

Republique Algerienne Democratique et Populaire
ministere de l'enseignement superieur et de la recherche scientifique
universite ammar thelidji laghouat
faculte de medecine



Mémoire de fin d'étude pour l'obtention du diplôme de docteur en médecine

**L'ÉVALUATION DES PERFORMANCES DIAGNOSTIQUES DE
DEUX MODÈLES D'INTELLIGENCE ARTIFICIELLE (MAIA AI
ET GPT-5.2) DANS L'INTERPRÉTATION DE LA
RADIOGRAPHIE THORACIQUE :
ÉTUDE COMPARATIVE À L'HÔPITAL MIXTE DE LAGHOUAT**

PRESENTE PAR :

GACEM FATIMA WAFAA

GHERIBI IMANE

ENCADRÉ PAR :

DR LITIM ATIF

EXAMINÉ PAR :

ZAABTA ABDELKARIM

Année universitaire : 2025/2026

REMERCIEMENTS

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

و الحمد لله حمدا كثيرا طيبا مباركا فيه

Nous adressons nos sincères remerciements à notre encadrant **DR LITIM** pour sa disponibilité, ses conseils éclairés et la rigueur scientifique dont il a fait preuve tout au long de l'élaboration de ce travail. Sa guidance a été déterminante dans la réalisation de cette étude.

Nous sommes profondément reconnaissants envers les membres du jury, qu'ont accepté d'évaluer ce travail et de nous faire l'honneur de leur expertise, et au **DR ZAABTA** pour sa participation à ce jury.

Nos remerciements vont également à l'équipe médicale du service de pneumologie de l'Hôpital Mixte de Laghouat, pour sa participation active à l'étude, son accueil chaleureux et le temps consacré à nos échanges cliniques.

Nous rendons hommage à l'ensemble du corps enseignant de la Faculté de Médecine de l'Université Amar Telidji de Laghouat, dont les enseignements, l'encadrement et l'exigence académique ont forgé notre formation médicale.

Nos remerciements s'adressent également à l'administration de la Faculté de Médecine et à celle de l'Hôpital Mixte de Laghouat, pour leur soutien logistique et leur accueil bienveillant tout au long de notre cursus et de la réalisation de ce travail.

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Dédicace

Louange à Allah qui nous a accordé la réussite pour mener à bien ce travail, que nous espérons être une contribution précieuse et un bienfait utile.

Je dédie ce travail,

À mes parents, mon père et ma mère, qui m'ont accompagné pas à pas tout au long du parcours éducatif et académique, sans lassitude ni ennui, et qui ont été ainsi les artisans de cette réussite, les premiers à avoir semé la graine de ce fruit.

À mon frère Oussama, pour son soutien constant et dévoué, ainsi qu'à tous mes chers frères et chères sœurs, Noussaiba, Abderrahmane, Roufaïda et Mohamed Amine.

À ma grande famille, à mes proches, à ma binôme Wafa, et à mes enseignements.

À l'âme de nos pieux martyrs, artisans de la gloire et de la liberté de l'Algérie EL-Mahroussa

À l'âme du docteur Adnan Al-Barech, martyr de l'humanité et du devoir médical.

Au docteur Hossam Abu Safiya, gardien de l'humanité, d'un courage et d'un dévouement professionnel sans égal.

GHERIBI Imane.

Dédicaces



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
الْحَمْدُ لِلَّهِ حَمْدًا كَثِيرًا طَيِّبًا مُبَارَكًا فِيهِ

*À CEUX DONT LES PRIÈRES DEMEURENT LE SECRET DE MA
RÉUSSITE, À LA SOURCE INTARISSABLE DE TENDRESSE ET DE
GÉNÉROSITÉ,*

MON PÈRE ET MA MÈRE...
QUE DIEU LES PROTÈGE ET LEUR PRÊTE LONGUE VIE.

*À CELLES QUI ONT CHEMINÉ À MES CÔTÉS ET M'ONT SOUTENUE
SANS FAILLIR*

ASMÂA, SAFAA, MARWA, NOUR EL HOUDA,
AINSI QU'À TOUTE MA FAMILLE.

*À MA BINOME IMANE, MA PLUS FIDÈLE COMPAGNE TOUT AU
LONG DE CE TRAVAIL.*

*À CEUX QUI M'ONT TRANSMIS LE SAVOIR GRÂCE AUQUEL JE ME
SUIS ÉLEVÉE,*

À MES CHERS ENSEIGNANTS,
*QUE DIEU LES RÉCOMPENSE EN MA FAVEUR DE LA PLUS BELLE
DES RÉCOMPENSES.*

*À CELLES DONT LA PRÉSENCE A RENDU CE CHEMIN PLUS
LÉGER ET PLUS SEREIN :*

IMANE, KHADIDJA ET WADAA ET TOUS MES AMIS
*À TOUS CEUX QUI M'ONT ACCOMPAGNÉE D'UNE PAROLE
BIENVEILLANTE ET D'UNE PRIÈRE SINCÈRE,*

WAFAA GACEM

RESUME

L'intelligence artificielle occupe une place croissante en imagerie médicale, notamment dans l'interprétation des radiographies thoraciques. L'objectif de ce travail était d'évaluer et de comparer les performances diagnostiques de deux modèles d'intelligence artificielle, GPT-5.2 et MAIA AI, dans l'identification de la pneumonie, du pneumothorax, de l'épanchement pleural et des clichés normaux. Il s'agissait d'une étude observationnelle analytique à recueil mixte, rétrospectif et prospectif, incluant 110 radiographies thoraciques. Les résultats des deux modèles ont été comparés à l'interprétation d'un pneumologue expert, considérée comme gold standard. MAIA AI a démontré des performances diagnostiques supérieures pour la pneumonie et le pneumothorax, tandis que les deux modèles présentaient des performances comparables dans la détection de l'épanchement pleural et des radiographies thoraciques normales. Ces résultats suggèrent que l'intelligence artificielle représente un outil prometteur pour l'aide au diagnostic radiologique, mais qu'elle doit être utilisée en complément de l'expertise humaine.

MOTS CLES : Intelligence artificielle, radiographie thoracique, gold standard, maia.ai, GPT 5.2, épanchement pleurale, pneumonie, pneumothorax, radiographie normale.

ملخص

شغل الذكاء الاصطناعي مكانةً متزايدة الأهمية في مجال التصوير الطبي، ولا سيما في تفسير الصور الشعاعية للصدر. هدفت هذه الدراسة إلى تقييم ومقارنة الأداء التشخيصي لنموذجين من نماذج الذكاء الاصطناعي، نموذج **GPT 5.2** ونموذج **MAIA AI** في التعرف على الالتهاب الرئوي، والاسترواح الصدري، والانصباب الجنبى، والصور الشعاعية الصدرية الطبيعية. أجريت هذه الدراسة وفق منهج رسدي تحليلي بجمع بيانات مختلف بين الأثر الرجعي والاستباقي، وشملت 110 صورة شعاعية للصدر. تمت مقارنة نتائج النموذجين بتفسير طبيب متخصص في أمراض الجهاز التنفسي ذي خبرة عالية، والذي اعتُبر معياراً للمقارنة. أظهر نموذج مايا أداءً تشخيصياً متفوقاً في تشخيص الالتهاب الرئوي والاسترواح الصدري، في حين أظهر النموذجان أداءً متقارباً في الكشف عن الانصباب الجنبى والصور الشعاعية الصدرية الطبيعية. تشير هذه النتائج إلى أن الذكاء الاصطناعي يمثل أداةً واعدة للمساعدة في التشخيص الإشعاعي، غير أنه ينبغي توظيفه بوصفه مكملاً للخبرة البشرية لا بديلاً عنها.

الكلمات المفتاحية: الذكاء الاصطناعي، الأشعة السينية للصدر، المعيار الذهبي، الانصباب الجنبى، الالتهاب الرئوي، استرواح الصدر، صورة الأشعة السينية الطبيعية للصدر

ABSTRACT

Artificial intelligence is playing an increasingly important role in medical imaging, particularly in the interpretation of chest radiographs. The aim of this study was to evaluate and compare the diagnostic performance of two artificial intelligence models, GPT-5.2 and MAIA AI, in identifying pneumonia, pneumothorax, pleural effusion, and normal chest X-rays. This was an analytical observational study with mixed retrospective and prospective data collection, including 110 chest radiographs. The results from both models were compared to the interpretation of an expert pulmonologist, which was considered the gold standard. MAIA AI demonstrated superior diagnostic performance for pneumonia and pneumothorax, while both models showed comparable performance in detecting pleural effusion and normal chest radiographs. These findings suggest that artificial intelligence represents a promising tool for aiding radiological diagnosis, but it should be used to complement human expertise.

KEYWORDS: Artificial intelligence, chest X-ray, gold standard, maia.ai, GPT 5.2, pleural effusion, pneumonia, pneumothorax, normal chest X-ray

TABLE DES MATIÈRES :

LISTE DES ABRÉVIATIONS.....	13
LISTE DES TABLEAUX	14
LISTE DES FIGURES.....	16
INTRODUCTION	18
PROBLÉMATIQUE	21

LA PREMIÈRE PARTIE : PARTIE THÉORIQUE

CHAPITRE I : L'INTELLIGENCE ARTIFICIELLE

1. DÉFINITION	23
2. HISTOIRE DE L'INTELLIGENCE ARTIFICIELLE	23
3. L'APPRENTISSAGE AUTOMATIQUE (MACHINE LEARNING)	29
4. L'APPRENTISSAGE PROFOND (DEEP LEARNING)	32
5. LES RÉSEAUX DE NEURONES ARTIFICIELS (RNA)	34
6. LES RÉSEAUX DE NEURONES CONVOLUTIFS (CNN)	36
7. MÉDECINE ET INTELLIGENCE ARTIFICIELLE	38

CHAPITRE II : RADIOGRAPHIE THORACIQUE STANDARD

1. GÉNÉRALITÉS SUR LE TÉLÉTHORAX	46
2. ANATOMIE RADIOLOGIQUE THORACIQUE NORMALE	47
3. SÉMIOLOGIE RADIOLOGIQUE THORACIQUE DE BASE	49
4. INDICATIONS CLINIQUES VALIDÉES : DU DEPISTAGE AU DIAGNOSTIC	52

CHAPITRE III – L'IA ET L'INTERPRÉTATION DE LA RADIOGRAPHIE THORACIQUE

1. FONCTIONNEMENT DES ALGORITHMES DÉDIÉS AU TÉLÉTHORAX	54
2. PATHOLOGIES DÉTECTÉES PAR L'IA SUR TÉLÉTHORAX	54
3. PERFORMANCES DE L'IA DANS LES ÉTUDES PUBLIÉES	56
4. LIMITES ET DÉFIS DE L'IA APPLIQUÉE AU TÉLÉTHORAX	57

5. PLACE DE L'IA DANS LE CONTEXTE CLINIQUE : RÔLE D'AIDE À LA DÉCISION ET NON DE REMPLACEMENT	58
6. MÉTHODES D'ÉVALUATION ET VALIDATION DES MODELES D'INTELLIGENCE ARTIFICIELLE EN RADIOGRAPHIE THORACIQUE	59

LA DEUXIÈME PARTIE : PARTIE PRATIQUE

I. MATÉRIEL ET MÉTHODES :

1. TYPE DE L'ETUDE	64
2. LIEU DE L'ETUDE	64
3. DURÉE DE L'ÉTUDE	64
4. LES OBJECTIFS DE L'ETUDE	64
5. POPULATION D'ETUDE	64
6. COLLECTE DES DONNEES	65
7. LES VARIABLES ÉTUDIÉES	65
8. CONSTITUTION DU STANDARD DE REFERENCE	68
9. STANDARDISATION DU PROMPT	68
10. DESCRIPTION DES MODÈLES D'IA ÉVALUÉS	68
11. PLAN D'ANALYSE STATISTIQUE DES DONNEES	69
12. CONSIDÉRATIONS ÉTHIQUES	70

II. RESULTATS :

1. FLUX DE SELECTION DES DONNEES	71
2. CARACTERISTIQUES DE LA POPULATION ET DU JEU DE DONNEES	71
3. ETUDE DES PERFORMANCES DIAGNOSTIC DU MODELE MAIA.IA : SENSIBILITE, SPECIFICITE, EXACTITUDE, VPP, VPN, COURBES ROC (AUC)	73
• DIAGNOSTIC DE L'ÉPANCHEMENT PLEURAL	73
• DIAGNOSTIC DE PNEUMONIE	75
• DIAGNOSTIC PNEUMOTHORAX	77
• DIAGNOSTIC DES CLICHES NORMAUX	80
• EVALUATION DE PERFORMANCE DIAGNOSTIC GLOBALE DE MAIA.IA	82
4. ETUDE DES PERFORMANCES DIAGNOSTIC DU MODELE GPT 5.2: SENSIBILITE, SPECIFICITE, EXACTITUDE, VPP, VPN, COURBES ROC (AUC).....	84
• DIAGNOSTIC DE L'ÉPANCHEMENT PLEURAL	84
• DIAGNOSTIC DE PNEUMONIE	86
• DIAGNOSTIC PNEUMOTHORAX	88
• DIAGNOSTIC DES CLICHES NORMAUX	91
• EVALUATION DE PERFORMANCE DIAGNOSTIC GLOBALE DE GPT 5.2	93

5. ANALYSE DE CONCORDANCE	93
• CONCORDANCE DIAGNOSTIQUE	93
• CONCORDANCE SÉMIOLOGIQUE	96
6. COMPARAISON INTER-MODÈLES	98
• COMPARAISON DES PERFORMANCES DIAGNOSTIQUES GLOBALES	98
• COMPARAISON DE CONCORDANCE DIAGNOSTIQUE	104
• COMPARAISON DE CONCORDANCE SEMIOLOGIQUE	105
7. ANALYSE DES ERREURS DIAGNOSTIQUES	106
III. DISCUSSION	107
IV. CONCLUSION	111
V. REFERENCES	113
VI. ANNEXES	117

TABLEAU DES ABREVIATIONS :

AUC	Area Under the Curve (Aire sous la courbe)
CNN	Réseau de neurones convolutif
DL	Deep Learning (Apprentissage profond)
EPL	Épanchement pleural liquidien
FN	Faux négatif
FP	Faux positif
GPT	Generative Pre-trained Transformer
GRU	Gated Recurrent Units (Unités récurrentes fermées)
GS	Gold Standard (Référence standard)
IA	Intelligence artificielle
IC	Intervalle de confiance
LLM	Large Language Models (Grands modèles de langage)
LSTM	Long Short-Term Memory (Mémoire à long et court terme)
ML	Machine Learning (Apprentissage automatique)
PNE	Pneumonie
PNO	Pneumothorax
RNA	Réseau de neurones artificiels
RNN	Réseau de neurones récurrents
ROC	Receiver Operating Characteristic (Courbe caractéristique de fonctionnement du récepteur)
RV +	Rapport de vraisemblance positif
RV -	Rapport de vraisemblance négatif
SADC	Systèmes d'aide à la décision clinique
TLT	Téléthorax
VN	Vrai négatif
VP	Vrai positif
VPN	Valeur prédictive négative
VPP	Valeur prédictive positive

LISTE DES TABLEAUX :

Tableau 1 : Vue d'ensemble des métriques de classification courantes

Tableau 2. Caractéristiques des deux modèles d'IA évalués

Tableau 3. Caractéristiques générales de la population d'étude.

Tableau 4. Répartition du jeu de données selon le diagnostic de référence.

Tableau 5. Matrice de confusion – MAIA AI, Épanchement pleural.

Tableau 6. Performances diagnostiques – MAIA AI, Épanchement pleural

Tableau 7. Matrice de confusion – MAIA AI, Pneumonie

Tableau 8. Performances diagnostiques – MAIA AI, Pneumonie

Tableau 9. Matrice de confusion – MAIA AI, Pneumothorax

Tableau 10. Performances diagnostiques – MAIA AI, Pneumothorax.

Tableau 11. Matrice de confusion – MAIA AI, Aspect normal.

Tableau 12. Performances diagnostiques – MAIA AI, Aspect normal.

Tableau 13. Performances diagnostiques – MAIA AI.

Tableau 14. Matrice de confusion – GPT-5.2, Épanchement pleural.

Tableau 15. Performances diagnostiques – GPT-5.2, Épanchement pleural

Tableau 16. Matrice de confusion – GPT-5.2, Pneumonie.

Tableau 17. Performances diagnostiques – GPT-5.2, Pneumonie.

Tableau 18. Matrice de confusion – GPT-5.2, Pneumothorax.

Tableau 19. Performances diagnostiques – GPT-5.2, Pneumothorax.

Tableau 20. Matrice de confusion – GPT-5.2, Aspect normal.

Tableau 21. Performances diagnostiques – GPT-5.2, Aspect normal.

Tableau 22. Performances diagnostiques globales –GPT 5.2-.

Tableau 23. Concordances diagnostiques de -MAIA.IA-.

Tableau 24. Concordances diagnostiques de -GPT 5.2 -.

Tableau 25. Concordances sémiologiques -MAIA.IA-.

Tableau 26. Concordances sémiologiques -GPT 5.2-.

Tableau 27. Tableau comparatif des performances diagnostique des 2 modèles MAIA.IA et GPT 5.2.

Tableau 28. Coefficient Kappa de Cohen par entité diagnostique pour chaque modèle d'IA.

Tableau 29. Répartition des faux positifs (FP) et faux négatifs (FN) par entité et par modèle.

LISTE DES FIGURES :

Figure 01. L'évolution historique de l'intelligence artificielle.

Figure 02. Les sciences fondamentales constitutives de l'intelligence artificielle.

Figure 03. Impact du Big Data sur la vision par ordinateur.

Figure 04. Les modalités d'Apprentissage automatique

Figure 05. Relations entre Intelligence Artificielle, Machine Learning et Deep Learning

Figure 06. Les modalités d'Apprentissage profond

Figure 07. Structure et fonctionnement du neurone artificiel.

Figure 08. L'architecture typique d'un réseau de neurones artificiels.

Figure 09. Architecture et fonctionnement des CNN.

Figure 10. Les 4 densités radiologiques

Figure 11. Flux de sélection des données

Figure 12. Courbe ROC d'entité diagnostique : Epanchement pleurale liquidien -MAIA AI

Figure 13. Courbe ROC d'entité diagnostique : Pneumonie -MAIA AI –

Figure 14. Courbe ROC d'entité diagnostique : Pneumothorax -MAIA AI

Figure 15. Courbe ROC d'entité diagnostique : TLT normal -MAIA AI -

Figure 16. Courbe ROC d'entité diagnostique : Epanchement pleurale -GPT 5.2-

Figure 17. Courbe ROC d'entité diagnostique : Pneumonie -GPT 5.2-

Figure 18. Courbe ROC d'entité diagnostique : Pneumothorax -GPT 5.2-

Figure 19. Courbe ROC d'entité diagnostique : TLT normale -GPT 5.2-

Figure 20. Concordances diagnostiques de -MAIA.IA-

Figure 21. Concordances diagnostiques de -GPT 5.2-

Figure 22. Courbes ROC comparatives (GPT-5.2 vs MAIA AI), Epanchement pleurale

Figure 23. Courbes ROC comparatives (GPT-5.2 vs MAIA AI), Pneumonie.

Figure 24. Courbes ROC comparatives (GPT-5.2 vs MAIA AI), Pneumothorax

Figure 25. Courbes ROC comparatives (GPT-5.2 vs MAIA AI), Radiographie thoracique normale.

Figure 26. Le coefficient Kappa de Cohen par signe radiologique GPT-5.2 vs MAIA AI

Figure 27. Répartition des erreurs diagnostiques par entité pathologique et par modèle d'IA

INTRODUCTION :

L'intelligence artificielle représente une avancée prometteuse dans le domaine de la santé, susceptible d'améliorer la prise en charge des patients et d'alléger la charge de travail des médecins sans compromettre la qualité diagnostique. L'essor des données de santé ouvre des perspectives encourageantes quant à ses applications médicales, dont le dépistage précoce du cancer constitue un exemple emblématique. Par ailleurs, la faillibilité inhérente à l'être humain justifie le développement d'outils d'aide à la décision capables de réduire les erreurs diagnostiques, en exploitant le Big Data — c'est-à-dire de très grands volumes de données médicales — afin de ne laisser échapper aucun indice clinique pertinent (1).

L'analyse d'images par intelligence artificielle a connu une expansion significative depuis les années 1990, portée par l'émergence des algorithmes de l'apprentissage profond (Deep Learning) fondés sur les réseaux de neurones, ainsi que par l'accroissement continu de la puissance de calcul. En médecine, leur application aux différentes modalités d'imagerie est particulièrement vaste, incluant la détection d'anomalies, la reconstruction d'images et l'identification de biomarqueurs radiologiques. Ces derniers contribuent à l'aide au diagnostic, à l'évaluation pronostique et à la prédiction de la réponse thérapeutique, dans un domaine désigné sous le terme de radiomique. Bien que de nombreux travaux demeurent au stade de la recherche, les résultats obtenus témoignent d'un potentiel clinique particulièrement prometteur.

L'intelligence artificielle est désormais au cœur de nombreuses avancées technologiques, et son impact sur les pratiques de soins est indéniable. Le nombre de publications consacrées à l'IA en radiologie a connu une croissance exponentielle au cours des dernières années, passant d'environ 100-150 publications annuelles à plus de 700-800 articles par an. Les algorithmes d'apprentissage automatique permettent d'optimiser différentes étapes du flux de travail radiologique, notamment la planification des examens, le triage des patients, l'aide à la décision clinique, la détection et l'interprétation des anomalies, le post-traitement des images, le contrôle de qualité et la génération automatisée de comptes rendus (2).

Cette technologie exerce une influence croissante sur la radiologie. Elle offre la possibilité d'assister en continu le radiologue ou le clinicien dans l'interprétation diagnostique des images, en réduisant les risques d'erreur, d'inattention et de variabilité inter-observateur, tout en facilitant la prise de décision médicale. Elle permet également d'améliorer la qualité des images, de segmenter des structures anatomiques et d'effectuer des mesures précises.

Cependant, ses applications en imagerie comportent certaines limites qu'il convient de ne pas négliger : les performances des algorithmes dépendent étroitement de la qualité et de la représentativité des données utilisées pour leur entraînement. De plus, leur domaine de validité reste limité à leur champ d'apprentissage et ne peut être extrapolé à d'autres populations ou environnements sans validation préalable. Leur utilisation nécessite une supervision médicale rigoureuse et soulève des enjeux éthiques, réglementaires et organisationnels importants (3).

Dans ce contexte, la radiographie thoracique — ou téléthorax —, examen d'imagerie de première intention largement utilisé en pratique clinique quotidienne, constitue un domaine d'application privilégié de l'intelligence artificielle. Plusieurs études ont montré que l'utilisation d'outils d'IA, en complément de l'interprétation radiologique, permet d'améliorer les performances diagnostiques des praticiens dans la détection de diverses anomalies, telles que les condensations pulmonaires, les épanchements pleuraux, les pneumothorax, les cardiomégalies et les nodules pulmonaires, tout en réduisant le temps d'analyse des clichés (4).

