7 révolutionne l'échange de données médicales dans le système de soins Suisse. Disponible sur: <https://www.linkedin.com/pulse/comment-hl7-revolutionne-lechange-de-donnees-medicales-eejlf/>
179. Les Echos. Numéro de sécu, mutuelle : 33 millions de Français victimes d'une cyberattaque au tiers payant. Disponible sur: <https://www.lesechos.fr/industrie-services/pharmacie-sante/numero-de-secu-mutuelle-33-millions-de-francais-victimes-dune-cyberattaque-au-tiers-payant-2074842>
180. Groupe Pasteur Mutualité [Internet]. Les cyber-attaques à l'hôpital. Disponible sur: <https://www.gpm.fr/cyberattaques-hopital-liberal/>
181. Ahmad MA, Eckert C, Teredesai A. Interpretable Machine Learning in Healthcare. In: Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics [Internet]. Washington DC USA: ACM; 2018 [cité 7 avr 2024]. p. 559-60. Disponible sur: <https://dl.acm.org/doi/10.1145/3233547.3233667>
182. Branellec G, Hadoussa S. Intelligence artificielle et santé, enjeux managériaux, juridiques et éthiques.
183. McLean S, Read GJM, Thompson J, Baber C, Stanton NA, Salmon PM. The risks associated with Artificial General Intelligence: A systematic review. *Journal of Experimental & Theoretical Artificial Intelligence*. 4 juill 2023;35(5):649-63.
184. Danaher J. Will Life Be Worth Living in a World Without Work? Technological Unemployment and the Meaning of Life. *Sci Eng Ethics*. févr 2017;23(1):41-64.
185. Smids J, Nyholm S, Berkers H. Robots in the Workplace: a Threat to—or Opportunity for—Meaningful Work? *Philos Technol*. sept 2020;33(3):503-22.
186. Van Wynsberghe A. Sustainable AI: AI for sustainability and the sustainability of AI. *AI Ethics*. août 2021;1(3):213-8.

Hamtache Inès

L'intelligence artificielle dans l'industrie des produits de santé : de la recherche à la commercialisation

Résumé :

L'avènement de l'intelligence artificielle (IA) sculpte un nouveau paysage dans notre société, et l'industrie des produits de santé n'est pas en reste. Dans un contexte de profusion des données de santé, l'IA se positionne comme l'outil clé pour les transformer en connaissances exploitables, améliorant ainsi la santé individuelle et collective.

Cette thèse plonge au cœur de cette technologie, depuis ses fondements théoriques jusqu'à ses applications concrètes dans l'industrie des produits de santé et les enjeux qui en découlent. Dans un premier temps, nous définissons la notion d'intelligence artificielle, explorons ses bases théoriques et l'importance des mégadonnées (*big data*) dans la génération d'algorithmes d'IA. Ensuite, nous présentons les applications de l'IA en santé, d'abord à travers le prisme du cycle du médicament, de la recherche préclinique à la commercialisation, démontrant les gains d'efficacité et de rentabilité apportés par l'IA. Nous mettons en lumière la manière dont l'IA érige les fondations d'une médecine du futur, préventive et personnalisée. Enfin, nous soulevons des questions essentielles quant aux défis liés à l'intégration de l'IA dans les industries de santé, allant des enjeux éthiques aux obstacles réglementaires et techniques, en passant par les implications économiques et sociétales.

Mots clés :

- **Intelligence artificielle**
- **Machine Learning**
- **Deep Learning**
- **Big Data**
- **Cycle du médicament**
- **Médecine 4P**