En Algérie, et particulièrement dans les wilayas du sud telles que Laghouat, l'accès à des radiologues spécialisés demeure insuffisant au regard des besoins de la population. Cette réalité engendre une surcharge de travail pour les pneumologues qui assurent, en plus de leurs missions cliniques, l'interprétation des examens radiologiques. Dans ce contexte, l'évaluation d'outils d'intelligence artificielle capables d'assister le praticien dans l'interprétation du téléthorax revêt une importance particulière, susceptible d'améliorer la qualité des soins et de réduire les délais diagnostiques dans les structures à ressources limitées.

C'est pour répondre à ce besoin local et contribuer à combler ce manque dans la littérature algérienne que s'inscrit notre étude, conduite à l'Hôpital Mixte de Laghouat, dont l'objectif est d'évaluer et de comparer les performances diagnostiques de deux modèles d'intelligence artificielle — MAIA AI et X-Ray & Radiologic GPT-5.2 (ci-après dénommé « GPT-5.2 ») — dans l'interprétation du téléthorax en conditions cliniques réelles.

Afin de répondre à ces objectifs, notre travail s'articule en deux grandes parties. La première partie, théorique, pose les fondements nécessaires à la compréhension de notre sujet à travers trois chapitres : le premier aborde les bases de l'intelligence artificielle, de son histoire à ses mécanismes d'apprentissage automatique et profond ; le deuxième est consacré à la

radiographie thoracique standard, son anatomie radiologique et sa s miologie ; le troisi me, au c ur de notre probl matique, traite de l'IA appliqu e   l'interpr tation du t l thorax, en examinant les algorithmes d di s, les performances rapport es dans la litt rature, les limites actuelles et les m thodes de validation des mod les. La seconde partie, pratique, pr sente l' tude que nous avons r alis e, en d taillant la m thodologie adopt e, les r sultats obtenus et leur discussion critique, dans le but d' valuer et de comparer les performances diagnostiques de MAIA AI et de GPT-5.2.

PROBLÉMATIQUE :

Si le téléthorax occupe une place centrale dans la pratique clinique quotidienne, son interprétation reste soumise à une variabilité inter-observateur non négligeable, liée à l'expérience du lecteur, aux conditions de lecture et à la charge de travail croissante des pneumologues et des radiologues. Dans ce contexte, l'émergence des systèmes d'intelligence artificielle appliqués à l'analyse des images médicales suscite un intérêt croissant, en raison de leur capacité potentielle à détecter certaines anomalies radiologiques avec une précision parfois comparable à celle d'un expert expérimenté. Cette situation est d'autant plus préoccupante dans des régions comme Laghouat, où la densité insuffisante en radiologues spécialisés et la surcharge de travail croissante des pneumologues rendent l'évaluation d'outils d'aide à l'interprétation automatisée particulièrement pertinente.

Cependant, malgré la multiplication des études publiées dans ce domaine, plusieurs interrogations persistent : les algorithmes d'intelligence artificielle sont-ils réellement performants dans l'interprétation de la radiographie thoracique ? Leurs performances diagnostiques sont-elles comparables à celles d'un pneumologue senior dans la détection des principales pathologies thoraciques ? Et dans quelle mesure ces systèmes peuvent-ils constituer un outil fiable d'aide à la décision clinique ?

Dans ce cadre, nous avons choisi d'évaluer deux modèles aux profils complémentaires : MAIA AI, système dédié à l'analyse de la radiographie thoracique, et GPT-5.2, modèle généraliste multimodal, afin de comparer leurs performances diagnostiques en conditions cliniques réelles à l'Hôpital Mixte de Laghouat.

C'est dans cette perspective que s'inscrit notre travail, dont l'objectif principal est d'évaluer et de comparer les performances diagnostiques de deux modèles d'IA (MAIA AI et GPT-5.2) dans l'interprétation du téléthorax, en analysant leur capacité à détecter les principales anomalies radiologiques à travers l'étude d'un ensemble de radiographies thoraciques représentatives des pathologies les plus fréquentes : épanchement pleural, pneumonie, pneumothorax — ainsi que des clichés thoraciques normaux utilisés comme groupe contrôle.

PARTIE THEORIQUE

CHAPITRE I : L'INTELLIGENCE ARTIFICIELLE :

1/ DEFINITION :

L'intelligence artificielle (IA) a été définie de diverses manières par ses créateurs pionniers et les références contemporaines. John McCarthy, l'inventeur du terme en 1956, la définissait comme « la science et l'ingénierie de la fabrication de machines intelligentes, en particulier de programmes informatiques intelligents » (5).

Selon Le Larousse (2019), il s'agit d'un « ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine » (6).

Pour Marvin Lee Minsky (2020), l'un des fondateurs du domaine, l'IA correspond à « la construction de programmes informatiques qui s'adonnent à des tâches qui ne sont, pour l'instant, accomplies de façon satisfaisante que 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 »(7).

2/ HISTOIRE DE L'INTELLIGENCE ARTIFICIELLE :

L'évolution historique de l'intelligence artificielle en tant que discipline académique remonte aux années 1950 et s'étend sur plus de soixante-dix ans, allant des fondements philosophiques antiques aux technologies contemporaines des modèles d'IA. Ce progrès a été marqué par l'alternance d'âges d'or et d'hivers, ainsi que par des avancées technologiques qui ont finalement mené à l'étape actuelle de récolte et de maturité (8).

On peut schématiser cette histoire en phases paradigmatiques successives : l'IA symbolique, fondée sur la logique formelle (années 1960-1980), les systèmes experts (années 1980-2010), et l'apprentissage profond (« deep learning ») avec les modèles de langage (années 2010 à aujourd'hui) (9).

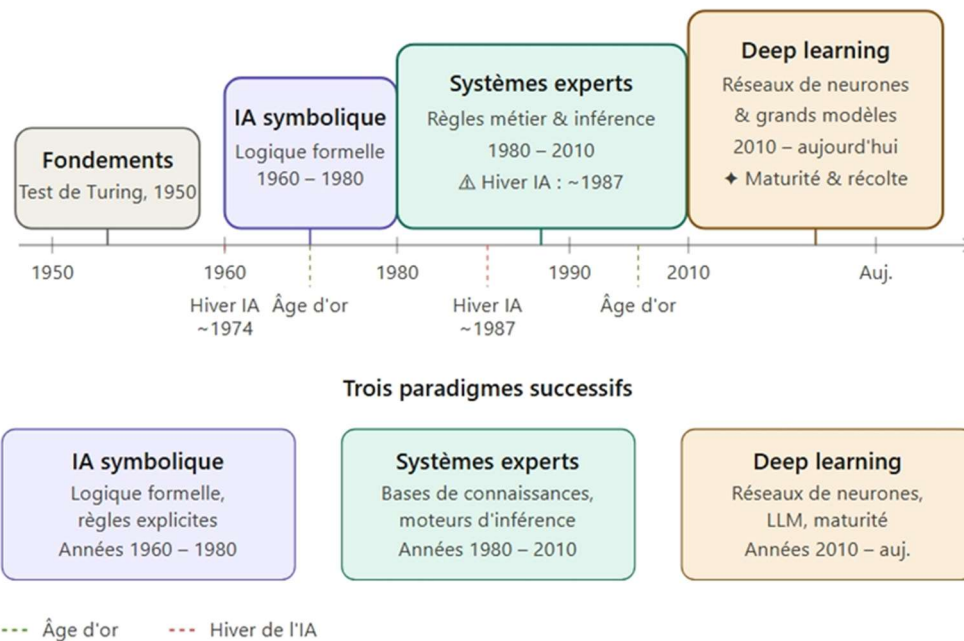


Figure 01. L'évolution historique de l'intelligence artificielle

2.1 Racines antiques et fondements d'IA :

Même avant la naissance de l'informatique, l'intelligence artificielle trouve ses origines dans la mythologie et la philosophie antiques. Cette aspiration s'est manifestée à travers la création des entités autonomes, cherchant à imiter le vivant et ses fonctions cognitives et motrices par la technique, illustrant le concept de “biotechnē” (la vie par l’artisanat). L'exemple le plus frappant est celui des Servantes d'Or d'Héphaïstos décrites par Homère. Celles-ci ne sont pas de simples automates mécaniques, mais possèdent l'intelligence et la raison (10) .

L'intelligence artificielle constitue un champ pluridisciplinaire issu de la convergence de plusieurs sciences fondamentales. Elle puise ses racines dans la philosophie et les mathématiques pour la logique du raisonnement et la probabilité et de calculabilité, dans les neurosciences et la psychologie pour la compréhension des mécanismes des fonctions cognitives et dans l'ingénierie informatique qui a rendu ces théories opérationnelles en fournissant les infrastructures matérielles indispensables à l'exécution des algorithmes d'IA (11).

Révéler les dimensions de l'intelligence artificielle

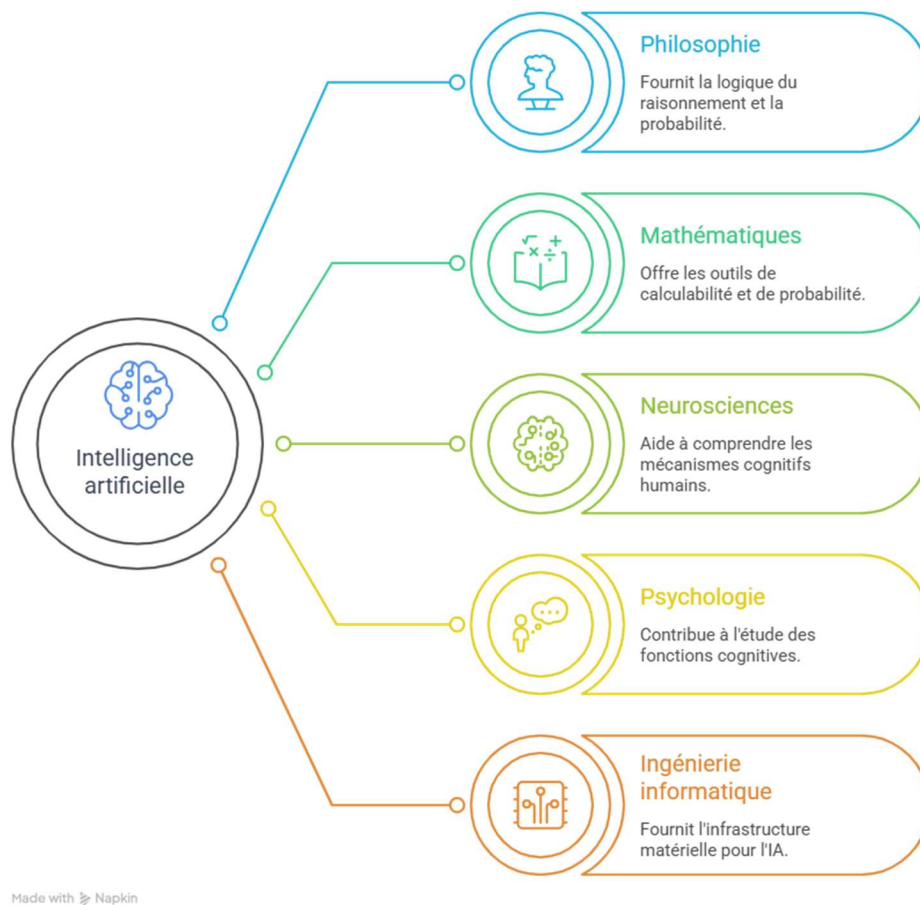


Figure 02. Les sciences fondamentales constitutives de l'intelligence artificielle.

2.2 La genèse d'IA et le modèle de neurones artificiels (1943–1956) :

Le premier travail fondateur de l'IA a été réalisé par les neurophysiologistes Warren McCulloch et le logicien Walter Pitts en 1943. Ils proposent un modèle de neurones artificiels basé sur la logique propositionnelle, la physiologie neuronale et la théorie de la calculabilité de Turing. Leur idée principale était de montrer que le système nerveux fonctionne comme une machine logique. Pour cela, ils ont transformé la loi biologique du « tout ou rien » et le mécanisme de seuil d'excitation en règles logiques binaires. Ils ont ainsi prouvé qu'un réseau des neurones artificiels peut effectuer n'importe quel calcul logique. En 1950, deux avancées majeures ont eu lieu : Alan Turing annonce l'IA avec son test pour évaluer l'intelligence d'une machine et ses idées sur l'apprentissage ; Marvin Minsky construit le premier ordinateur à réseau neuronal, le SNARC (11,12).

2.3 Naissance Formelle d'IA et paradigme Symbolique (années 1960-1980) :

La discipline de l'intelligence artificielle prend officiellement naissance lors de la conférence de Dartmouth en 1956. Organisée par John McCarthy, qui lui donne son nom et entreprend de formaliser les mécanismes de l'apprentissage et de l'intelligence pour les reproduire sur une machine. Cet événement historique a instauré le paradigme de l'IA symbolique — selon lequel l'intelligence peut être reproduite par la manipulation de symboles et de règles logiques — en démontrant, grâce au programme *Logic Theorist* de Newell et Simon, que l'ordinateur pouvait dépasser le simple calcul numérique pour manipuler des concepts logiques. Ces travaux fondateurs ont été soutenus par la création du langage Lisp par McCarthy en 1958, qui est devenu dominant en IA au cours des trente années suivantes, et par le développement de programmes symboliques comme : le *General Problem Solver* (GPS) de Newell et Simon, et le jeu de dames d'Arthur Samuel (11).

Entre 1966 et 1973, l'intelligence artificielle a traversé sa première crise majeure, où elle était capable de résoudre des problèmes simples avec succès, mais n'arrivait pas à résoudre des problèmes un peu plus complexes en raison des limitations de vitesse des ordinateurs, incapables de faire l'inventaire de toutes les possibilités. Cette déception a donné lieu au rapport Lighthill de 1973, qui a critiqué l'IA de manière très sévère, avec des réelles conséquences en termes de financement, ce qui a donné lieu au premier "hiver de l'IA" (11).

2.4 Les systèmes experts (années 1980-2010) :

Face aux limites des premières méthodes de l'IA, qui échouaient face à la complexité du monde réel, les systèmes experts ont émergé entre 1969 et 1986 en passant de programmes cherchant des solutions "à l'aveugle" à des programmes exploitant une connaissance approfondie et spécialisée, démontrant que la puissance d'un système intelligent réside avant tout dans son savoir. Les pionniers du domaine furent DENDRAL en chimie, MYCIN pour le diagnostic médical des infections sanguines qui égalait les performances d'experts humains, et R1, premier système commercial rentable dans la configuration d'ordinateurs. Cette période a également vu l'émergence d'outils de représentation des connaissances comme le langage Prolog, et de grands projets nationaux, comme le projet japonais de la Cinquième Génération (1981) (11).

Cependant, ces systèmes experts, bien que très sophistiqués, n'ont pas réussi à tenir leurs promesses en raison de leur complexité extrême et de leurs performances limitées surtout à

automatiser l'acquisition des connaissances, notamment face aux données numériques comme en reconnaissance de formes ou en synthèse vocale. Leur déclin a été accéléré par des facteurs économiques, comme le prix très élevé de leurs machines LISP spécialisées face aux ordinateurs personnels moins chers, mettant fin à la domination de cette méthode et de l'IA symbolique en général (9).

2.5 Le retour des réseaux de neurones (1986-présent) :

Dans les années 1950, les premiers chercheurs cybernéticiens réalisent des modèles qui s'inspirent du cerveau d'où s'arrive l'IA connexionniste (ou réseaux de neurones), qui permet aux machines d'apprendre par elles-mêmes à partir d'exemples ; mais cette approche est brutalement rejetée en 1969 par les travaux de Minsky et Papert qui montrent ses limites, permettant à l'IA symbolique de dominer le domaine pendant plusieurs décennies (11).

Au milieu des années 1980, la méthode d'apprentissage appelée rétropropagation a été redécouverte et popularisée, notamment grâce à l'ouvrage "*Parallel Distributed Processing*", relançant l'intérêt pour les réseaux de neurones. Ces modèles "connexionnistes" ont deux avantages importants : premièrement, ils peuvent gérer des concepts flous, ce qui est utile car dans la vie réelle les choses ne sont pas toujours parfaitement définies ; deuxièmement, ils apprennent par l'exemple en comparant leurs résultats avec les bonnes réponses et en s'ajustant pour faire moins d'erreurs à l'avenir (9).

2.6 Le raisonnement probabiliste et l'apprentissage automatique (1987-présent) :

L'évolution de l'intelligence artificielle depuis 1987, marquée par un abandon progressif des systèmes experts fragiles et des approches symboliques pour une approche scientifique basée sur les probabilités et l'apprentissage automatique. Les chercheurs utilisent des données réelles et des référentiels précis pour valider leurs résultats comme MNIST et ImageNet, pour la reconnaissance d'image.

Comme le dit David McAllester en 1998, l'apprentissage automatique se rapproche de la théorie de l'information, le raisonnement avec incertitude rejoint les modèles statistiques, et la recherche de solutions s'inspire de l'optimisation classique. Tous les domaines recommencent à collaborer.

Deux avancées majeures en 1988 illustrent ce tournant : les réseaux bayésiens qui offrent un cadre formel pour le raisonnement probabiliste, et les travaux de Rich Sutton liant l'apprentissage par renforcement aux fondements mathématiques. La reconnaissance vocale, dominée par les modèles de Markov cachés (HMM), évoque cette transition (11).

2.7 L'ère de big data (2001-présent) :

Grâce à Internet et à des ordinateurs de plus en plus puissants, on dispose d'énormes quantités de données – un phénomène parfois appelé big data- : des milliards de mots, d'images, de vidéos. Des chercheurs montrent qu'augmenter la quantité de données est plus efficace qu'améliorer l'algorithme. L'exemple le plus évident est La base d'images ImageNet, avec ses millions de photos, déclenche une révolution en vision par ordinateur (11).

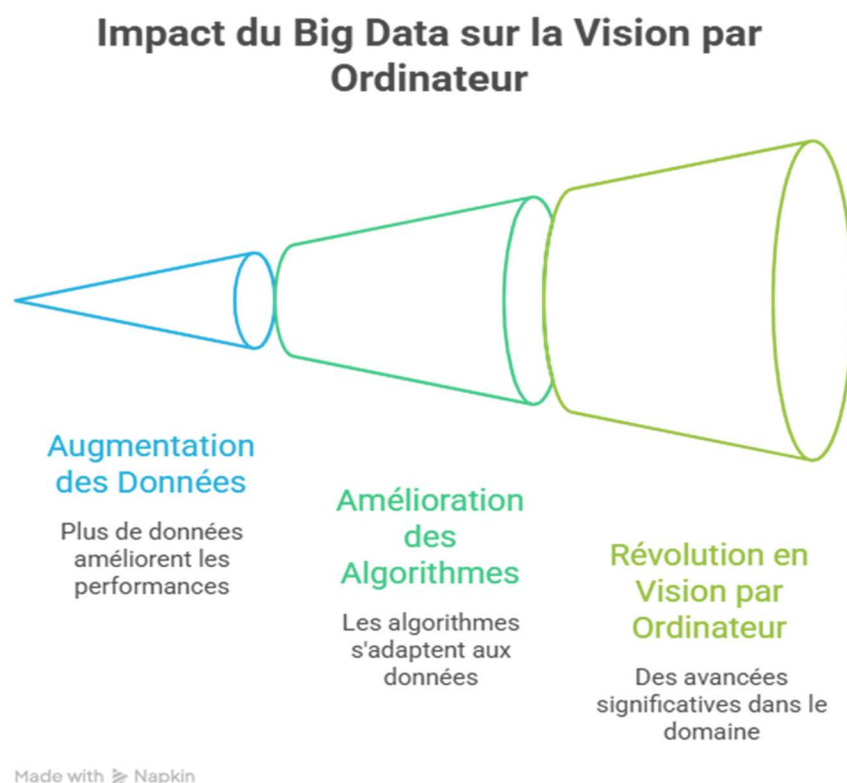


Figure 03. Impact du Big Data sur la vision par ordinateur.

2.8 La Révolution du Deep Learning (2010 - présent) :

L'apprentissage profond, ou *Deep Learning*, repose sur l'utilisation de réseaux de neurones à couches multiples capables d'apprendre des représentations très complexes. Bien que l'idée existe depuis les années 1970, son succès a véritablement commencé en 2012, lorsque les recherches de Geoffrey Hinton ont montré que les machines pouvaient apprendre des caractéristiques par elles-mêmes, dépassant ainsi les modèles créés manuellement par les

experts humains. Aujourd'hui, cette technologie obtient des résultats exceptionnels dans des domaines variés comme la reconnaissance vocale, la médecine ou les jeux (AlphaGo). Cette révolution repose sur la réunion de trois facteurs clés : l'accès à d'immenses quantités de données "Le Big Data", l'invention de nouvelles méthodes de calcul et l'utilisation de matériel informatique ultra-puissant (GPU) capable de traiter des milliards d'opérations (11).

3/ L'APPRENTISSAGE AUTOMATIQUE (MACHINE LEARNING):

3.1 Définition :

En 1959, Arthur Samuel fut le premier à introduire le terme « *Machine Learning* » (apprentissage automatique). Il le définit comme « *le domaine de l'étude et de la conception d'algorithmes qui permettent aux ordinateurs d'apprendre sans être explicitement programmés* » (13).

L'apprentissage automatique peut également être décrit comme « *un ensemble de méthodes permettant aux ordinateurs d'automatiser la construction de modèles et la programmation à partir des données, grâce à la découverte systématique de motifs statistiquement significatifs dans les données disponibles* » (14).

3.2 Modalités d'apprentissage automatique :

Il existe plusieurs scénarios d'apprentissage qui diffèrent par la nature des données d'entraînement disponibles pour l'apprenant, l'outil et la stratégie selon lesquels les données d'entraînement sont reçues. Ci-dessous quelques scénarios courants (15,16).

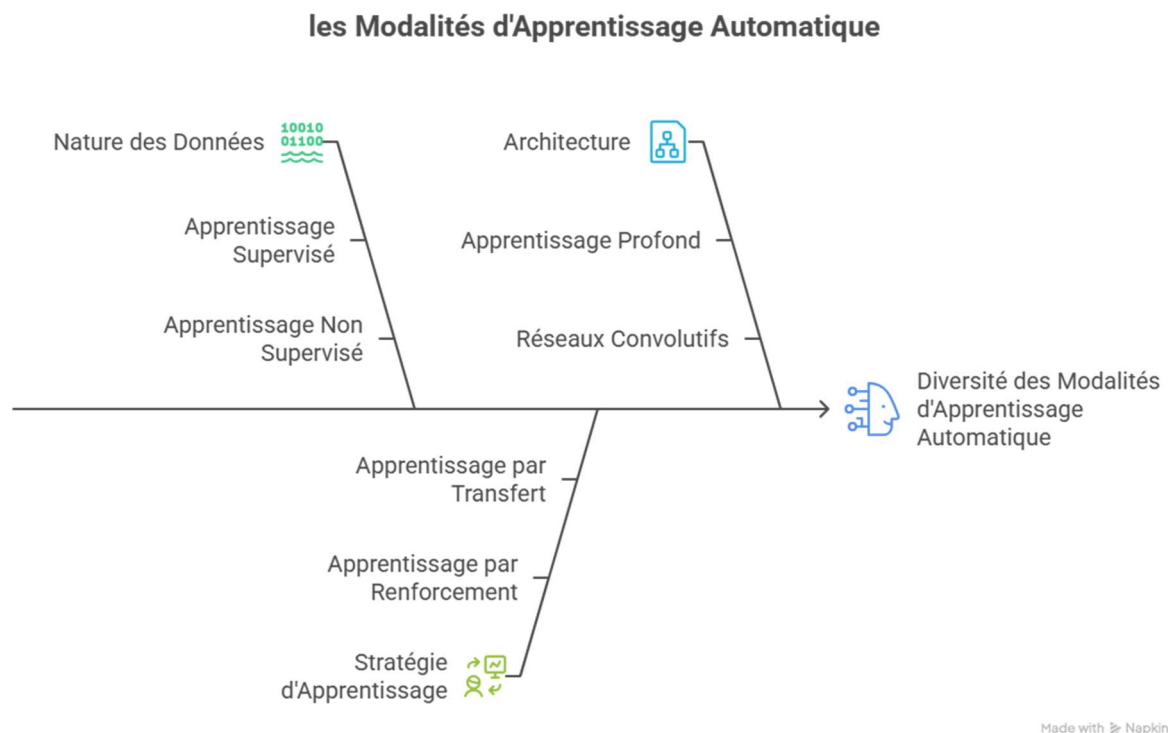


Figure 04. Les modalités d'Apprentissage automatique.

- **Selon la nature des données :**

- L'Apprentissage Supervisé : L'apprenant observe des exemples d'entrée et de sortie, et apprend une fonction qui, à partir d'une entrée, produit une sortie. La sortie peut être une valeur continue (régression) — ce qui consiste à prédire une valeur réelle pour chaque élément — ou discrète (classification) — ce qui consiste à assigner une catégorie à chaque élément. C'est le scénario le plus courant (11,15).

- L'apprentissage non supervisé : L'apprenant n'observe que des entrées, sans sorties associées. L'objectif est de trouver des structures régulières dans les données, comme des regroupements (11).

- L'apprentissage semi supervisé : L'apprenant reçoit un échantillon d'entraînement composé à la fois de données étiquetées (des entrées avec ses sorties) et non étiquetées (des entrées sans sorties), et fait des prédictions pour tous les points non observés. L'apprentissage semi-supervisé est courant dans les situations où les données non étiquetées sont facilement accessibles mais où les étiquettes sont coûteuses à obtenir. Divers types de problèmes rencontrés dans les applications, y compris les tâches de classification, de régression ou de classement (15).

- L'apprentissage auto supervisé : L'apprentissage auto-supervisé est une approche de pré-entraînement sans étiquettes qui apprend des représentations visuelles à partir de données non annotées. L'objectif est d'apprendre un espace de représentation où des images similaires (selon le contexte) sont proches, sans jamais utiliser de labels. Ce pré-entraînement est ensuite suivi d'un réglage fin supervisé sur un petit nombre d'images étiquetées pour la tâche cible (classification dermatologique ou radiographique) (17).

- **Selon la stratégie d'apprentissage :**

-L'apprentissage par renforcement : L'agent apprend à partir d'une série de récompenses ou de punitions. Par exemple, un agent apprend à jouer aux échecs en jouant des parties et en recevant une récompense (+1 pour la victoire, -1 pour la défaite). L'objectif de l'apprenant est de maximiser sa récompense au fil d'une séquence d'actions et d'itérations avec l'environnement. Cependant, aucune récompense à long terme n'est fournie par l'environnement (11,15).

-L'apprentissage par transfert : L'apprentissage par transfert consiste essentiellement à utiliser des réseaux pré-entraînés (typiquement sur des images naturelles comme ImageNet) pour tenter de contourner l'exigence de grands ensembles de données pour l'entraînement de réseaux profonds. Deux stratégies d'apprentissage par transfert ont été identifiées : utiliser un réseau pré-entraîné comme extracteur de caractéristiques, et ajuster finement (fine-tuner) un réseau pré-entraîné sur des données médicales. L'intérêt est de permettre d'obtenir des performances élevées, voire proches de celles d'experts humains, même lorsque le jeu de données médicales étiquetées est de taille limitée (18).

- **Selon architecture (outil) :**

L'apprentissage profond (Deep learning) se présente comme une modalité de l'apprentissage automatique caractérisée par une architecture de couches multiples, et comme une collection d'outils tels que : le perceptron multicouche (MLP), les réseaux convolutifs (CNN) et les réseaux récurrents (RNN), qui partagent le même principe de profondeur hiérarchique. Les CNN, avec l'entraînement supervisé, sont devenus aujourd'hui l'approche privilégiée pour l'interprétation de l'imagerie médicale (16,18).

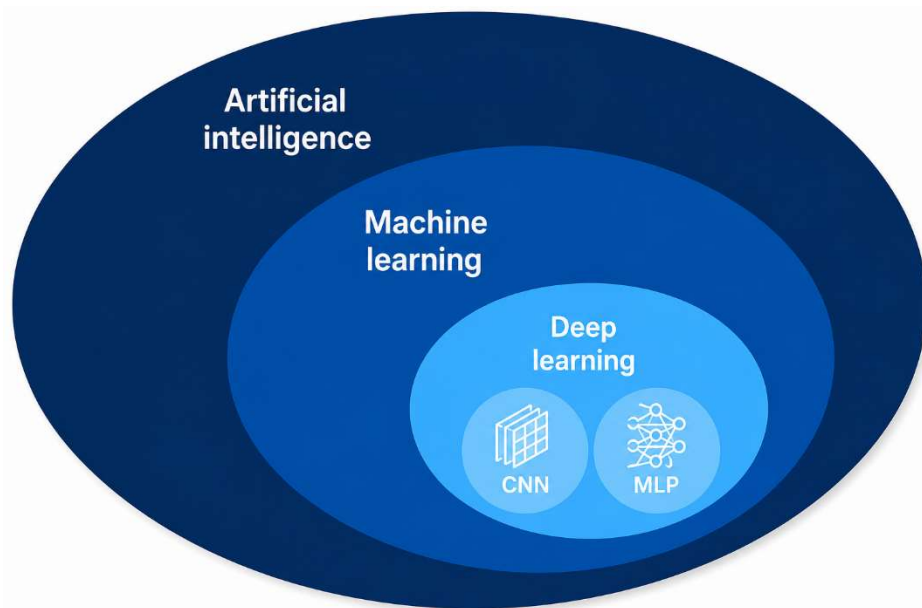


Figure 05. Relations entre Intelligence Artificielle, Machine Learning et Deep Learning

4/ L'APPRENTISSAGE PROFOND « DEEP LEARNING » :

4.1 Définition :

Dans l'ouvrage de référence *Deep Learning*, Ian Goodfellow donne cette définition : « L'apprentissage profond est une forme particulière d'apprentissage automatique qui atteint une grande puissance et une grande flexibilité en apprenant à représenter le monde comme une hiérarchie imbriquée de concepts, chaque concept étant défini par rapport à des concepts plus simples » (16).

Selon François Chollet, l'un des piliers pédagogiques du domaine : « Le « profond » dans « apprentissage profond » ne fait pas référence à une quelconque compréhension plus approfondie obtenue par cette approche ; il représente plutôt cette idée de couches successives de représentations. Le nombre de couches qui contribuent à un modèle des données est appelé la profondeur du modèle » (19).

4.2 Le lien avec les réseaux de neurones artificiels :

L'apprentissage profond est avant tout une architecture composée de nombreuses couches de traitement. La longueur totale de cette chaîne de couches détermine la profondeur du modèle, et c'est de cette terminologie que provient le nom « *deep learning* ». Cette architecture s'appuie presque toujours sur des réseaux de neurones à couches multiples, c'est

pourquoi les premières formes d'apprentissage profond sont apparues sous le nom de « réseaux de neurones artificiels » (ANN), ces modèles étant directement inspirés du fonctionnement du cerveau. Cependant, le terme moderne « *deep learning* » dépasse la perspective neuroscientifique sur la génération actuelle de modèles d'apprentissage automatique qui ne sont pas nécessairement inspirés par les neurones biologiques (16).

4.3 les principales modalités d'apprentissage profond :

L'apprentissage profond n'est pas un modèle unique mais une famille d'architectures qui partagent toutes le même principe de composition de multiples couches de traitement, dont les principales modalités sont : les perceptrons multicouches (MLP), réseaux profonds génériques considérés comme le modèle profond par excellence ; les réseaux convolutifs (CNN), spécialisés dans le traitement de données en grille (images, volumes) ; et les réseaux récurrents (RNN, LSTM, GRU), adaptés aux données séquentielles (parole, texte) (16).

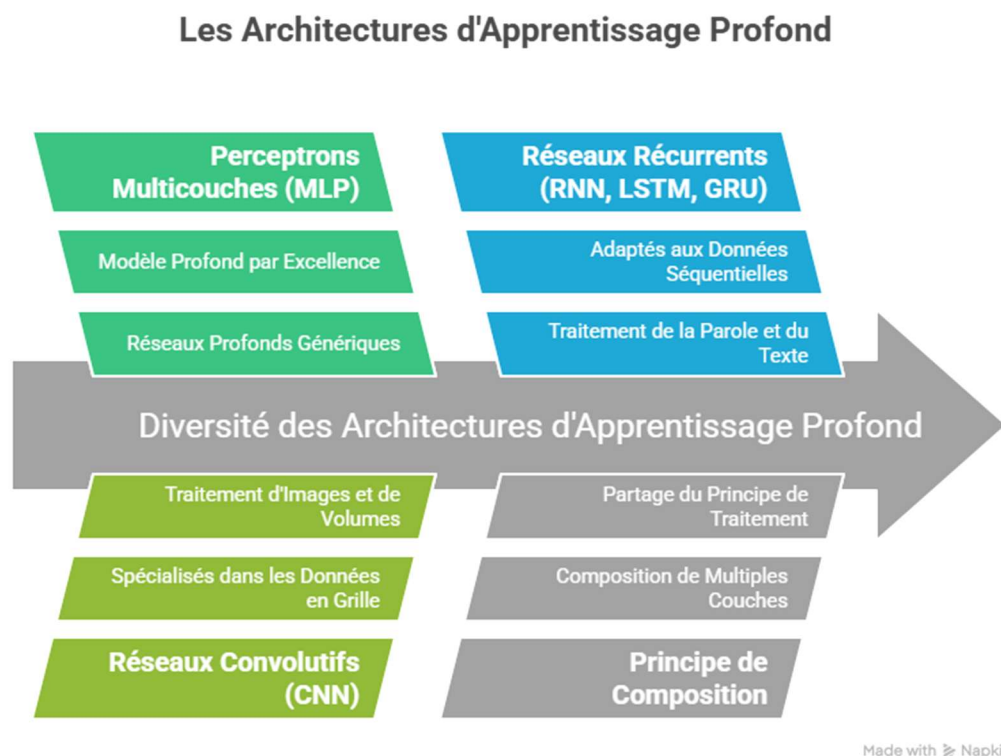


Figure 06. Les modalités d'Apprentissage profond

5/ LES RÉSEAUX DE NEURONES ARTIFICIELS (RNA) :

5.1 Définition :

Le réseau de neurones artificiel (RNA) est défini comme : « un système de traitement de l'information ou du signal composé d'un grand nombre d'éléments de traitement simples, interconnectés par des liaisons directes et qui coopèrent pour effectuer un traitement distribué parallèle afin de résoudre une tâche de calcul souhaitée. » (20).

5.2 Les origines biologiques et l'architecture de neurone artificiel :

Dès 1943, McCulloch et Pitts ont proposé des neurones artificiels mimant les neurones biologiques et capables de mémoriser des fonctions logiques simples, les réseaux de neurones artificiels réalisés à partir de ce type de neurones sont ainsi inspirés du système nerveux. Ils sont conçus pour reproduire certaines caractéristiques des mémoires biologiques par le fait qu'ils sont :

- massivement parallèles ;
- capables d'apprendre ;
- capables de mémoriser l'information dans les connexions entre les neurones ;
- capables de traiter des informations incomplètes (21).

Le neurone artificiel présente une unité qui fonctionne en quatre étapes : **Entrées** — comme les dendrites reçoivent des signaux, le neurone prend des valeurs numériques en entrée ; **Poids** — chaque entrée est multipliée par un poids représentant la force de la connexion ; **Sommation** — les entrées pondérées sont additionnées ; **Fonction d'activation** — la somme passe par une fonction d'activation. (Transmettre l'information). Informatiquement, cette action binaire se traduit par : 0 = pas d'activation / 1 = activation (22).

Le Neurone Artificiel : Fonctionnement en 4 étapes

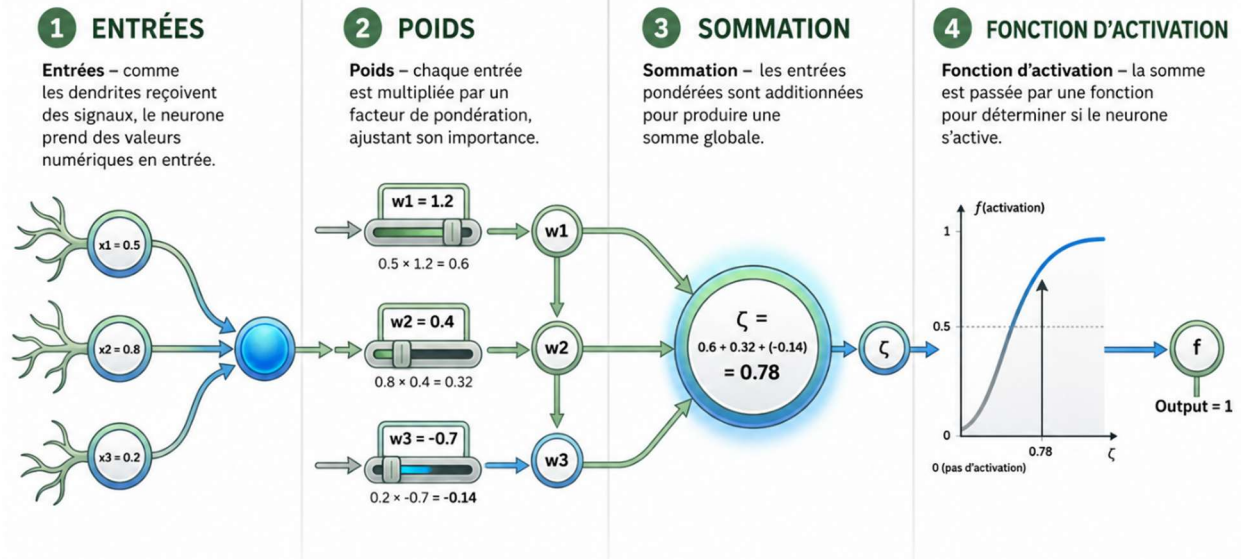


Figure 07. Structure et fonctionnement du neurone artificiel.

5.3 Architecture et fonctionnement des RNA :

L'architecture d'un réseau de neurones artificiels désigne l'arrangement des neurones en couches, les schémas de connexion entre couches, les fonctions d'activation et les méthodes d'apprentissage. La manière dont les neurones d'un réseau de neurones sont structurés est intimement liée à la modalité d'apprentissage utilisé pour entraîner le réseau (22).

Architecture typique d'un Réseau de Neurones Artificiel (RNA)

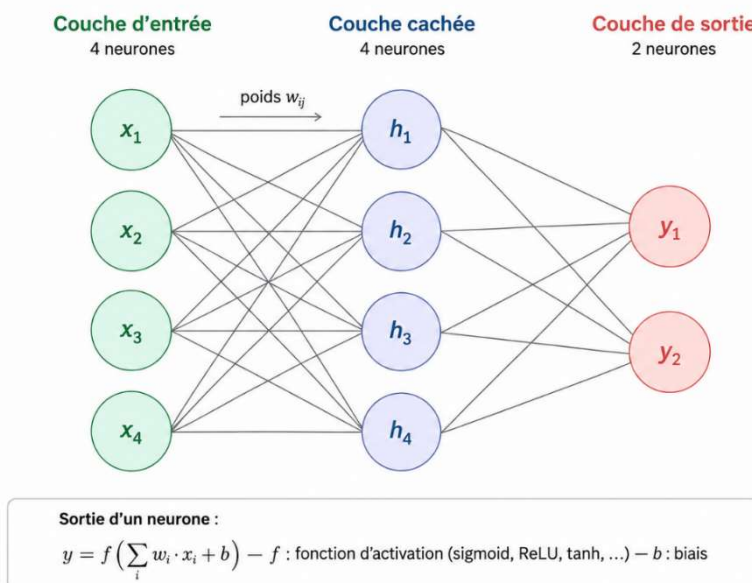


Figure 08. L'architecture typique d'un réseau de neurones artificiels.

Selon Haykin (2009), les architectures des réseaux de neurones artificiels se répartissent en deux grandes familles. D'une part, les réseaux à propagation avant (*feedforward*) se caractérisent par un cheminement strictement direct de l'information, de la couche d'entrée vers la couche de sortie, sans aucune connexion de rétroaction. Cette famille inclut aussi bien les réseaux à une seule couche — comme le Perceptron de Rosenblatt — que les réseaux multicouches. D'autre part, les réseaux récurrents se distinguent par la présence d'au moins une boucle de rétroaction (22).

Les travaux plus récents en apprentissage profond (Goodfellow et al., 2016) conservent cette distinction fondamentale, mais y ajoutent la notion de profondeur (un nombre élevé de couches) et introduisent des architectures spécialisées comme : les réseaux convolutifs (CNN) pour les données spatiales, les réseaux récurrents à mémoire (LSTM, GRU) pour les séquences, ainsi que les autoencodeurs et les modèles génératifs (VAE, GAN) (16).

6/ LES RÉSEAUX DE NEURONES CONVOLUTIFS (CNN) :

6.1 Définition et intérêt :

Dans l'ouvrage de référence *Deep Learning*, Ian Goodfellow donne la définition suivante : “Les réseaux convolutifs, également appelés réseaux de neurones convolutifs ou CNN, sont un type spécialisé de réseau de neurones conçu pour traiter des données possédant une topologie en grille connue (comme les images)” (16).

Les réseaux de neurones convolutifs sont spécialement conçus pour traiter des données structurées en grille, comme des images (grille 2D de pixels) ou des séquences temporelles (grille 1D). Leur nom vient de l'opération mathématique de convolution, qui est l'élément central de leur fonctionnement (11).

6.2 Architecture et fonctionnement des CNN :

L'architecture d'un CNN comprend plusieurs blocs de couches, typiquement des couches de convolution, des couches de pooling et des couches entièrement connectées. Une architecture classique répète une pile composée de plusieurs couches de convolution suivies d'une couche de pooling. Ces deux types de couches sont responsables de l'extraction automatique des caractéristiques à partir des données d'entrée, avant une ou plusieurs couches entièrement connectées qui donnent la sortie finale. Le nom CNN vient de l'opération mathématique de convolution, qui repose sur le principe suivant : au lieu de

connecter chaque neurone à tous les neurones de la couche précédente (connexion dense), les couches convolutives appliquent un ensemble de filtres (ou noyaux) de petite taille qui glissent sur l'image d'entrée et calculent à chaque position un produit de convolution. Chaque filtre détecte un motif local spécifique : un bord, une texture, une forme simple. Une propriété clé est le partage des poids : le même filtre est utilisé sur toutes les positions, ce qui réduit drastiquement le nombre de paramètres et assure une invariance à la translation, c'est-à-dire la reconnaissance d'un motif où qu'il se trouve (11,23).

En empilant plusieurs couches convolutives, le réseau apprend une hiérarchie de caractéristiques : des motifs simples (premières couches) jusqu'aux objets complexes (couches profondes). Les couches de pooling réduisent la dimension spatiale des représentations, rendant le modèle plus robuste aux petites variations de position et diminuant encore le nombre de paramètres. Enfin, après plusieurs blocs de convolution-pooling, une ou plusieurs couches entièrement connectées utilisent les caractéristiques extraites pour produire la décision finale (classification, régression, etc.) (11,23).

Processus d'extraction de caractéristiques dans un CNN

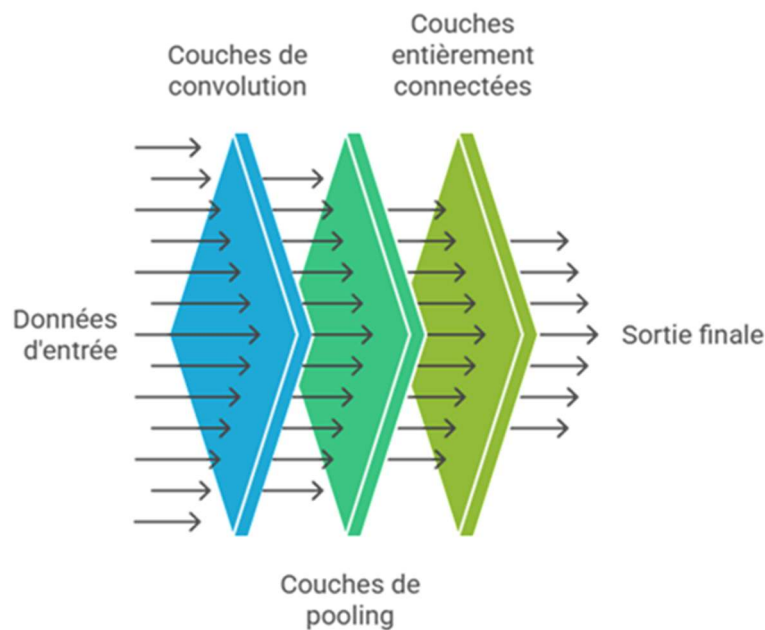


Figure 09. Architecture et fonctionnement des CNN.

7/ MÉDECINE ET INTELLIGENCE ARTIFICIELLE :

L'intelligence artificielle (IA) représente aujourd'hui l'une des mutations les plus profondes que connaisse la médecine depuis l'avènement de l'imagerie médicale. Il s'est progressivement imposé dans l'ensemble du champ biomédical, des soins primaires aux disciplines médicales et chirurgicales hautement spécialisées. L'explosion des données médicales, des imageries de haute résolution et des dossiers de santé électroniques (Big Data) a permis l'émergence des modèles d'IA capables d'égaler, voire de dépasser les performances humaines dans certaines tâches diagnostiques ou pronostiques (24–26).

7.1 Fondements et évolution de l'IA médicale :

L'histoire de l'IA médicale remonte aux années 1970 avec les premiers systèmes experts comme MYCIN, qui utilisaient des règles logiques pour suggérer des traitements antibiotiques. Les limites de ces approches symboliques, qui reposent sur des connaissances encodées manuellement, ont conduit au développement de l'apprentissage automatique, capable d'apprendre directement à partir des données. Aujourd'hui, le machine learning (apprentissage automatique) constitue le socle de la plupart des applications médicales contemporaines ; grâce aux réseaux de neurones convolutifs (CNN) pour l'imagerie médicale et aux réseaux récurrents (RNN) ou transformers pour le traitement des séries temporelles (comme l'ECG et l'EEG) et du langage (extraction d'information à partir de comptes-rendus, génération de texte en produisant un résumé de dossier médical), il a permis des progrès spectaculaires (27–29).

7.2 Applications de l'IA dans les différentes spécialités :

- **Dans les spécialités médicales :**

Les technologies d'IA ont été intégrées dans de nombreuses spécialités médicales, chacune bénéficiant d'une précision diagnostique, d'une modélisation prédictive et d'une planification personnalisée du traitement améliorées (24).

- **Cardiologie :** La détection de la fibrillation auriculaire (FA) représente l'une des applications les plus précoces et les plus validées de l'IA en cardiologie. L'essai REHEARSE-AF a démontré que le dispositif Kardia, un ECG mobile assisté par IA, surpassait la surveillance standard pour le dépistage de la FA en soins primaires (24).

Au-delà de l'arythmologie, Vela et al. (2025) ont documenté dans une revue narrative les apports de l'IA en cardiologie : analyse automatisée d'échocardiographies, prédiction du risque de syndrome coronarien aigu et d'insuffisance cardiaque à partir des dossiers médicaux électroniques, avec des performances supérieures aux scores cliniques traditionnels (30).

- **Neurologie** : En neurologie, les capteurs wearables alimentés par des algorithmes d'IA permettent une quantification objective et continue des symptômes moteurs associés à la maladie de Parkinson, à la chorée de Huntington et à la sclérose en plaques, avec une haute sensibilité pour la détection des anomalies de la marche, des tremblements et des troubles posturaux (24).

La surveillance ambulatoire continue de l'épilepsie constitue une autre application prometteuse : des systèmes IA embarqués sont capables de détecter les crises convulsives avec une fiabilité supérieure aux méthodes conventionnelles (28).

La revue systématique interdisciplinaire de 2025 (PMC12736181) a également mis en évidence les apports de l'IA dans les maladies neurodégénératives, en particulier pour l'intégration multimodale de données d'imagerie, d'électrophysiologie et de biomarqueurs numériques dans l'aide au diagnostic précoce de la maladie d'Alzheimer (31).

- **Oncologie et médecine de précision** : L'oncologie représente le domaine où l'IA exerce son impact profondément. Des modèles d'apprentissage profond sont désormais capables de prédire la réponse individuelle des patients aux traitements anticancéreux à partir de leur profil génomique tumoral, avec une précision supérieure à 80 % dans certaines études. Ces approches s'inscrivent dans le paradigme de la médecine de précision, qui vise à personnaliser les stratégies thérapeutiques en fonction des caractéristiques moléculaires de chaque tumeur. La valeur prédictive positive élevée suggère que l'IA pourrait être utilisée pour identifier des options de traitement de deuxième ligne prometteuses pour les patients en échec des thérapies de première ligne standard (24,28).

Encore, l'intelligence artificielle (IA) change la façon de travailler avec l'histologie numérique. Des scanners produisent des images très détaillées de lames entières. Ensuite, des algorithmes d'apprentissage automatique analysent ces images pour repérer les cellules cancéreuses, mesurer certains biomarqueurs et évaluer le pronostic de la maladie. Parfois,

ces algorithmes sont encore plus précis que des médecins pathologistes très expérimentés (27).

- **Santé mentale** : Les plateformes assistées par IA soutiennent la détection précoce et la stratification du risque psychiatrique et la planification du traitement. Des modèles d'IA analysent les schémas comportementaux et les indices linguistiques pour aider les cliniciens à diagnostiquer la dépression, l'anxiété et la schizophrénie. De plus, les outils de thérapie cognitive et comportementale en ligne augmentés par l'IA ont démontré une efficacité clinique équivalente aux interventions en présentiel pour les troubles émotionnels courants (24,28).

- **Endocrinologie** : La surveillance continue de la glycémie permet aux patients de suivre en temps réel leur taux de glucose interstitiel et ses variations. En 2018, le modèle d'IA [Sugar.IQ](#) a été développé pour aider à prévenir les hypoglycémies en s'appuyant sur des mesures répétées. La surveillance continue, assistée par ce type de modèles, peut aider les patients à mieux contrôler leur glycémie (28).

- **Imagerie et diagnostic radiologique** : La radiologie est le domaine où l'IA a le plus rapidement démontré sa valeur. Des études prospectives ont validé l'usage de l'IA en imagerie thoracique pour générer des comptes rendus radiologiques, accélérant le flux de travail tout en maintenant une qualité comparable à celle des radiologues seniors, voire en dépassant parfois les performances diagnostiques humaines. Malgré ces promesses, une méta-analyse de 2019 a révélé que 99 % de ces études manquaient de méthodologie rigoureuse, limitant leur fiabilité clinique. Ces résultats soulignent le besoin critique d'essais cliniques de haute qualité pour valider les applications de l'IA en radiologie avant une adoption en routine (24).

- **Dans les spécialités chirurgicales :**

- **Chirurgie générale** : L'IA permet une stratification du risque opératoire en temps réel par identification automatique des structures anatomiques, réduisant ainsi le risque de complication peropératoire. La prise de décision préopératoire a également été considérablement améliorée grâce à la capacité de l'IA à perfectionner les modèles de prédiction des risques existants (24).

- **Endoscopie digestive** : En endoscopie digestive basse, les systèmes de détection assistée par ordinateur (CAD e) ont significativement amélioré le taux de détection des adénomes

coliques dans plusieurs essais randomisés contrôlés et facilité la différenciation des lésions bénignes et malignes. De plus, les outils d'IA appliqués à la gastroscopie haute ont atteint une précision de 89 % (sensibilité 90 %, spécificité 88 %) pour la détection de l'œsophage de Barrett néoplasique (24).

- **Chirurgie oncologique** : En chirurgie oncologique, l'IA a permis de prédire les résultats chez les patients atteints de cancers en intégrant de multiples variables liées à la tumeur et au patient. Les modèles présentaient une forte capacité de prédiction de la survie des patients et permettaient d'identifier les variables ayant un impact sur les résultats. Ceci a été réalisé en intégrant des données cliniques, d'imagerie et histopathologiques, permettant une stratégie de traitement personnalisée, représentant un progrès significatif dans la pratique de l'oncologie de précision (24).

- **Chirurgie plastique et reconstructive** : Dans le domaine de la chirurgie plastique, les robots assistés par IA apportent plus de précision, notamment en microchirurgie. L'IA évalue objectivement la symétrie postopératoire, le volume et les résultats esthétiques. De plus, les systèmes de diagnostic pilotés par l'IA aident à identifier précocement des complications telles que l'ischémie des lambeaux, détectant souvent les problèmes plus tôt que les méthodes cliniques traditionnelles (24).

7.3 L'IA en médecine générale et soins primaires :

Bien que la médecine générale soit souvent considérée comme une pratique fondée sur la relation humaine, elle est également touchée par la transformation numérique. Face à la pénurie de médecins et à la surcharge des cabinets libéraux, l'IA pourrait aider à améliorer les soins primaires, à condition qu'elle réponde aux besoins réels des médecins et des patients (28,32).

- **Outils d'autodiagnostic numérique :**

Les outils numériques qui utilisent l'IA pour aider les patients à évaluer leurs symptômes (« symptom checkers ») représentent une nouvelle approche entre prévention et soins primaires. Mentzou et al. (2025) ont publié une revue narrative sur ces outils. Selon eux, ils ont un certain potentiel dans un contexte de manque de soignants, mais ils soulèvent des problèmes : fiabilité diagnostique parfois faible, biais des algorithmes, et questions éthiques liées à l'autonomie et à la sécurité des patients (32).

- **IA générative et grands modèles de langage en pratique clinique :**

L'intelligence artificielle générative et les grands modèles de langage, y compris ChatGPT, Med-PaLM et BioGPT gagnent un intérêt croissant en pratique clinique pour l'automatisation des tâches documentaires : rédaction de comptes rendus de consultation, génération de résumés de sortie, transcription de dictées médicales. Ces systèmes montrent une efficacité prometteuse pour automatiser des tâches expertes, mais présentent des risques de « hallucinations » (génération d'informations médicalement incorrectes) et peuvent amplifier les biais présents dans les données d'entraînement. Un cadre rigoureux de validation, d'explicabilité et de gouvernance est donc indispensable avant leur déploiement clinique à large échelle (24,26,28).

- **Systèmes d'aide à la décision clinique :**

Les systèmes d'aide à la décision clinique (SADC) fondés sur l'intelligence artificielle exploitent de grandes quantités de données de santé (dossiers électroniques, signes vitaux, analyses biologiques) pour aider les cliniciens à poser un diagnostic, à évaluer un pronostic ou à choisir un traitement. Des modèles d'apprentissage automatique ont montré de bonnes performances pour prédire la mortalité hospitalière, la réadmission ou la survenue d'un sepsis (33).

Cependant, l'impact réel de ces outils sur la qualité des décisions médicales reste incertain. Une revue portant sur 37 études a constaté que seule la moitié d'entre elles rapportaient une amélioration avec l'aide de l'IA, sans bénéfice net démontré en situation clinique réelle. Plusieurs obstacles expliquent ce résultat : des problèmes d'intégration dans le flux de travail, un manque de confiance ou une mauvaise utilisation par les soignants, ainsi que des biais issus des données d'entraînement. En définitive, l'IA offre des capacités prédictives prometteuses, mais son utilité concrète dépend étroitement de facteurs humains et organisationnels, et non uniquement de la performance technique des algorithmes (33).

7.4 L'IA en recherche biomédicale :

- **Génomique et médecine personnalisée :**

L'analyse de données génomiques à grande échelle constitue l'une des applications les plus prometteuses du deep learning en recherche clinique. Les algorithmes d'IA facilitent l'identification de variants pathogènes, la stratification des patients sur la base de signatures moléculaires, et l'élaboration de stratégies thérapeutiques personnalisées. Ces avancées sont au cœur des initiatives de médecine de précision (24).

- **Découverte de médicaments :**

Un exemple historique est Alpha Fold de DeepMind, qui a considérablement fait progresser la prédiction de la structure des protéines, une composante essentielle de l'identification des cibles pour le développement de médicaments. Par ailleurs, des systèmes d'intelligence artificielle et d'apprentissage profond ont réussi à reconnaître de nouvelles structures moléculaires dotées de propriétés antibactériennes, contribuant ainsi à élargir l'arsenal des antibiotiques. Aujourd'hui, les modèles d'apprentissage automatique sont utilisés pour concevoir de nouvelles molécules, prédire les interactions entre un médicament et sa cible, et évaluer, par simulation informatique, leur toxicité ainsi que leur devenir dans l'organisme (pharmacocinétique). Ces approches réduisent le temps et les coûts liés aux premières phases de la découverte de médicaments (24).

- **Optimisation des essais cliniques :**

L'intelligence artificielle contribue à améliorer l'efficacité et le succès des essais cliniques. Elle aide au recrutement des patients, par exemple en comparant les caractéristiques physiques ou biologiques (phénotypes) des participants et en les répartissant en sous-groupes (stratification). Elle utilise aussi la modélisation prédictive, qui consiste à établir des prévisions à partir des données disponibles. Par ailleurs, divers appareils et applications mobiles permettent de vérifier que les patients suivent correctement leur traitement (observance). Les grands modèles de langage facilitent l'exploration de données cliniques non structurées, comme des comptes rendus médicaux, ce qui rend les essais plus inclusifs et plus précis. Enfin, la modélisation prédictive rend possible les essais dits adaptatifs : le protocole peut être modifié en cours d'étude en fonction des résultats intermédiaires, ce qui accroît l'efficacité globale de l'essai et réduit les coûts opérationnels (24).

7.5 Limites, défis et considérations éthiques :

- **Acceptabilité professionnelle :**

Briganti et Le Moine (2020) ont mis en évidence que l'IA rencontre une résistance plus marquée de la part des médecins que des patients, notamment en raison d'un déficit de formation aux outils numériques dans les cursus médicaux traditionnels. Ce phénomène est corroboré par l'observation que les cliniciens adoptent historiquement les innovations technologiques de manière prudente et progressive. Plusieurs facultés de médecine ont depuis intégré des modules d'enseignement à la médecine numérique dans leurs programmes, reconnaissant la nécessité de former des praticiens capables d'évaluer de manière critique les outils d'IA et d'en superviser l'utilisation (27,28).

- **Protection des données et cadres réglementaires :**

L'intégration de l'IA dans les systèmes de santé pose des questions majeures de protection des données personnelles de santé. Dans l'Union européenne, le Règlement Général sur la Protection des Données impose des conditions strictes d'anonymisation, de consentement éclairé et de traçabilité pour l'utilisation de données médicales dans l'entraînement de modèles d'IA. L'EU AI Act, applicable depuis 2025, classe les IA médicales dans la catégorie des risques élevés. Elles doivent donc être évaluées et obtenir le marquage CE avant d'être commercialisées. Aux États-Unis, la loi HIPAA encadre la confidentialité des données de santé traitées par des systèmes d'IA (24).

- **Biais algorithmiques et Validité des modèles :**

Le biais algorithmique est l'un des problèmes majeurs de l'IA en santé. Les systèmes d'IA apprennent à partir de données. Si ces données contiennent des inégalités sociales déjà existantes, les algorithmes risquent de les renforcer. Par exemple, un modèle entraîné principalement avec des données provenant d'hommes blancs peut donner des résultats inexacts ou injustes pour les femmes ou pour les patients d'autres origines ethniques.(24)

Un autre problème important est la validité externe des modèles d'IA. Cela signifie qu'un modèle entraîné sur des données issues d'une seule base de données peut moins bien fonctionner quand on l'utilise dans un autre service clinique. Une étude qui a rassemblé plusieurs recherches a montré que, pour l'imagerie médicale, 81 % des algorithmes d'apprentissage profond voyaient leurs résultats baisser lorsqu'ils étaient testés sur de nouvelles données. Parmi ces algorithmes, près d'un quart perdaient beaucoup en performance. Cette fragilité montre qu'il est essentiel de réaliser des essais cliniques

rigoureux et de valider les modèles sur plusieurs centres différents avant de les utiliser à grande échelle (33,34).

- **Transparence et IA explicable (XAI) :**

La majorité des systèmes de l'apprentissage profond (Deep Learning) fonctionnent comme des « boîtes noires » dont les mécanismes décisionnels sont opaques, ce qui heurte frontalement les exigences de transparence propres à la médecine fondée sur les preuves. L'IA explicable (XAI) vise à remédier à cette limitation en développant des méthodes permettant aux cliniciens et aux patients de comprendre la logique sous-jacente aux décisions algorithmiques. Cette transparence est non seulement souhaitable du point de vue éthique, mais constitue une exigence réglementaire croissante dans de nombreuses juridictions (24,33).

CHAPITRE II : RADIOGRAPHIE THORACIQUE

STANDARD :

1. GÉNÉRALITÉS SUR LE TÉLÉTHORAX :

Le téléthorax constitue l'un des examens radiologiques les plus fréquemment prescrits en pratique médicale. En raison de sa facilité de réalisation, de sa large disponibilité et de son faible coût, il occupe une place essentielle dans l'exploration initiale des pathologies thoraciques. Il est ainsi sollicité dans de nombreuses situations cliniques, tant en urgence que dans le cadre du suivi ou du dépistage (35,36).

La fréquence élevée de la demande de téléthorax s'explique par la diversité des informations qu'il permet d'obtenir sur les structures pulmonaires, pleurales, cardiaques et médiastinales. Cet examen est particulièrement utile dans le diagnostic et la surveillance des affections respiratoires, infectieuses, cardiovasculaires et tumorales. Il constitue souvent un examen de première intention, permettant d'orienter la démarche diagnostique et de guider la prise en charge thérapeutique (35,36).

Malgré le développement et l'accessibilité croissante des techniques d'imagerie avancées, notamment la tomodensitométrie, le téléthorax demeure un examen de référence. Son excellent rapport bénéfice–risque, associé à une irradiation relativement faible, justifie sa place centrale dans la stratégie diagnostique (35,36).

La radiographie thoracique repose sur l'utilisation des rayons X, qui traversent le corps humain et sont atténués différemment selon la densité des tissus traversés. Cette atténuation différentielle permet la formation d'une image en niveaux de gris, traduisant les contrastes entre l'air, les tissus mous, les liquides et les structures osseuses.

Ainsi : l'air apparaît en hyperclarté (noir), les tissus mous et liquides en opacité intermédiaire (gris), alors que les structures osseuses apparaissent en hyperopacité (blanc) (35,36).



Figure 10. Les 4 densités radiologiques

Le téléthorax est indiqué dans de nombreuses situations cliniques, notamment en cas de suspicion de pathologie infectieuse pulmonaire, telle que la pneumonie ou la tuberculose. Il est également utilisé dans l'exploration de symptômes respiratoires comme la dyspnée, la toux ou la douleur thoracique. Il joue un rôle important dans le dépistage et le suivi des pathologies tumorales pulmonaires, ainsi que dans l'évaluation des atteintes pleurales, notamment l'épanchement pleural et le pneumothorax. Il permet également d'explorer certaines affections cardiovasculaires, telles que la cardiomégalie ou l'œdème aigu pulmonaire. Aussi le téléthorax est réalisé dans le cadre du bilan préopératoire (35,36).

2. ANATOMIE RADIOLOGIQUE THORACIQUE NORMALE :

La connaissance de l'anatomie radiologique thoracique normale constitue une étape fondamentale dans l'interprétation du téléthorax. Elle permet d'identifier les structures normales, de reconnaître leurs variations physiologiques et de différencier ces aspects des anomalies pathologiques. Le téléthorax offre une représentation globale du thorax, caractérisée par la superposition des structures anatomiques, ce qui nécessite une analyse rigoureuse méthodique et systématisée.

2.1 La paroi thoracique

La paroi thoracique comprend les structures osseuses, les parties molles ainsi que les articulations visibles sur le téléthorax. Les éléments osseux incluent les clavicules, visibles dans leurs portions médiale et latérale, les côtes antérieures et postérieures qui apparaissent sous forme d'arcs réguliers, les scapulas normalement projetées en dehors des champs pulmonaires, ainsi que la colonne vertébrale thoracique dont les corps vertébraux peuvent être visualisés à travers la silhouette cardiaque. Les parties molles correspondent aux muscles et aux tissus sous-cutanés. Chez la femme, les ombres mammaires peuvent modifier la densité des champs pulmonaires inférieurs. L'étude de la paroi thoracique permet ainsi de détecter d'éventuelles anomalies osseuses, des lésions traumatiques ou des masses pariétales (36,38).

2.2 Les champs pulmonaires

Les champs pulmonaires représentent les zones les plus étendues du téléthorax. Ils apparaissent normalement en hyperclarté, traduisant la richesse en air du parenchyme pulmonaire. Le poumon droit est divisé en trois lobes (supérieur, moyen et inférieur), tandis que le poumon gauche en comporte deux (supérieur et inférieur). Toutefois, cette organisation lobaire n'est pas directement visible sur la radiographie standard. L'image pulmonaire normale est dominée par les trames vasculo-bronchiques, plus marquées au niveau des régions péri-hilaires et s'atténuant progressivement vers la périphérie. Un champ pulmonaire normal doit être homogène et dépourvu d'opacités focales ou d'hyperclartés anormales (36,37).

2.3 Les hiles pulmonaires

Les hiles pulmonaires correspondent aux zones de convergence des bronches, des artères et des veines pulmonaires. Sur le téléthorax de face, le hile gauche est généralement situé légèrement plus haut que le hile droit. Ils apparaissent sous forme d'opacités de densité intermédiaire, aux contours relativement flous mais réguliers. Leur analyse repose sur plusieurs éléments, notamment leur position, leur taille, leur symétrie et leur densité. Toute anomalie de ces paramètres peut évoquer une atteinte d'origine vasculaire, ganglionnaire ou tumorale (36,37).

2.4 Le médiastin

Le médiastin est la région centrale du thorax située entre les deux poumons. Il contient le cœur, les gros vaisseaux, la trachée, l'œsophage ainsi que les structures lymphatiques. Sur le téléthorax de face, son analyse repose principalement sur l'étude de la silhouette cardiaque, des contours médiastinaux et de la trachée, qui est normalement médiane. Un médiastin normal présente des contours nets, sans élargissement ni déviation. La trachée apparaît sous forme d'une clarté aérique verticale, légèrement déviée vers la droite au niveau de la crosse aortique (36,37).

2.5 La silhouette cardiaque

La silhouette cardiaque correspond à la projection du cœur sur le téléthorax. Elle est majoritairement visible du côté gauche. L'évaluation de la taille cardiaque repose sur le rapport cardio-thoracique, qui doit être inférieur à 50 % sur un téléthorax réalisé en incidence postéro-anérieure chez l'adulte. Les contours cardiaques doivent être réguliers et bien définis. Toute augmentation de la taille ou modification des contours peut évoquer une cardiopathie sous-jacente (36,37).

2.6 Le diaphragme et les angles costo-phréniques

Le diaphragme apparaît sous la forme de deux coupes, la coupole droite étant généralement plus élevée que la gauche. Leurs contours doivent être réguliers et bien individualisés. Les angles costo-phréniques sont normalement aigus et bien visibles. Leur comblement constitue un signe évocateur d'un épanchement pleural (36,37).

3. SÉMIOLOGIE RADIOLOGIQUE THORACIQUE DE BASE :

La sémiologie radiologique thoracique regroupe l'ensemble des signes élémentaires observables sur la radiographie thoracique, permettant d'orienter le diagnostic vers une pathologie donnée. L'interprétation correcte d'un cliché thoracique repose sur une analyse systématique et méthodique, fondée sur la reconnaissance des anomalies de densité, de forme, de localisation et de distribution des images radiologiques. Cette sémiologie constitue la base de l'interprétation humaine et représente également le fondement des systèmes d'intelligence artificielle appliqués à la radiographie thoracique.

3.1 Les opacités pulmonaires

Une opacité pulmonaire correspond à une augmentation de la densité radiologique du parenchyme pulmonaire, se traduisant par une zone plus blanche que le poumon normal.

a. Opacités alvéolaires

Les opacités alvéolaires résultent du comblement des espaces alvéolaires par du liquide, du pus, du sang ou des cellules. Elles se caractérisent généralement par des contours flous et une tendance à la confluence. Ce type d'opacité est classiquement observé dans les pneumonies, l'œdème aigu pulmonaire ou encore les hémorragies alvéolaires (36,38).

b. Opacités interstitielles

Les opacités interstitielles traduisent une atteinte du tissu de soutien pulmonaire. Elles se manifestent le plus souvent sous forme de lignes réticulées ou de micronodules. Ces anomalies sont rencontrées dans les fibroses pulmonaires, les pneumopathies interstitielles et certaines infections virales (36,38).

3.2 Les clartés pulmonaires anormales :

Une hyperclarté pulmonaire correspond à une diminution de la densité radiologique, traduisant un excès d'air ou une raréfaction du parenchyme pulmonaire.

a. Pneumothorax

Le pneumothorax se caractérise par la présence d'une ligne pleurale visible associée à une absence de trame vasculaire au-delà de cette ligne, ainsi qu'à une hyperclarté périphérique. Il peut être d'origine spontanée, traumatique ou iatrogène (36,38).

b. Emphysème pulmonaire

L'emphysème pulmonaire se manifeste par une hyperclarté diffuse, associée à une diminution de la trame vasculaire périphérique et à un aplatissement des coupes diaphragmatiques (36,38).

3.3 Les anomalies pleurales :

a. Épanchement pleural

L'épanchement pleural apparaît sous forme d'une opacité homogène déclive, comblant les angles costo-phréniques. Dans les formes abondantes, il peut entraîner un refoulement des structures médiastinales (36,38).

b. Épaississement pleural

L'épaississement pleural se traduit par une opacité linéaire ou nodulaire le long de la paroi thoracique. Il est souvent observé dans les séquelles inflammatoires ou dans certaines pathologies tumorales (36,38).

3.4 Les anomalies médiastino-cardiaques :

a. Cardiomégalie

La cardiomégalie est définie par un rapport cardio-thoracique supérieur à 50 % sur un téléthorax de face en incidence postéro-antérieure. Elle peut être le reflet d'une pathologie cardiaque aiguë ou chronique (36,38).

b. Élargissement du médiastin

Un élargissement du médiastin peut être observé dans diverses pathologies d'origine vasculaire, tumorale ou ganglionnaire. Il nécessite généralement la réalisation d'examens complémentaires pour en préciser l'étiologie (36,38).

3.5 Les nodules et masses pulmonaires :

Un nodule pulmonaire correspond à une opacité arrondie de moins de 3 cm de diamètre, tandis qu'une masse pulmonaire dépasse cette taille. Leur analyse repose sur plusieurs critères, notamment la taille, les contours, la densité et la localisation, éléments essentiels pour orienter le diagnostic (36,38).

4. INDICATIONS CLINIQUES VALIDÉES : DU DEPISTAGE AU DIAGNOSTIC :

Le téléthorax est indiqué dans de nombreuses situations cliniques, en particulier lorsqu'une pathologie thoracique est suspectée. Il constitue un examen de première intention dans l'évaluation des symptômes respiratoires tels que la dyspnée, la toux ou la douleur thoracique. Il est fréquemment utilisé dans la recherche de pathologies infectieuses pulmonaires, notamment la pneumonie ou la tuberculose, ainsi que dans l'exploration des atteintes pleuropulmonaires telles que le pneumothorax ou l'épanchement pleural ; aussi en cas de traumatisme thoracique ou poly traumatisme (39).

Le téléthorax permet également d'évaluer certaines affections cardiovasculaires, telles que la cardiomégalie ou l'œdème aigu pulmonaire, en analysant la silhouette cardiaque et les signes de surcharge pulmonaire (39).

En pratique clinique, cet examen est aussi indiqué dans le cadre du bilan préopératoire ou pour le suivi de certaines pathologies chroniques (39).

Il peut également être utilisé en cas d'inhalation d'un corps étrange ; et pour vérifier la position de dispositifs médicaux, tels que les sondes d'intubation ou les cathéters. Ainsi, grâce à sa rapidité d'exécution et à sa large accessibilité, le téléthorax demeure un outil essentiel dans la prise en charge diagnostique initiale des patients (39).

Afin de limiter son utilisation, la HAS a établi en 2009 une liste des principales non-indications de la radiographie thoracique :

Dans les maladies respiratoires non tumorales, la radiographie thoracique n'est pas indiquée dans les infections des voies aériennes hautes, la bronchite aiguë, la bronchiolite non compliquée de l'enfant ou la douleur thoracique non spécifique en dehors de l'urgence (35).

Dans les maladies cardiovasculaires, la radiographie thoracique n'est pas indiquée en cas d'hypertension artérielle (HTA) ou pour le suivi de l'insuffisance cardiaque (35).

Dans les maladies tumorales, la radiographie thoracique n'est pas indiquée pour le dépistage et ne garde que de rares indications pour le suivi, entre deux examens tomodensitométriques (35).

Dans le bilan pré-anesthésie, la radiographie thoracique n'est pas indiquée chez les sujets de moins de 60 ans sans comorbidités cardiopulmonaires (35).

Dans les services de réanimation, la radiographie thoracique n'est pas indiquée de manière systématique chez un patient stable non ventilé (35).

Dans les services d'urgences, la radiographie thoracique n'est pas indiquée systématiquement, et ne doit pas être réalisée lorsque l'on suspecte une perforation œsophagienne, une dissection aortique ou une rupture d'anévrisme (35).

CHAPITRE III – L’INTELLIGENCE ARTIFICIELLE ET L’INTERPRÉTATION DU TÉLÉTHORAX :

Les radiographies thoraciques sont largement utilisées en pratique clinique ; cependant, leur interprétation peut être compromise par des erreurs humaines et le manque de radiologues thoraciques expérimentés. L’apprentissage profond pourrait améliorer la précision de l’interprétation des radiographies thoraciques.

L’IA, notamment basée sur l’apprentissage profond (Deep-learning), peut aider à standardiser l’interprétation ; détecter des anomalies subtiles ; réduire les erreurs humaines ; et améliorer l’efficacité du flux de travail clinique. L’intérêt de l’IA dans les téléthorax repose sur l’existence de bases de données massives, qui permettent d’entraîner des modèles performants capables de repérer des motifs pathologiques complexes.

1. FONCTIONNEMENT DES ALGORITHMES DÉDIÉS AU TÉLÉTHORAX :

Les systèmes IA modernes utilisent principalement des réseaux de neurones convolutifs (CNN) pour analyser les images radiographiques. Ces modèles apprennent à partir de milliers à des millions d’images annotées par des experts. Ils repèrent des motifs visuels associés à des pathologies ; consolidations, pneumothorax, nodules, etc. Un exemple de modèle d’IA performant est VinDr-CXR, qui a été entraîné sur plus de 50 000 clichés de thorax annotés et a montré une performance comparable à celle des radiologues pour la classification de plusieurs maladies thoraciques (AUROC $\approx$ 0,967) (40).

2.PATHOLOGIES DÉTECTÉES PAR L’IA SUR TÉLÉTHORAX :

Les pathologies suivantes ont été évaluées dans la plupart des études et revues récente :

2.1 Pneumonie :

La méta-analyse a révélé une sensibilité globale de 88 % et une spécificité de 90 % pour la détection de la pneumonie par l’IA (41).

2.2 Nodules pulmonaires :

Concernant les nodules pulmonaires, la sensibilité globale de l'IA était d'environ 72 % et sa spécificité d'environ 95 %. Un modèle d'apprentissage profond représentatif pour la détection des nodules est présenté. L'IA avait tendance à ne pas détecter les très petits nodules ou les nodules centraux, mais détectait environ 90 % des nodules plus volumineux. De manière cruciale, l'utilisation de l'IA comme second lecteur a amélioré les performances du radiologue, augmentant la sensibilité d'environ 9 à 10 points de pourcentage (41).

2.3 Pneumothorax :

Globalement, le système d'IA a atteint une précision de 74 % (IC à 95 % : 68-79 %), avec une sensibilité de 66 % (IC à 95 % : 59-73 %) et une spécificité de 93 % (IC à 95 % : 85-97 %). Les radiologues humains ont présenté une précision comparable (77 %, IC à 95 % : 71-82 %, $p = 0,355$), avec une sensibilité supérieure (73 %, IC à 95 % : 66-79 %, $p = 0,040$), mais une spécificité inférieure (85 %, IC à 95 % : 75-91 %, $p = 0,034$). Les performances de l'IA étaient influencées par le positionnement du patient lors de la radiographie thoracique ($p = 0,040$). L'outil proposé pourrait constituer une aide précieuse aux radiologues pour la détection du pneumothorax, en améliorant la spécificité. Un perfectionnement supplémentaire, grâce à un entraînement sur des cas plus complexes, pourrait permettre son utilisation comme outil de dépistage ou comme outil autonome (42).

2.4 Tuberculose pulmonaire et autres pathologies infectieuses :

La précision, la sensibilité et la spécificité combinées, ainsi que les intervalles de confiance à 95 % des modèles d'IA, étaient respectivement de 83,44 % (83,13–83,75 %), 99,88 % (99,86–99,90 %) et 68,19 % (67,95–68,43 %), les modèles TBNet, CAD et ResNet50 présentant la meilleure précision. La validation externe a démontré une précision supérieure [96,51 % (96,35–96,66 %) contre 88,58 % (88,24–88,92 %)] à celle de la validation interne. Les études rétrospectives ont montré une précision, une spécificité et une sensibilité supérieures à celles des études prospectives, tandis que les études menées auprès de populations en zones d'endémie tuberculeuse ont montré une spécificité et une précision supérieures à celles observées auprès de populations non endémiques, avec une sensibilité similaire (43).

Les modèles d'IA se sont révélés très prometteurs pour le dépistage précis de la tuberculose et pourraient transformer sa détection dans les contextes à faibles ressources. Il est toutefois crucial d'intégrer les résultats générés par l'IA au tableau clinique avant d'initier un traitement, afin de garantir une approche holistique et éclairée des soins (43).

2.5 Épanchement pleurale :

Dans une étude multicentrique publiée dans JAMA Network Open, l'utilisation d'un algorithme d'intelligence artificielle pour l'interprétation de radiographies thoraciques a montré une sensibilité élevée pour la détection d'épanchements pleuraux ($\approx 87\%$) en analyse stand-alone. De plus, lorsque les radiologues ont bénéficié de l'aide de l'IA, la sensibilité de la détection d'épanchement pleural est légèrement passée de 0,889 à 0,895, démontrant que l'IA peut améliorer la performance diagnostique humaine sans compromettre la spécificité (44).

3. PERFORMANCES DE L'IA DANS LES ÉTUDES PUBLIÉES :

L'utilisation de l'IA dans l'interprétation des téléthorax a montré plusieurs résultats intéressants :

3.1 Amélioration des performances des lecteurs humains :

Une étude publiée dans JAMA Network Open a montré que l'utilisation d'un outil d'IA augmente la sensibilité des médecins pour détecter des anomalies thoraciques comme pneumothorax, pneumonie ou nodules, sans diminuer la spécificité (40).

Le système de double lecture assistée par IA a permis d'identifier des anomalies non détectées sur les radiographies thoraciques après validation des comptes rendus. Cette approche a nécessité l'intervention d'un radiologue externe pour examiner les discordances détectées par l'IA. Le nombre d'anomalies cliniquement significatives non détectées par les radiologues est resté très faible (40).

Pertinence clinique : Il a été démontré que le flux de travail de double lecture assisté par IA permet de détecter les erreurs de diagnostic et pourrait être utilisé comme outil d'assurance qualité. Bien que les anomalies cliniquement pertinentes non détectées soient rares, leur impact potentiel est à prendre en compte tenu de l'utilisation fréquente de la radiographie thoracique (40).

3.2 Réduction du temps d'interprétation :

L'IA peut également réduire le temps nécessaire pour lire et interpréter les radiographies, améliorant ainsi l'efficacité du service radiologique. Le temps de compte rendu avec l'IA était inférieur de 10 % à celui sans IA (40,8 secondes contre 36,9 secondes ; différence : 3,9 secondes ; IC à 95 % : 2,9-5,2 secondes ; $p < 0,001$ (44).

3.3 Performance diagnostique élevée :

Le système d'IA, approuvé par la FDA, a démontré d'excellentes performances en autonomie, détectant les anomalies thoraciques sur les radiographies avec une précision comparable à celle des radiologues. Ces performances se sont généralisées à un autre jeu de données de radiographies thoraciques accessible au public. Lors des essais cliniques, nous avons montré que la précision globale des médecins s'améliorait grâce à l'assistance du système d'IA, et que les médecins non radiologues atteignaient une précision équivalente à celle des radiologues dans l'interprétation des radiographies thoraciques lorsqu'ils étaient assistés par ce système. En résumé, nos résultats indiquent que le système d'IA soutient les différentes spécialités médicales dans l'interprétation précise des radiographies thoraciques (45).

4. LIMITES ET DÉFIS DE L'IA APPLIQUEE AU TÉLÉTHORAX :

Malgré ses performances prometteuses, l'IA présente plusieurs limites importantes.

4.1 Faux positifs et faux négatifs fréquents :

Les systèmes IA tendent parfois à générer davantage de faux positifs que les radiologues, ce qui peut augmenter la charge de travail plutôt que la réduire. Les auteurs ont montré que les taux de faux positifs étaient plus élevés pour les outils d'IA que pour les comptes rendus radiologiques, tandis que les taux de faux négatifs étaient similaires. Une variation du compromis entre la sensibilité et la spécificité de chaque outil d'IA. De nombreuses recherches fondamentales et cliniques sur l'IA seront menées à partir de données de radiographies thoraciques ont noté que les performances des outils d'IA se dégradent en cas de lésions cibles de petite taille et/ou de lésions multiples et que de nombreuses erreurs commises par les outils d'IA peuvent être difficiles, voire impossibles à détecter pour les lecteurs humains sans accès à un traitement d'image supplémentaire ou à l'historique du

patient. La qualité des résultats dépend fortement des données d'apprentissage. Si ces données sont biaisées (par âge, sexe, population), l'IA risque de reproduire ces biais cliniques (46).

4.2 Dépendance aux données d'entraînement :

La qualité des résultats dépend fortement des données d'apprentissage. Si ces données sont biaisées (par âge, sexe, population), l'IA risque de reproduire des biais cliniques. Les applications de l'IA en imagerie comportent de nombreuses limites et risques qu'il est nécessaire de prévenir : les performances d'un algorithme dépendent de la qualité des données qui ont servi à l'entraîner. Les domaines de compétence et les champs d'application d'une application d'IA donnée sont dépendants de son « apprentissage » et sont très étroits (3).

4.3 Pas de raisonnement clinique :

L'IA ne remplace pas l'interprétation clinique complète, car elle ne prend pas en compte les symptômes, l'historique médical, ni d'autres éléments cliniques essentiels.

5. PLACE DE L'IA DANS LE CONTEXTE CLINIQUE : RÔLE D'AIDE À LA DÉCISION ET NON DE REMPLACEMENT :

L'intelligence artificielle appliquée à l'interprétation du téléthorax n'est pas destinée à remplacer le radiologue, mais plutôt à améliorer la prise de décision clinique. Dans une étude multicentrique récente publiée dans JAMA Network Open, l'utilisation d'un système d'intelligence artificielle basé sur l'apprentissage profond comme support d'interprétation de radiographies thoraciques a été associée à une amélioration significative de la sensibilité diagnostique des lecteurs humains, sans diminuer la spécificité. Les radiologues, fellows et résidents participant à l'étude ont obtenu de meilleurs taux de détection pour des anomalies telles que nodules, pneumonie, épanchements pleuraux et pneumothorax lorsqu'ils utilisaient l'outil IA, tout en réduisant le temps de lecture des examens. Cela illustre que l'IA fonctionne comme un outil d'aide à la décision clinique, renforçant l'efficacité diagnostique sans remplacer l'expertise humaine (44).

6. MÉTHODES D'ÉVALUATION ET VALIDATION DES MODELES D'INTELLIGENCE ARTIFICIELLE EN RADIOGRAPHIE

THORACIQUE :

Le développement d'outils cliniques basés sur l'IA accroît le besoin d'un cadre d'évaluation rigoureux, fiable et pertinent afin de garantir leur mise en œuvre sûre et efficace. Actuellement, les performances des modèles d'IA sont évaluées à l'aide de métriques, bien que ces mesures soient limitées par la tâche, les propriétés des jeux de données et les hypothèses méthodologiques. Pour évaluer une IA, il est nécessaire de la comparer à une « référence », qui peut être un ensemble d'experts indépendants ou un expert du domaine, tel qu'un radiologue, et de l'évaluer à l'aide de nombreux exemples sur plusieurs cas. Si ses performances sont équivalentes à celles de la référence, elle peut être approuvée (1).

6.1 Métriques de classification

L'évaluation robuste des modèles de classification est primordiale en IA appliquée à l'imagerie médicale, car le choix des métriques peut considérablement influencer l'interprétation des performances du modèle et les décisions qui en découlent. Les métriques de classification évaluent la capacité d'un modèle à attribuer des entrées aux catégories appropriées et sont généralement divisées en celles dérivées d'un seuil de décision fixe (souvent basées sur la matrice de confusion) et celles qui résument les performances sur plusieurs seuils (47).

6.2 Matrice de confusion

La matrice de confusion constitue la base de nombreuses métriques d'évaluation pour la classification. Elle résume les résultats des prédictions en quatre catégories :

- **Vrais positifs (True Positives, TP)** : cas positifs correctement identifiés.
- **Faux positifs (False Positives, FP)** : cas négatifs incorrectement identifiés comme positifs.
- **Vrais négatifs (True Negatives, TN)** : cas négatifs correctement identifiés.
- **Faux négatifs (False Negatives, FN)** : cas positifs incorrectement identifiés comme négatifs.

Ces quantités permettent de dériver plusieurs métriques de performance (47).

6.3 Exactitude (Accuracy) :

L'exactitude représente la proportion de prédictions correctes parmi l'ensemble des observations :

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

Bien que facile à interpréter, l'exactitude peut être trompeuse dans des ensembles de données déséquilibrés, où une classe domine largement l'autre (47).

6.4 Sensibilité et spécificité :

La sensibilité (également appelée rappel ou taux de vrais positifs) mesure la proportion de cas positifs correctement détectés :

$$\text{Sensibilité} = \text{TP} / (\text{TP} + \text{FN})$$

La spécificité mesure la proportion de cas négatifs correctement identifiés :

$$\text{Spécificité} = \text{TN} / (\text{TN} + \text{FP})$$

Ces métriques sont particulièrement importantes en médecine, où les conséquences des faux négatifs et des faux positifs diffèrent souvent (47).

6.5 Précision (Précision) :

La précision mesure la proportion de prédictions positives qui sont réellement positives :

$$\text{Précision} = \text{TP} / (\text{TP} + \text{FP})$$

Elle reflète la fiabilité des prédictions positives du modèle (47).

6.6 Score F1 :

Le score F1 correspond à la moyenne harmonique entre la précision et la sensibilité :

$$\text{F1} = 2 \times (\text{Précision} \times \text{Sensibilité}) / (\text{Précision} + \text{Sensibilité})$$

Il est particulièrement utile lorsque les classes sont déséquilibrées (47).

6.7 Courbe ROC et aire sous la courbe (AUC) :

La courbe ROC (Receiver Operating Characteristic) représente la relation entre la sensibilité et le taux de faux positifs à différents seuils de décision. L'aire sous la courbe ROC (AUC) résume la capacité globale du modèle à distinguer les classes indépendamment du seuil choisi. Cependant, l'AUC peut masquer des performances médiocres dans des régions cliniquement pertinentes du seuil (47).

6.8 Pièges et mauvaises applications courantes :

L'utilisation exclusive d'une seule métrique peut conduire à une interprétation incorrecte des performances du modèle. Par exemple, une exactitude élevée peut être obtenue simplement en prédisant toujours la classe majoritaire dans des ensembles de données fortement déséquilibrés. De plus, comparer les valeurs d'AUC entre études utilisant différentes populations ou prévalences peut être trompeur (47).

6.9 Recommandations :

Il est recommandé de rapporter plusieurs métriques complémentaires, incluant la sensibilité, la spécificité, la précision et le score F1, ainsi que d'évaluer les performances à des seuils cliniquement pertinents. Les intervalles de confiance et les méthodes statistiques appropriées devraient également être fournis afin de refléter l'incertitude des estimations de performance (47).

	Métrique	Définition et caractéristiques	Formule
Basées sur la matrice de confusion (seuil de décision fixe)	Exactitude (Accuracy)	Proportion des prédictions correctes	$(TP + TN) / (TP + TN + FP + FN)$
	Sensibilité (Recall / Rappel)	Taux de vrais positifs	$TP / (TP + FN)$
	Spécificité	Taux de vrais négatifs	$TN / (TN + FP)$
	Exactitude équilibrée (Balanced Accuracy)	Moyenne de la sensibilité et de la spécificité ; utile pour les jeux de données déséquilibrés	$(\text{Sensibilité} + \text{Spécificité}) / 2 = [(TP / (TP + FN)) + (TN / (TN + FP))] / 2$
	Valeur prédictive positive (VPP ; Précision)	Probabilité de présence de la maladie étant donné un test positif	$TP / (TP + FP)$
	Valeur prédictive négative (VPN)	Probabilité d'absence de maladie étant donné un test négatif	$TN / (TN + FN)$
	Score F1	Moyenne harmonique de la précision et du rappel ; équilibre les faux positifs et les faux négatifs	$2 \times TP / (2 \times TP + FP + FN)$
	Coefficient de corrélation de Matthews (MCC)	Mesure équilibrée de la classification binaire tenant compte de tous les éléments de la matrice de confusion (TP, TN, FP, FN)	$(TP \times TN - FP \times FN) / \sqrt{[(TP + FP)(TP + FN)(TN + FP)(TN + FN)]}$
Multi-seuils	AUROC	Aire sous la courbe ROC (Receiver Operating Characteristic) ; indépendante du seuil ; mesure la capacité globale de discrimination	Voir la formule dans la Réf. [17]
	AUPRC	Aire sous la courbe précision-rappel ; indépendante du seuil ; met l'accent sur les performances de la classe positive	Voir la formule dans la Réf. [17]

Tableau 1 : Vue d'ensemble des métriques de classification courantes

PARTIE PRATIQUE

I.MATÉRIEL ET MÉTHODES :

1.Type d'étude :

Il s'agit d'une étude diagnostique rétrospective et prospective, analytique et comparative, intitulée « L'évaluation des performances diagnostiques de deux modèles d'intelligence artificielle (MAIA AI et GPT-5.2) dans l'interprétation de la radiographie thoracique », réalisée sur 110 clichés thoraciques. La réalisation et le rapport de cette étude ont été effectués conformément aux recommandations internationales STARD-AI.

2. Lieu de l'étude : Service de pneumologie et de radiologie de l'Hôpital Mixte de Laghouat.

3. Durée de l'étude : Pendant la période s'étendant du mois d'octobre 2025 au mois d'avril 2026.

4. Les objectifs de l'étude :

4.1 L'objectif principal :

Évaluer et comparer les performances diagnostiques des deux modèles MAIA AI et GPT-5.2, en calculant la sensibilité, la spécificité, la VPP et la VPN dans la détection de la pneumonie, du pneumothorax, de l'épanchement pleural et de la radiographie normale, par rapport au gold standard (Pneumologue senior)

4.2 Les objectifs secondaires :

- Tracer et analyser les courbes ROC.
- Évaluer la concordance diagnostique et sémiologique des deux modèles avec le gold standard.
- Comparaison inter-modèles : Comparer statistiquement les performances diagnostiques globales entre GPT-5.2 et MAIA AI.

5. Population d'étude :

La population cible était constituée de patients adultes âgés de 18 ans ou plus, ayant bénéficié d'une radiographie thoracique standard au service de radiologie de l'Hôpital Mixte de Laghouat durant la période d'étude.

➤ **Critères d'inclusion :**

- Radiographies thoraciques de face de bonne qualité.
- Patients des deux sexes, âgés de plus de 18 ans.
- Indication clinique documentée.
- Radiographies comportant un diagnostic confirmé de pneumonie, de pneumothorax, d'épanchement pleural ou un aspect radiologique normal.

➤ **Critères de non inclusion :**

- Radiographie thoracique de profil ou de mauvaise qualité.
- Téléthorax d'un patient de moins de 18 ans.
- Radiographies présentant des pathologies non ciblées par l'étude.
- Patients porteurs de séquelles ou d'antécédents de chirurgie thoracique modifiant significativement l'anatomie radiologique (pneumonectomie, lobectomie...).

6. Collecte des données :

La collecte des données a été réalisée de manière rétrospective et prospective à partir des radiographies thoraciques standard de face réalisées durant la période de l'étude. Les clichés inclus ont été sélectionnés selon les critères d'inclusion et de non inclusion préalablement définis. Certaines radiographies présentaient plusieurs anomalies concomitantes ; par conséquent, les catégories diagnostiques n'étaient pas mutuellement exclusives. Les images ont été exportées au format JPEG haute résolution.

7. Les variables étudiées :

7.1 Variables démographiques :

- Âge.
- Sexe.

7.2 Variables diagnostiques principales :

- Diagnostic retenu par le gold standard ;
- Diagnostic proposé par X-Ray & Radiologic GPT-5.2 ;
- Diagnostic proposé par MAIA AI.

Les diagnostics étudiés étaient :

- Pneumonie ;
- Pneumothorax ;
- Épanchement pleural ;
- Cliché thoracique normal.

7.3 Variables de performance diagnostique :

- Vrais positifs (VP).
- Vrais négatifs (VN).
- Faux positifs (FP).
- Faux négatifs (FN).
- Sensibilité.
- Spécificité.
- Valeur prédictive positive.
- Valeur prédictive négative.
- Exactitude diagnostique (Accuracy).
- Rapport de vraisemblance positif (RV+).
- Rapport de vraisemblance négatif (RV-)
- AUC (l'aire sous la courbe ROC).

7.4 Variables sémiologiques :

➤ Épanchement pleural

- Côté atteint. (Droit / gauche / bilatéral)
- Présence ou non d'opacité pleurale.

- Abondance de l'épanchement.
- Présence ou non d'un refoulement médiastinal.

➤ **Pneumonie**

- Côté atteint.
- Présence d'opacité pulmonaire.

➤ **Pneumothorax**

- Côté atteint.
- Hyperclarté avasculaire.
- Abondance du pneumothorax.
- Présence ou non d'un hydropneumothorax.
- Présence d'un liseré pleural.

➤ **Cliché thoracique normal**

- Interprétation globale du cliché.
- Concordance du diagnostic de normalité.

7.5 Variables de concordance :

- Concordance entre les signes radiologiques décrits par l'IA et le gold standard.

8. Constitution du standard de référence :

Le standard de référence (gold standard) a été constitué par une seule lecture experte indépendante. Un pneumologue senior a interprété chaque radiographie de façon indépendante et en aveugle des résultats des modèles d'IA (STARD-AI, item 17a), avec accès aux seules informations cliniques minimales (âge et sexe). Les étiquettes diagnostiques de référence ont été attribuées et consignées sur une fiche standardisée, puis archivées.

9. Standardisation du prompt :

Conformément aux exigences de transparence et de reproductibilité définies par les lignes directrices STARD-AI (items 15a et 15b) pour l'évaluation des grands modèles de langage, un prompt unique et standardisé a été employé pour toutes les soumissions au modèle GPT-5.2.

Ce prompt comportait les éléments suivants : L'instruction est de diagnostiquer et décrire les signes radiologiques : identifier la ou les anomalies radiologiques avec un score de probabilité et une interprétation radiologique détaillée sous forme d'un rapport radiologique.

En revanche, MAIA AI ne requiert aucune invite textuelle : l'image soumise est analysée automatiquement sans instruction complémentaire de la part de l'utilisateur.

10. Description des modèles d'IA évalués :

Paramètre	Modèle 1	Modèle 2
Nom	X-Ray & Radiologic GPT-5.2	MAIA AI
Éditeur	OpenAI (modèle multimodal GPT-5.2)	maiahub.my
Type	Grand modèle de langage multimodal (LMM)	Modèle d'IA spécialisé en imagerie médicale
Base technologique	Transformer multimodal (Deep Learning)	Vision médicale + LLM médical (Deep Learning)

Paramètre	Modèle 1	Modèle 2
Entraînement	Données générales + vision	Données médicales/radiologiques
Modalité d'entrée	Image JPEG + invite textuelle (prompt)	Image JPEG / DICOM
Mode d'accès	Interface web / API	Interface web (maiahub.my)
Usage déclaré	Aide à l'interprétation radiologique	Aide au diagnostic en imagerie médicale

Tableau 2. Caractéristiques des deux modèles d'IA évalués.

11. Plan d'analyse statistique des données :

✓ Statistiques descriptives :

Les variables qualitatives ont été décrites par des fréquences et pourcentages. Les variables quantitatives ont été décrites par la moyenne $\pm$ écart-type ou la médiane selon la distribution.

✓ Mesures de performance diagnostique :

Pour chaque modèle et chaque entité, ont été calculés (STARD-AI, item 18) : sensibilité (Se), spécificité (Sp), valeur prédictive positive (VPP), valeur prédictive négative (VPN), rapport de vraisemblance positif (RV+), rapport de vraisemblance négatif (RV-), exactitude globale et aire sous la courbe ROC (AUC). Les intervalles de confiance à 95 % ont été calculés par la méthode de Wilson.

✓ Comparaison inter-modèles :

Les sensibilités, les spécificités et les exactitudes ont été comparées par test de McNemar. Le seuil de significativité retenu était $p < 0,05$.

✓ Concordance :

Toutes les analyses de concordance réalisées dans cette étude - incluant la concordance diagnostique, la concordance sémiologique (findings radiologiques) - ont été évaluées à l'aide du coefficient Kappa de Cohen avec intervalle de confiance à 95 %.

✓ **Logiciels :**

Les analyses ont été réalisées avec le logiciel SPSS version 27.0.

12. Considérations éthiques :

L'étude a été soumise à l'approbation de la direction de l'Hôpital Mixte de Laghouat et du comité d'éthique de la Faculté de Médecine de Laghouat. Toute donnée identifiante a été remplacée par un code numérique anonyme avant traitement. L'étude n'a impliqué aucun acte supplémentaire pour les patients. On déclare l'absence de tout conflit d'intérêt financier avec les éditeurs des modèles évalués (STARD-AI).

II. RESULTATS :

1. FLUX DE SÉLECTION DES DONNÉES

Un total de 136 radiographies thoraciques avait initialement été prévu. Après application des critères d'exclusion, 26 radiographies ont été écartées (mauvaise qualité technique jugée par le gold standard : $n = 26$), portant l'effectif final analysé à 110 radiographies.

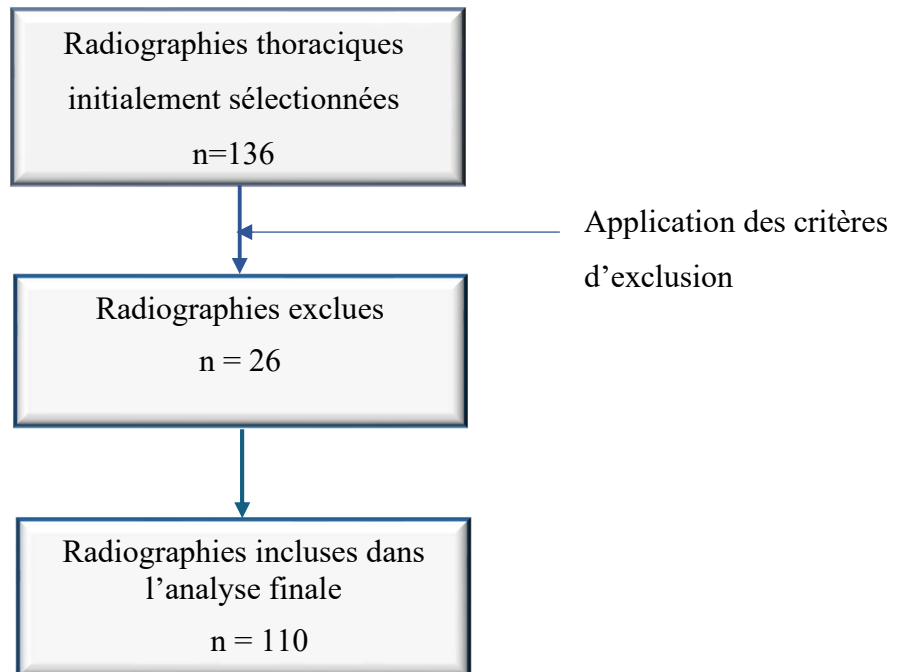


Figure 11. Flux de sélection des données

2. CARACTERISTIQUES DE LA POPULATION ET DU JEU DE DONNEES :

2.1. Caractéristiques générales

La population étudiée était majoritairement masculine, avec 76 hommes représentant 69.1% de l'effectif total, contre 34 femmes (30.9 %). L'âge moyen des participants était de 54,25 $\pm$ 19,7 ans, indiquant une population relativement âgée avec une grande variabilité des âges.

Caractéristique	Effectif (n = 110)	Proportion (%) ou Moy ± ET
Sexe masculin	76	69.1 %
Sexe féminin	34	30,9 %
Âge moyen (années)	54.25	54.25± 19.7

Tableau 3. Caractéristiques générales de la population d'étude.

2.2. Répartition selon les entités diagnostiques (gold standard)

La pneumonie constituait l'entité diagnostique la plus fréquente dans le jeu de données, avec 47 cas (42,7 %), suivie de l'épanchement pleural liquidien avec 35 cas (31,8 %). Le pneumothorax représentait 28 cas (25,5 %), tandis que 20 examens (18,2 %) étaient considérés comme radiologiquement normaux.

Note : Le total des proportions dépasse 100 % en raison de la présence de clichés présentant plusieurs anomalies concomitantes (catégories non mutuellement exclusives).

Entité diagnostique	n	Proportion (%)
Épanchement pleural liquidien	35	31,8
Pneumonie	47	42,7
Pneumothorax	28	25,5
Aspect radiologique normal	20	18,2

Tableau 4. Répartition du jeu de données selon le diagnostic de référence.

3. ÉTUDE DES PERFORMANCES DIAGNOSTIQUES DU MODELE

MAIA.IA :

- **DIAGNOSTIC DE L'ÉPANCHEMENT PLEURAL**

	Gold standard : Positif (n=35)	Gold standard : Négatif (n=75)	Total
MAIA AI : Positif	VP = 31	FP = 4	35
MAIA AI : Négatif	FN = 4	VN = 71	75
Total	35	75	110

Tableau 5. Matrice de confusion – MAIA AI, Épanchement pleural.

Mesure	Valeur	IC 95 %
Sensibilité	88,6 %	74,0 – 95,5 %
Spécificité	94,7 %	87,1 – 97,9 %
VPP	88,6 %	74,0 – 95,5 %
VPN	94,7 %	87,1 – 97,9 %
Exactitude	92,7 %	86,3 – 96,3 %
RV+	16,61	–
RV–	0,12	–
AUC (ROC)	0,916	84,8 – 98,4 %

Tableau 6. Performances diagnostiques – MAIA AI, Épanchement pleural

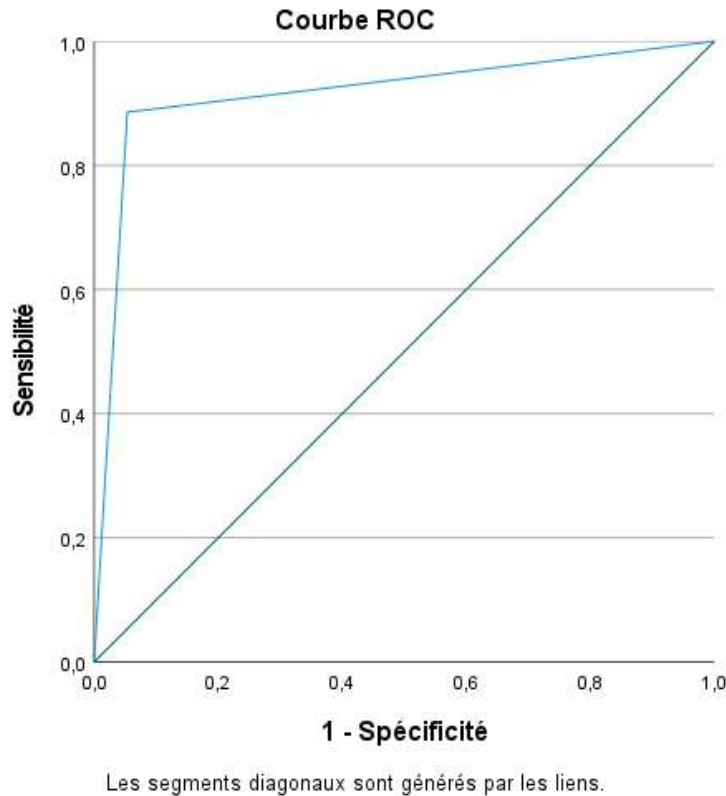


Figure 12. Courbe ROC d'entité diagnostique : Epanchement pleural -MAIA AI

➤ **Commentaires :**

Le tableau ci-dessus compare les résultats diagnostiques de MAIA AI à ceux du gold standard pour la détection de l'épanchement pleural sur 110 radiographies thoraciques. Parmi les 35 cas confirmés d'épanchement pleural par le gold standard, MAIA AI en a correctement identifié 31, tandis que 4 cas n'ont pas été détectés (faux négatifs). Parmi les 75 cas ne présentant pas d'épanchement pleural, 71 ont été correctement classés comme négatifs, alors que 4 ont été à tort considérés comme positifs (faux positifs). Ces résultats correspondent à une sensibilité de 88,6 %, une spécificité de 94,7 %, une valeur prédictive positive de 88,6 %, une valeur prédictive négative de 94,7 % et une exactitude diagnostique de 92,7 %. Ces performances témoignent d'une excellente capacité de MAIA AI à détecter l'épanchement pleural, avec un faible taux d'erreurs diagnostiques. Le rapport de vraisemblance positif élevé ($RV+ = 16,61$) indique une forte valeur confirmatoire d'un résultat positif, tandis que le $RV-$ faible (0,12) suggère une bonne capacité à exclure la pathologie en cas de résultat négatif. Enfin, l'aire sous la courbe ROC : $AUC = 0,916$ (IC 95% 84,8 – 98,4 % ; $p < 0,001$), reflète une excellente capacité discriminative globale du modèle pour cette entité diagnostique.

- **DIAGNOSTIC DE PNEUMONIE**

	Gold standard : Positif (n=47)	Gold standard : Négatif (n=63)	Total
MAIA AI : Positif	VP = 34	FP = 4	38
MAIA AI : Négatif	FN = 13	VN = 59	72
Total	47	63	110

Tableau 7. Matrice de confusion – MAIA AI, Pneumonie.

Mesure	Valeur	IC 95 %
Sensibilité	72,3 %	58,2 – 83,1 %
Spécificité	93,7 %	84,8 – 97,5 %
VPP	89,5 %	75,9 – 95,8 %
VPN	81,9 %	71,5 – 89,1 %
Exactitude	84,5 %	76,6 – 90,1 %
RV+	11,39	—
RV–	0,30	—
AUC (ROC)	0,830	74.5 – 91.5 %

Tableau 8. Performances diagnostiques – MAIA AI, Pneumonie

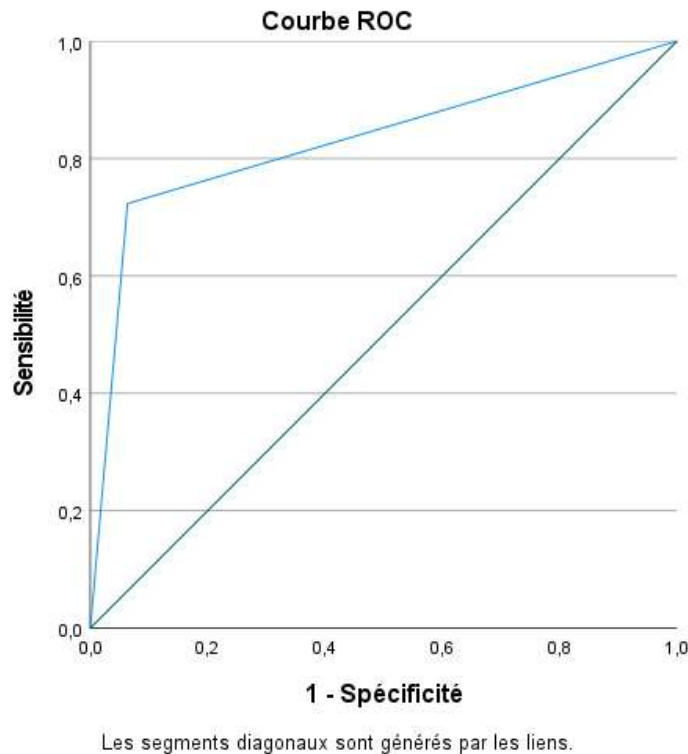


Figure 13. Courbe ROC d'entité diagnostique : Pneumonie -MAIA AI –

➤ **Commentaires :**

Le tableau ci-dessus compare les résultats diagnostiques de MAIA AI à ceux du gold standard pour la détection de la pneumonie sur 110 radiographies thoraciques. Parmi les 47 cas de pneumonie confirmés par le gold standard, le modèle MAIA AI en a correctement identifié 34, tandis que 13 cas n'ont pas été détectés, correspondant à des faux négatifs. Parmi les 63 cas ne présentant pas de pneumonie, 59 ont été correctement classés comme négatifs, alors que 4 ont été à tort considérés comme positifs. Ces résultats correspondent à une sensibilité de 72,3 %, une spécificité de 93,7 %, une valeur prédictive positive de 89,5 %, une valeur prédictive négative de 81,9 % et une exactitude diagnostique globale de 84,5 %. Ces performances indiquent que MAIA AI présente une très bonne spécificité et une excellente valeur prédictive positive, traduisant une bonne capacité du modèle à confirmer la maladie lorsqu'il est positif. En revanche, la sensibilité plus modérée traduit une tendance du modèle à méconnaître certains cas de pneumonie. Le rapport de vraisemblance positif élevé ($RV+ = 11,39$) indique une forte valeur confirmatoire d'un résultat positif, tandis que le $RV-$ (0,30) témoigne d'une capacité modérée à réduire la probabilité de pneumonie en cas de test négatif. Enfin, l'aire sous la courbe ROC : $AUC = 0,830$ (IC 95% 74.5 – 91.5 % ; $p < 0,001$), reflète une bonne capacité discriminative globale du modèle pour cette pathologie.

- **DIAGNOSTIC PNEUMOTHORAX**

	Gold standard : Positif (n=28)	Gold standard : Négatif (n=82)	Total
MAIA AI : Positif	VP = 18	FP = 1	19
MAIA AI : Négatif	FN = 10	VN = 81	91
Total	28	82	110

Tableau 9. Matrice de confusion – MAIA AI, Pneumothorax.

Mesure	Valeur	IC 95 %
Sensibilité	64,3 %	45,8 – 79,3 %
Spécificité	98,8 %	93,4 – 99,8 %
VPP	94,7 %	75,4 – 99,1 %
VPN	89,0 %	80,9 – 93,9 %
Exactitude	90,0 %	83,0 – 94,3 %
RV+	52,71	–
RV–	0,36	–
AUC (ROC)	0,815	70,3 – 92,8 %

Tableau 10. Performances diagnostiques – MAIA AI, Pneumothorax.

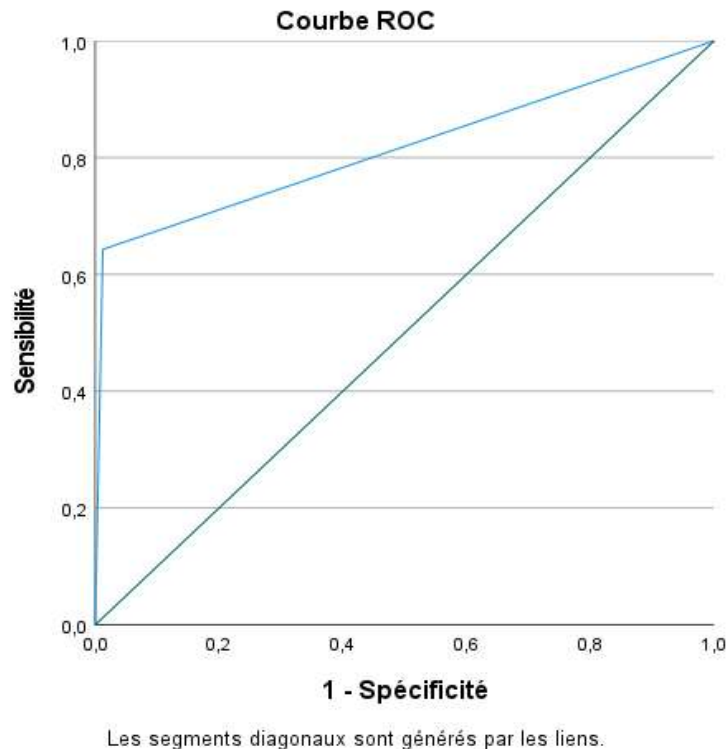


Figure 14. Courbe ROC d'entité diagnostique : Pneumothorax -MAIA AI

➤ **Commentaires :**

Le tableau ci-dessus compare les résultats diagnostiques de MAIA AI à ceux du gold standard pour la détection du pneumothorax sur 110 radiographies thoraciques. Parmi les 28 cas de pneumothorax confirmés par le gold standard, le modèle en a correctement identifié 18, tandis que 10 cas n'ont pas été détectés, correspondant à des faux négatifs. Parmi les 82 cas ne présentant pas de pneumothorax, 81 ont été correctement classés comme négatifs, alors qu'un seul cas a été à tort considéré comme positif. Ces résultats correspondent à une sensibilité de 64,3 %, une spécificité de 98,8 %, une valeur prédictive positive de 94,7 %, une valeur prédictive négative de 89,0 % et une exactitude diagnostique globale de 90,0 %. Ils montrent que MAIA AI possède une excellente spécificité et une valeur prédictive positive élevée, indiquant qu'un diagnostic positif de pneumothorax est généralement très fiable. En revanche, la sensibilité modérée traduit une capacité limitée à détecter l'ensemble des cas, avec plus d'un tiers des pneumothorax non reconnus). Le rapport de vraisemblance positif très élevé ($RV+ = 52,71$) indique une valeur confirmatoire exceptionnelle d'un résultat positif, tandis que le $RV-$ (0,36) reflète une capacité modérée à réduire la probabilité de pneumothorax en cas de test négatif. Enfin, l'aire sous la courbe ROC : $AUC = 0,815$ (IC

95% 70,3 – 92,8 % ; $p < 0,001$), traduit une bonne capacité discriminative globale du modèle pour cette pathologie, malgré une sensibilité relativement limitée.

- **DIAGNOSTIC DES CLICHES NORMAUX**

	Gold standard : Positif (n=20)	Gold standard : Négatif (n=90)	Total
MAIA AI : Positif	VP = 20	FP = 8	28
MAIA AI : Négatif	FN = 0	VN = 82	82
Total	20	90	110

Tableau 11. Matrice de confusion – MAIA AI, Aspect normal.

Mesure	Valeur	IC 95 %
Sensibilité	100,0 %	83,9 – 100 %
Spécificité	91,1 %	83,4 – 95,4 %
VPP	71,4 %	52,9 – 84,7 %
VPN	100,0 %	95,5 – 100 %
Exactitude	92,7 %	86,3 – 96,3 %
RV+	11,25	–
RV–	0,00	–
AUC (ROC)	0,956	91,9 – 99,2 %

Tableau 12. Performances diagnostiques – MAIA AI, Aspect normal.

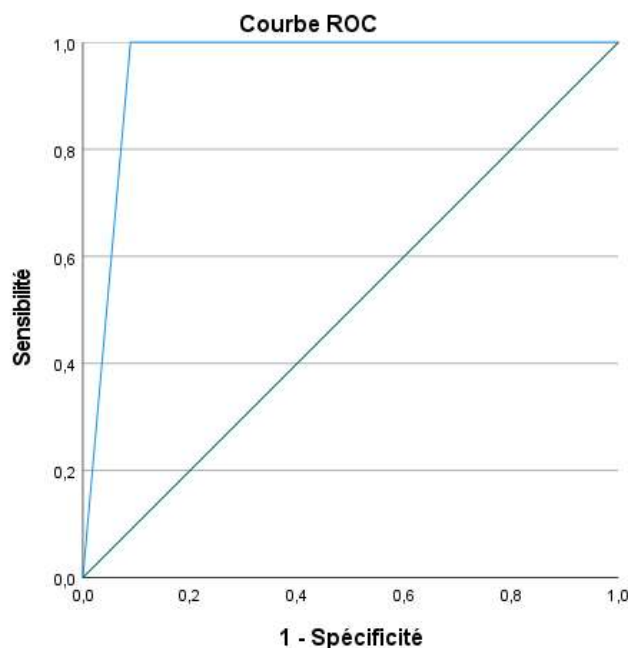


Figure 15. Courbe ROC d'entité diagnostique : TLT normal -MAIA AI -

➤ **Commentaires :**

Le tableau de contingence ci-dessus compare les résultats diagnostiques de MAIA AI à ceux du gold standard pour l'identification des clichés thoraciques normaux sur 110 radiographies. Parmi les 20 clichés considérés comme normaux par le gold standard, le modèle MAIA AI les a tous correctement identifiés, sans aucun faux négatif. En revanche, parmi les 90 clichés comportant une anomalie radiologique, 82 ont été correctement classés comme non normaux, tandis que 8 ont été à tort interprétés comme normaux, correspondant à des faux positifs. Ces résultats correspondent à une sensibilité de 100 %, une spécificité de 91,1 %, une valeur prédictive positive de 71,4 %, une valeur prédictive négative de 100 % et une exactitude diagnostique globale de 92,7 %. Ils montrent que MAIA AI possède une excellente capacité à reconnaître les clichés réellement normaux, sans méconnaître aucun cas. Toutefois, la valeur prédictive positive plus modérée indique qu'une proportion non négligeable de clichés jugés normaux par le modèle présentait en réalité une anomalie radiologique. Le rapport de vraisemblance positif ($RV+ = 11,25$) témoigne d'une bonne valeur confirmatoire d'un résultat indiquant un aspect normal, tandis que le $RV-$ nul (0,00) traduit une capacité maximale à exclure toute pathologie en cas de résultat négatif pour l'anomalie. Enfin, l'aire sous la courbe ROC : $AUC = 0,956$ (IC 95% 91,9 – 99,2 % ; $p < 0,001$) reflète une excellente capacité discriminative globale du modèle pour distinguer un aspect normal des images pathologiques.

- **EVALUATION DE PERFORMANCE DIAGNOSTIC GLOBALE DE MAIA.IA**

	EPANCHEMEN T	PNEUMONI E	PNEUMOTHORA X	TLT NORMAL
Sensibilité (%)	88.6	72.3	64.3	100
Spécificité (%)	94.7	93.6	98.8	91.1
VPP (%)	88.6	89.5	94.7	71.4
VPN (%)	94.7	81.9	89.0	100
Exactitude	92.7	84.5	90.0	92.7

Tableau 13. Performances diagnostiques – MAIA AI.

➤ **Commentaires :**

L'analyse des performances diagnostiques de l'IA montre des résultats satisfaisants pour l'ensemble des pathologies étudiées. La détection de l'épanchement pleural présente les meilleures performances avec une sensibilité de 88,6 % et une spécificité de 94,7 %, témoignant d'une bonne capacité du modèle à identifier cette pathologie. Pour le téléthorax normal, l'IA atteint une sensibilité parfaite de 100 % et une VPN de 100 %, ce qui signifie qu'aucun cas normal n'a été classé à tort comme pathologique, bien que la spécificité soit légèrement inférieure à 91,1 %. En revanche, la pneumonie enregistre la sensibilité la plus basse (72,3 %), suggérant une difficulté relative du modèle à détecter les condensations pulmonaires, notamment dans les formes peu typiques. Le pneumothorax, quant à lui, affiche une spécificité très élevée de 98,8 % et une VPP de 94,7 %, indiquant que lorsque le modèle détecte un pneumothorax, ce diagnostic est presque toujours confirmé. L'exactitude globale varie entre 84,5 % pour la pneumonie et 92,7 % pour l'épanchement pleural et le téléthorax normal, reflétant des performances diagnostiques globalement bonnes de MAIA AI dans l'interprétation du téléthorax. Enfin, les AUC reflètent une capacité discriminative globale bonne à excellente du modèle pour les quatre entités diagnostiques étudiées.

Note : L'AUC a été calculée à partir des sorties binaires du modèle (0/1). Bien que mathématiquement valide, l'interprétation de la courbe ROC reste limité en raison de

l'absence de scores probabilistes continus et de seuils décisionnels multiples dans l'interprétation du modèle MAIA AI.

4. ÉTUDE DES PERFORMANCES DIAGNOSTIQUES DU MODELE GPT 5.2

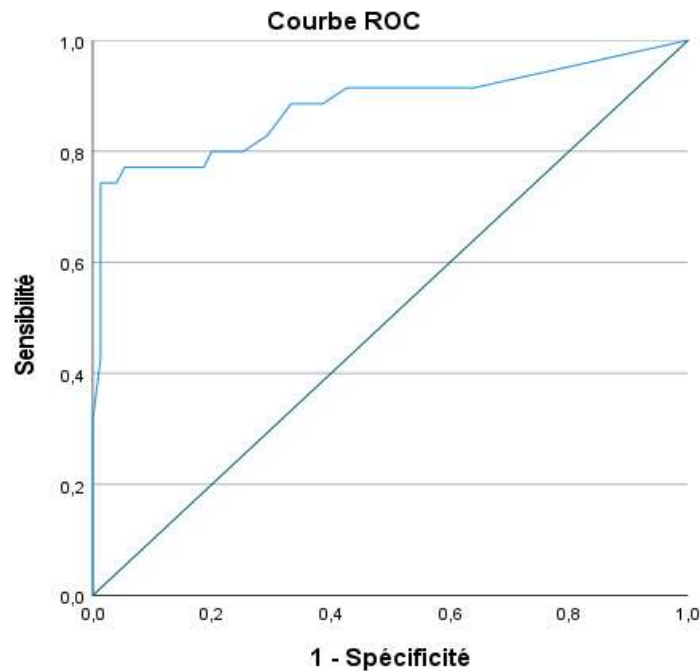
- **DIAGNOSTIC DE L'ÉPANCHEMENT PLEURAL**

	Gold standard : Positif (n=35)	Gold standard : Négatif (n=75)	Total
GPT-5.2 : Positif	VP = 27	FP = 3	30
GPT-5.2 : Négatif	FN = 8	VN = 72	80
Total	35	75	110

Tableau 14. Matrice de confusion – GPT-5.2, Épanchement pleural.

Mesure	Valeur	IC 95 %
Sensibilité	77,1 %	61,0 – 87,9 %
Spécificité	96,0 %	88,9 – 98,6 %
VPP	90,0 %	74,4 – 96,5 %
VPN	90,0 %	81,5 – 94,8 %
Exactitude	90,0 %	83,0 – 94,3 %
RV+	19,29	—
RV–	0,24	—
AUC (ROC)	0,881	79,8 – 96,7 %

Tableau 15. Performances diagnostiques – GPT-5.2, Épanchement pleural



Les segments diagonaux sont générés par les liens.

Figure 16. Courbe ROC d'entité diagnostique : Epanchement pleural -GPT 5.2-

➤ **Commentaires :**

Le tableau de contingence ci-dessus compare les résultats diagnostiques de X-Ray & Radiologic GPT-5.2 à ceux du gold standard pour la détection de l'épanchement pleural sur 110 radiographies thoraciques. Parmi les 35 cas d'épanchement pleural confirmés par le gold standard, le modèle en a correctement identifié 27, tandis que 8 cas n'ont pas été détectés, correspondant à des faux négatifs. Parmi les 75 cas sans épanchement pleural, 72 ont été correctement classés comme négatifs et 3 ont été à tort considérés comme positifs. Ces résultats correspondent à une sensibilité de 77,1 %, une spécificité de 96,0 %, une valeur prédictive positive de 90,0 %, une valeur prédictive négative de 90,0 % et une exactitude diagnostique globale de 90,0 %. Ils montrent que GPT-5.2 présente de bonnes performances diagnostiques pour la détection de l'épanchement pleural. La spécificité élevée indique que le modèle génère peu de faux positifs, tandis que la valeur prédictive positive de 90 % signifie qu'un diagnostic positif posé par le modèle est généralement fiable. La sensibilité, bien que satisfaisante, reste inférieure à la spécificité, traduisant l'existence de quelques cas non détectés. Par ailleurs, le rapport de vraisemblance positif élevé ($RV+ = 19,29$) suggère une forte valeur confirmatoire d'un résultat positif, tandis que le $RV-$ de 0,24 indique une capacité modérée à exclure la maladie en cas de résultat négatif. L'analyse ROC a montré une bonne capacité discriminante du modèle pour cette pathologie avec une aire sous la courbe (AUC) de 0,881 (IC95% : 0,798–0,964 ; $p < 0,001$).

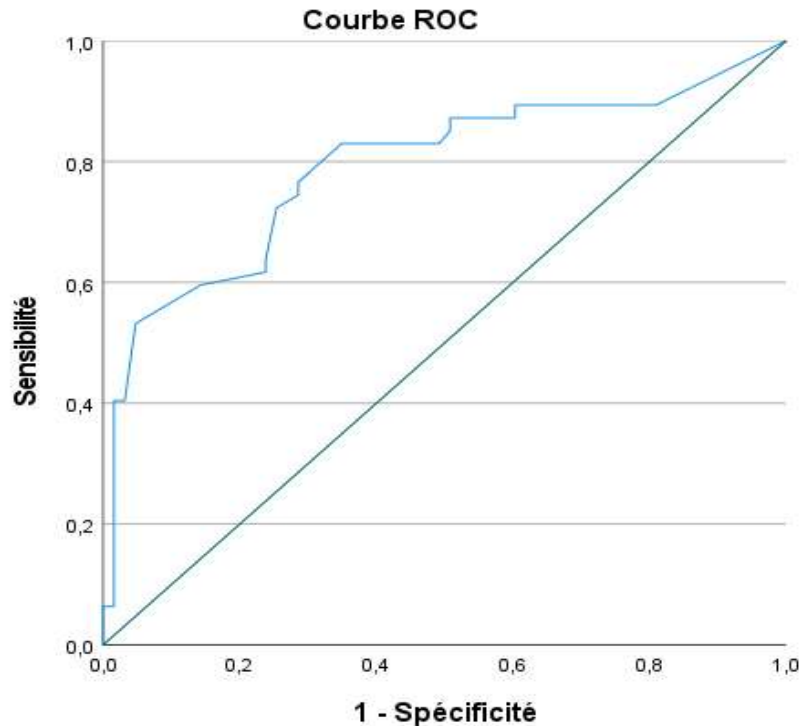
- **DIAGNOSTIC DE PNEUMONIE**

	Gold standard : Positif (n=47)	Gold standard : Négatif (n=63)	Total
GPT-5.2 : Positif	VP = 39	FP = 20	59
GPT-5.2 : Négatif	FN = 8	VN = 43	51
Total	47	63	110

Tableau 16. Matrice de confusion – GPT-5.2, Pneumonie.

Mesure	Valeur	IC 95 %
Sensibilité	83,0 %	69,9 – 91,1 %
Spécificité	68,3 %	56,0 – 78,4 %
VPP	66,1 %	53,4 – 76,9 %
VPN	84,3 %	72,0 – 91,8 %
Exactitude	74,5 %	65,7 – 81,8 %
RV+	2,61	—
RV–	0,25	—
AUC (ROC)	0,791	69,9 – 88.2 %

Tableau 17. Performances diagnostiques – GPT-5.2, Pneumonie.



Les segments diagonaux sont générés par les liens.

Figure 17. Courbe ROC d'entité diagnostique : Pneumonie -GPT 5.2-

➤ Commentaires :

Le tableau ci-dessus compare les résultats diagnostiques de X-Ray & Radiologic GPT-5.2 à ceux du gold standard pour la détection de la pneumonie sur 110 radiographies thoraciques. Parmi les 47 cas de pneumonie confirmés par le gold standard, le modèle en a correctement identifié 39, tandis que 8 cas n'ont pas été détectés, correspondant à des faux négatifs. Parmi les 63 cas ne présentant pas de pneumonie, 43 ont été correctement classés comme négatifs, alors que 20 cas ont été à tort considérés comme positifs. Ces résultats correspondent à une sensibilité de 83,0 %, une spécificité de 68,3 %, une valeur prédictive positive de 66,1 %, une valeur prédictive négative de 84,3 % et une exactitude diagnostique globale de 74,5 %. Ils montrent que X-Ray & Radiologic GPT-5.2 présente une bonne sensibilité pour la détection de la pneumonie, permettant d'identifier la majorité des cas confirmés. En revanche, la spécificité plus modérée traduit un nombre relativement important de faux positifs, ce qui réduit la fiabilité d'un diagnostic positif. Le rapport de vraisemblance positif ($RV+ = 2,61$) indique une valeur confirmatoire limitée d'un résultat positif, alors que le $RV-$ de 0,25 reflète une capacité modérée à exclure la maladie en cas de résultat négatif. L'analyse ROC a montré une capacité discriminante acceptable du modèle pour cette pathologie avec une aire sous la courbe (AUC) de 0,791 (IC95% 69,9 – 88.2 % ; $p < 0,001$).

- **DIAGNOSTIC DE PNEUMOTHORAX**

	Gold standard : Positif (n=28)	Gold standard : Négatif (n=82)	Total
GPT-5.2 : Positif	VP = 7	FP = 0	7
GPT-5.2 : Négatif	FN = 21	VN = 82	103
Total	28	82	110

Tableau 18. Matrice de confusion – GPT-5.2, Pneumothorax.

Mesure	Valeur	IC 95 %
Sensibilité	25,0 %	12,7 – 43,4 %
Spécificité	100,0 %	95,5 – 100 %
VPP	100,0 %	64,6 – 100 %
VPN	79,6 %	70,8 – 86,3 %
Exactitude	80,9 %	72,6 – 87,2 %
RV+	Indéfini (Sp = 100 %)	–
RV–	0,75	–
AUC (ROC)	0,755	63.8 – 87.3 %

Tableau 19. Performances diagnostiques – GPT-5.2, Pneumothorax.

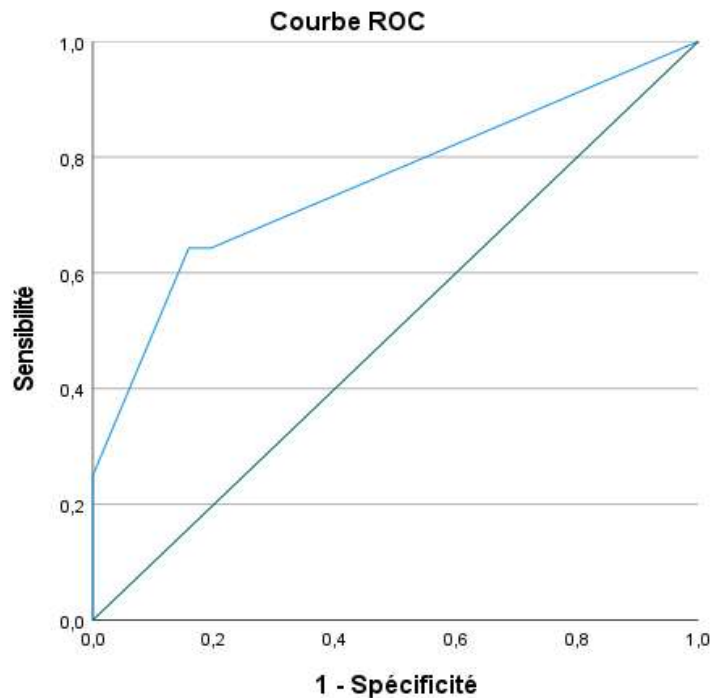


Figure 18. Courbe ROC d'entité diagnostique : Pneumothorax -GPT 5.2-

➤ **Commentaires :**

Le tableau ci-dessus compare les résultats diagnostiques de X-Ray & Radiologic GPT-5.2 à ceux du gold standard pour la détection du pneumothorax sur 110 radiographies thoraciques. Parmi les 28 cas de pneumothorax confirmés par le gold standard, le modèle en a correctement identifié 7, tandis que 21 cas n'ont pas été détectés, correspondant à des faux négatifs. Parmi les 82 cas ne présentant pas de pneumothorax, tous ont été correctement classés comme négatifs, sans aucun faux positif observé. Ces résultats correspondent à une sensibilité de 25,0 %, une spécificité de 100,0 %, une valeur prédictive positive de 100,0 %, une valeur prédictive négative de 79,6 % et une exactitude diagnostique globale de 80,9 %. Ils montrent que X-Ray & Radiologic GPT-5.2 présente une très faible sensibilité pour la détection du pneumothorax, n'identifiant qu'un quart des cas confirmés. En revanche, sa spécificité maximale traduit une excellente capacité à reconnaître les examens ne présentant pas cette pathologie, sans générer de faux positifs. Ainsi, lorsqu'un pneumothorax est identifié par le modèle, le diagnostic apparaît hautement fiable. Le rapport de vraisemblance positif ($RV+ = 2,61$) indique une valeur confirmatoire limitée d'un résultat positif, alors que le $RV-$ de 0,25 reflète une capacité modérée à exclure la maladie en cas de résultat négatif.

L'analyse ROC a montré une capacité discriminante acceptable du modèle pour cette pathologie avec une aire sous la courbe (AUC) de 0,791 (IC95% 69,9 – 88.2 % ; $p < 0,001$).

Globalement, le modèle apparaît très spécifique mais peu sensible pour le pneumothorax. Il est donc plus adapté comme outil de confirmation diagnostique que comme outil de dépistage, où une sensibilité élevée est indispensable.

- **DIAGNOSTIC DES CLICHES NORMAUX**

	Gold standard : Positif (n=20)	Gold standard : Négatif (n=90)	Total
GPT-5.2 : Positif	VP = 17	FP = 2	19
GPT-5.2 : Négatif	FN = 3	VN = 88	91
Total	20	90	110

Tableau 20. Matrice de confusion – GPT-5.2, Aspect normal.

Mesure	Valeur	IC 95 %
Sensibilité	85,0 %	64,0 – 94,8 %
Spécificité	97,8 %	92,3 – 99,4 %
VPP	89,5 %	68,6 – 97,1 %
VPN	96,7 %	90,8 – 98,9 %
Exactitude	95,5 %	89,8 – 98,0 %
RV+	38,25	—
RV–	0,15	—
AUC (ROC)	0,933	85.2 – 100 %

Tableau 21. Performances diagnostiques – GPT-5.2, Aspect normal.

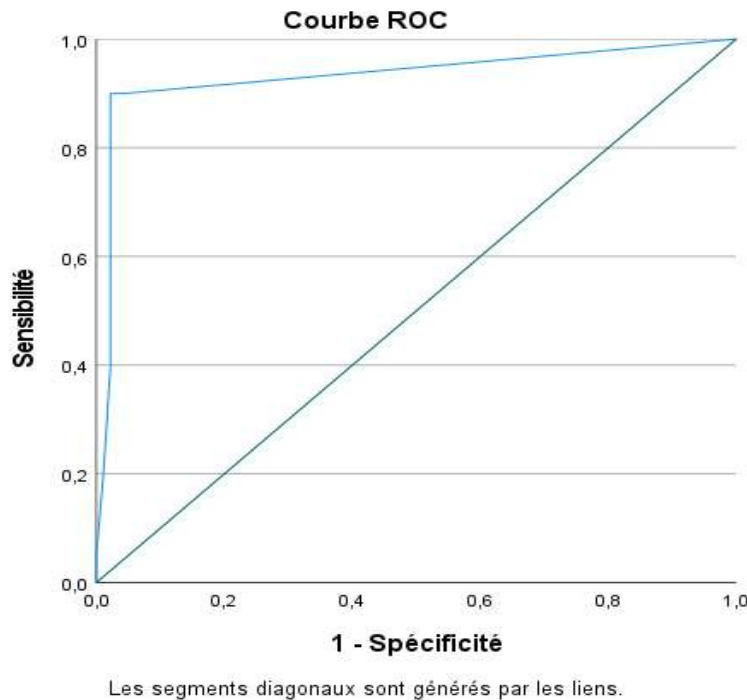


Figure 19. Courbe ROC d'entité diagnostique : tlt normale -GPT 5.2-

➤ **Commentaires :**

Le tableau ci-dessus compare les résultats diagnostiques de GPT-5.2 à ceux du gold standard pour l'identification des clichés thoraciques d'aspect normal sur 110 radiographies thoraciques. Parmi les 20 clichés considérés comme normaux selon le gold standard, le modèle en a correctement identifié 17, tandis que 3 cas n'ont pas été reconnus, correspondant à des faux négatifs. Parmi les 90 clichés présentant une anomalie, 88 ont été correctement classés comme non normaux, alors que 2 cas ont été à tort considérés comme normaux, correspondant à des faux positifs. Ces résultats correspondent à une sensibilité de 85,0 %, une spécificité de 97,8 %, une valeur prédictive positive de 89,5 %, une valeur prédictive négative de 96,7 % et une exactitude diagnostique globale de 95,5 %. Ils montrent que GPT-5.2 présente de très bonnes performances pour l'identification des radiographies thoraciques normales, avec une capacité élevée à reconnaître correctement les examens sans anomalie tout en limitant les erreurs diagnostiques. Le rapport de vraisemblance positif élevé ($RV+ = 38,25$) indique une forte valeur diagnostique d'un résultat positif, tandis que le $RV-$ de 0,15 suggère une bonne capacité à exclure un aspect normal en cas de résultat négatif. L'analyse ROC a montré une excellente capacité discriminante du modèle avec une aire sous la courbe (AUC) de 0,933 (IC 95% 85.2 – 100 % ; $p < 0,001$).

• **EVALUATION DE PERFORMANCE DIAGNOSTIC GLOBALE**
DE GPT 5.2

	EPANCHEMENT	PNEUMONIE	PNEUMOTHORAX	TLT NORMAL
Sensibilité (%)	77.1	83.0	25.0	90.0
Spécificité (%)	96.0	68.3	100	97.8
VPP (%)	90.0	66.1	100	90.0
VPN (%)	90.0	84.3	79.6	97.8
Exactitude (%)	90.0	74.5	80.9	96.4

Tableau 22. Performances diagnostiques globales –GPT 5.2-.

➤ **Commentaires :**

Le tableau ci-dessus présente les performances diagnostiques du GPT pour les différentes anomalies radiologiques étudiées. Les meilleurs résultats ont été observés pour les clichés thoraciques normaux, avec une exactitude de 96,4 %, une sensibilité de 90 % et une spécificité de 97,8 %. Le modèle a également montré de bonnes performances dans le diagnostic de l'épanchement pleural, avec une exactitude de 90 %.

Concernant la pneumonie, la sensibilité était satisfaisante (83 %), mais la spécificité et la valeur prédictive positive étaient plus faibles (68,3 % et 66,1 % respectivement), traduisant un nombre plus élevé de faux positifs. Pour le pneumothorax, l'IA présentait une excellente spécificité et une VPP de 100 %, indiquant l'absence de faux positifs, mais une faible sensibilité (25 %), suggérant une capacité limitée à détecter le pneumothorax. Enfin, les AUC reflètent une capacité discriminative globale acceptable à excellente du modèle pour les quatre entités diagnostiques étudiées.

Globalement, les performances diagnostiques de GPT-5.2 variaient selon la pathologie étudiée, avec de meilleurs résultats pour les anomalies évidentes radiologiquement et les clichés normaux.

5. ANALYSE DE CONCORDANCE :

• CONCORDANCE DIAGNOSTIQUE :

➤ MAIA.AI :

Entité diagnostique	MAIA AI : Kappa (IC 95 %)
Épanchement pleural	0,832
Pneumonie	0,676
Pneumothorax	0,705
Aspect normal	0,788

Tableau N23 : concordance diagnostique de MAIA.AI

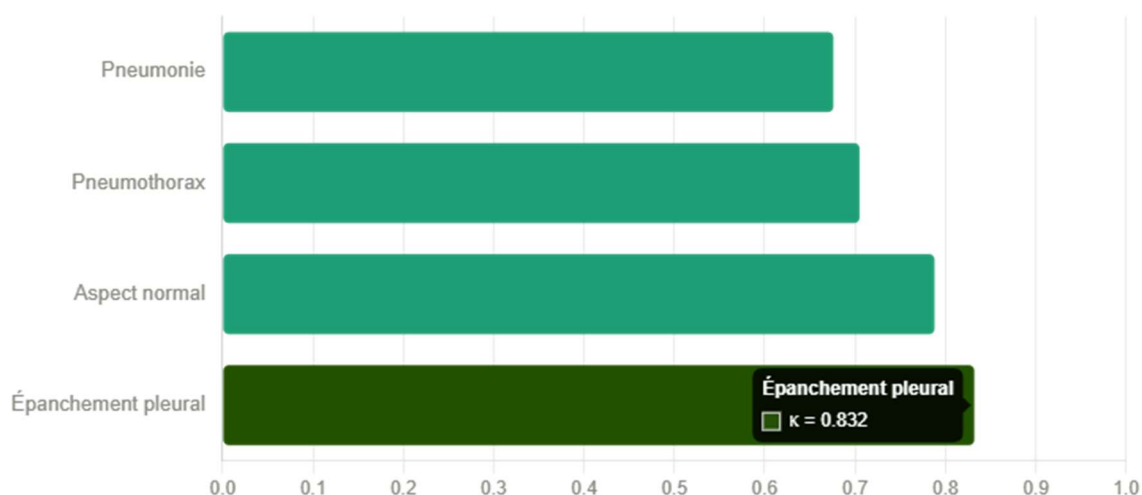


Figure 20 : concordances diagnostiques de MAIA.AI

➤ Commentaires :

MAIA AI se distingue par sa capacité à maintenir une concordance substantielle à quasi-parfaite sur l'ensemble du spectre diagnostique. Sa progression sur le pneumothorax (+0,373 vs GPT-5.2) est cliniquement la plus importante, car il s'agit d'une urgence radiologique où un faux négatif peut engager le pronostic vital. Sa seule légère faiblesse relative concerne l'aspect normal (0,788 vs 0,844), à surveiller pour éviter les faux positifs inutiles.

➤ **GPT 5.2 :**

Entité diagnostique	GPT-5.2 : Kappa (IC 95 %)
Épanchement pleural	0,760
Pneumonie	0,496
Pneumothorax	0,332
Aspect normal	0,844

Tableau N24 : Concordances diagnostiques de GPT 5.2

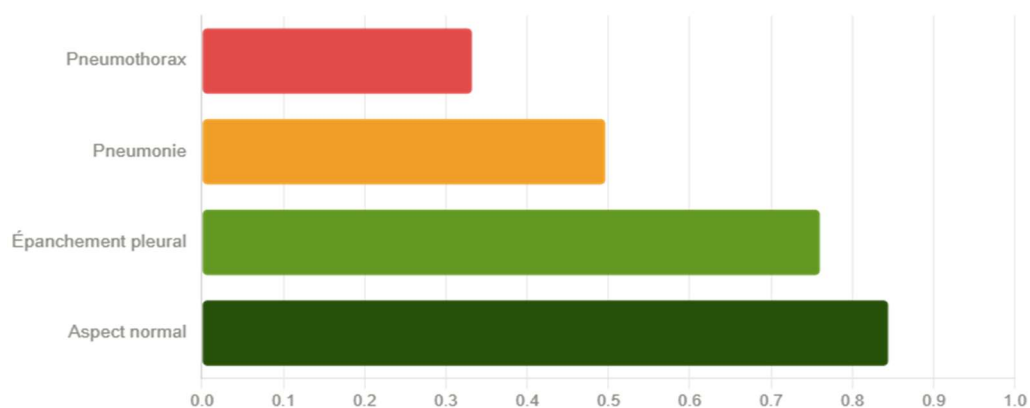


Figure 21 : Concordances diagnostiques de GPT 5.2

➤ **Commentaires :**

GPT-5.2 excelle sur les entités visuellement marquées et bien définies — aspect normal (0,844) et épanchement pleural (0,760) — mais chute significativement sur le pneumothorax (0,332), une pathologie urgente. Cette hétérogénéité limite sa fiabilité en tant qu'outil d'aide au diagnostic polyvalent.

- **CONCORDANCE SÉMIOLOGIQUE :**

➤ **MAIA.IA**

Entité diagnostique	Signes radiologiques	MAIA AI : Kappa / Kappa pondéré (IC 95 %)	Interprétation
Épanchement pleural	Opacité pleurale	0,876	Très bon accord
	Côté	0,782	Bon accord
	Abondance	0,885	Très bon accord
	Signe de refoulement	0,316	Accord passable
Pneumonie	Opacité alvéolaire/interstitielle	0,676	Bon accord
	Côté	0,535	Accord modéré
Pneumothorax	Hyper clarté	0,695	Bon accord
	Côté	0,742	Bon accord
	Liseré	0,628	Bon accord
	Abondance	0,803	Bon accord
	Hydropneumothorax	0,798	Bon accord
Aspect normal	Interprétation sans anomalie.	0,788	Bon accord

Tableau 25. Concordances sémiologiques -MAIA.IA-

➤ **Commentaires :**

Dans l'ensemble, la concordance sémiologique de MAIA AI est bonne à très bonne pour la majorité des signes radiologiques étudiés, avec des valeurs de Kappa comprises entre 0,628 et 0,885. Le modèle démontre une cohérence interne satisfaisante entre ses diagnostics et les signes sémiologiques qu'il décrit, ce qui renforce sa valeur en tant qu'outil d'aide à la décision clinique. La seule faiblesse notable concerne le signe de refoulement médiastinal ($\kappa = 0,316$) dans l'épanchement pleural, et la latéralisation de la pneumonie ($\kappa = 0,535$), deux signes sémiologiques dont la détection nécessite une analyse contextuelle plus fine que ce que le modèle semble capable de fournir de façon systématique.

➤ GPT 5.2

Entité diagnostique	Signes radiologiques	GPT-5.2 : Kappa / Kappa pondéré (IC 95 %)	Interprétation
Épanchement pleural	Opacité pleurale	0,760	Bon accord
	Côté	0,551	Accord modéré
	Abondance	0,858	Très bon accord
	Signe de refoulement	0,218	Accord passable
Pneumonie	Opacité alvéolaire/interstitielle	0,469	Accord modéré
	Côté	0,469	Accord modéré
Pneumothorax	Hyper clarté	0,217	Accord passable
	Côté	0,298	Accord passable
	Liseré	0,246	Accord passable
	Abondance	0,432	Accord modéré
	Hydropneumothorax	0,645	Bon accord
Aspect normal	Interprétation sans anomalie.	0,844	Très bon accord

Tableau 26. Concordances sémiologiques -GPT 5.2 -

➤ Commentaires :

De manière globale, la concordance sémiologique de GPT-5.2 est variable selon les entités diagnostiques, avec des valeurs de Kappa allant de 0,217 à 0,858. Le modèle démontre de bonnes performances sémiologiques pour l'aspect radiologique normal et l'abondance de l'épanchement pleural, mais présente des lacunes importantes pour le pneumothorax et la pneumonie. Ces résultats indiquent que la cohérence interne du raisonnement diagnostique de GPT-5.2 est inégale, performante pour les tableaux radiologiques simples et bien définis, mais insuffisante pour les pathologies nécessitant une analyse sémiologique fine et multidimensionnelle comme le pneumothorax.

6. COMPARAISON INTER-MODELES :

• COMPARAISON DES PERFORMANCES DIAGNOSTIQUES GLOBALES

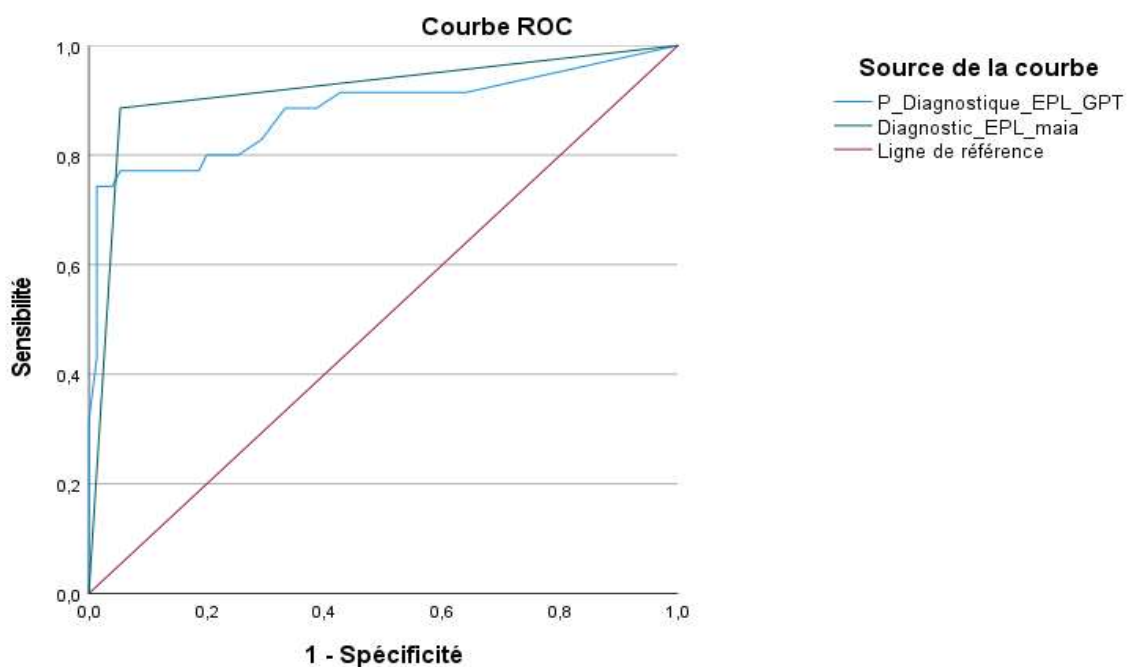
Entité diagnostique	Mesure	GPT-5.2	MAIA AI	Valeur P (McNemar)
Épanchement pleural	Sensibilité (%)	77,1	88,6	0,344
	Spécificité (%)	96,0	94,7	1
	Exactitude (%)	90,0	92,7	0,267
	AUC	0,881	0,916	-
Pneumonie	Sensibilité (%)	83,0	72,3	0,227
	Spécificité (%)	68,3	93,7	Inférieur à 0,001
	Exactitude (%)	74,5	84,5	Inférieur à 0,001
	AUC	0,791	0,830	-
Pneumothorax	Sensibilité (%)	25,0	64,3	Inférieur à 0,001
	Spécificité (%)	100,0	98,8	Inférieur à 0,001
	Exactitude (%)	80,9	90,0	Inférieur à 0,001
	AUC	0,755	0,815	-
Aspect normal	Sensibilité (%)	85,0	100,0	Inférieur à 0,001
	Spécificité (%)	97,8	91,1	0,109
	Exactitude (%)	95,5	92,7	0.022
	AUC	0,933	0,956	-

Tableau 27. Tableau comparatif des performances diagnostique des 2 modèles MAIA.IA et GPT 5.2.

➤ Commentaires :

1. Épanchement pleural

Les deux modèles présentent des performances comparables dans la détection de l'épanchement pleural. GPT-5.2 affiche une sensibilité de 77,1 % contre 88,6 % pour MAIA AI, suggérant une meilleure capacité de détection des vrais positifs pour ce dernier. En revanche, la spécificité est similaire (96,0 % vs 94,7 %). L'exactitude est légèrement supérieure pour MAIA AI (92,7 % vs 90,0 %), tout comme l'AUC (0,916 vs 0,881). Le test de McNemar ne montre pas de différence statistiquement significative entre les deux modèles pour cette pathologie ($p = 0,344$ et $p = 1$), indiquant des performances globalement équivalentes.

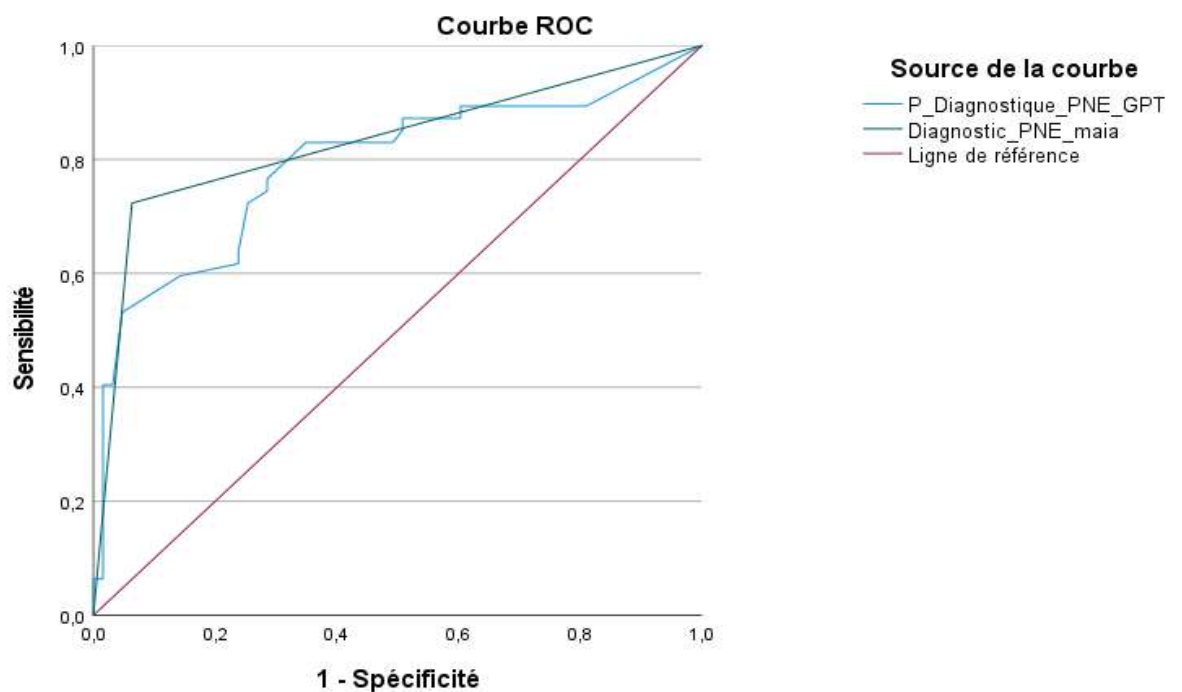


Les segments diagonaux sont générés par les liens.

Figure 22. Courbes ROC comparatives (GPT-5.2 vs MAIA AI), Epanchement pleurale

2. Pneumonie

C'est pour la pneumonie que les différences entre les deux modèles sont les plus marquées et statistiquement significatives. GPT-5.2 présente une sensibilité supérieure (83,0 % vs 72,3 %), mais une spécificité nettement inférieure (68,3 % vs 93,7 %), ce qui traduit un taux de faux positifs plus élevé. L'exactitude est significativement meilleure pour MAIA AI (84,5 % vs 74,5 %, $p < 0,001$), tout comme l'AUC (0,830 vs 0,791). Ces résultats suggèrent que MAIA AI offre un meilleur équilibre sensibilité/spécificité pour la détection de la pneumonie, avec une différence statistiquement significative confirmée par le test de McNemar ($p < 0,001$).



Les segments diagonaux sont générés par les liens.

Figure 23. Courbes ROC comparatives (GPT-5.2 vs MAIA AI), Pneumonie.

3. Pneumothorax

C'est l'entité la plus discriminante. La sensibilité de GPT-5.2 est alarmante : 25,0 % seulement, signifiant que 3 pneumothorax sur 4 ne sont pas détectés. MAIA AI fait significativement mieux (64,3 %, $p < 0,001$) avec une spécificité quasi-parfaite (98,8 %). L'exactitude de GPT-5.2 (80,9 %) est trompeuse : elle est artificiellement gonflée par les vrais négatifs nombreux dans un échantillon déséquilibré. Ce résultat contre-indique formellement l'utilisation isolée de GPT-5.2 pour cette urgence potentiellement létale.

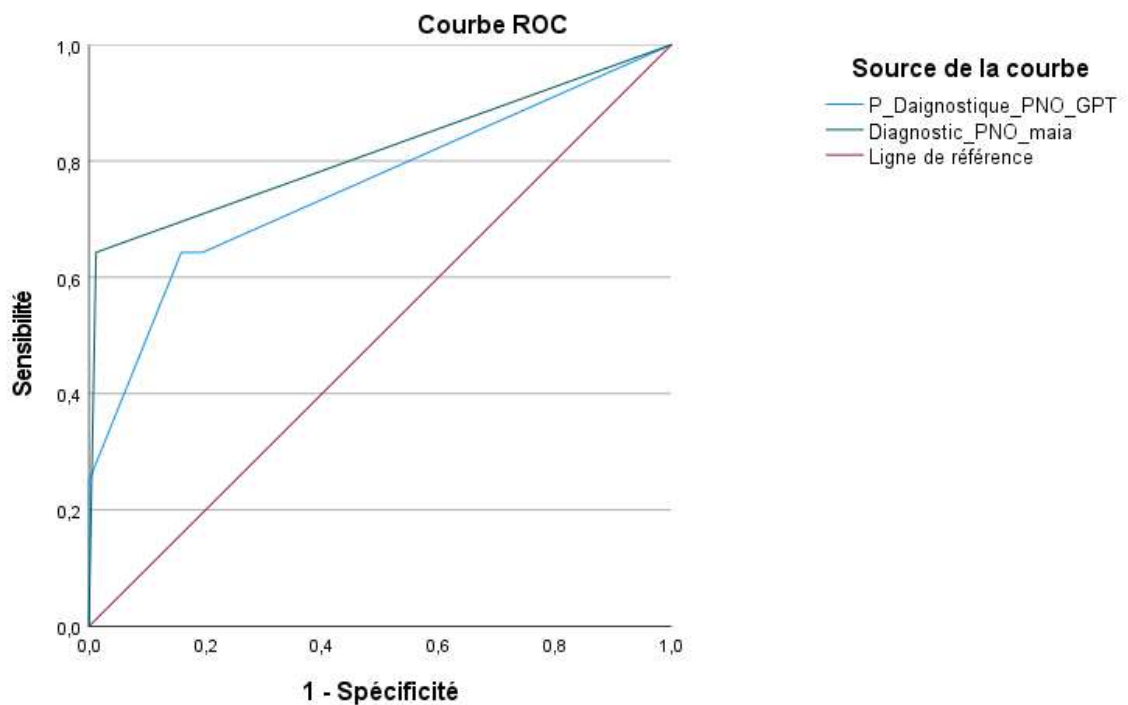
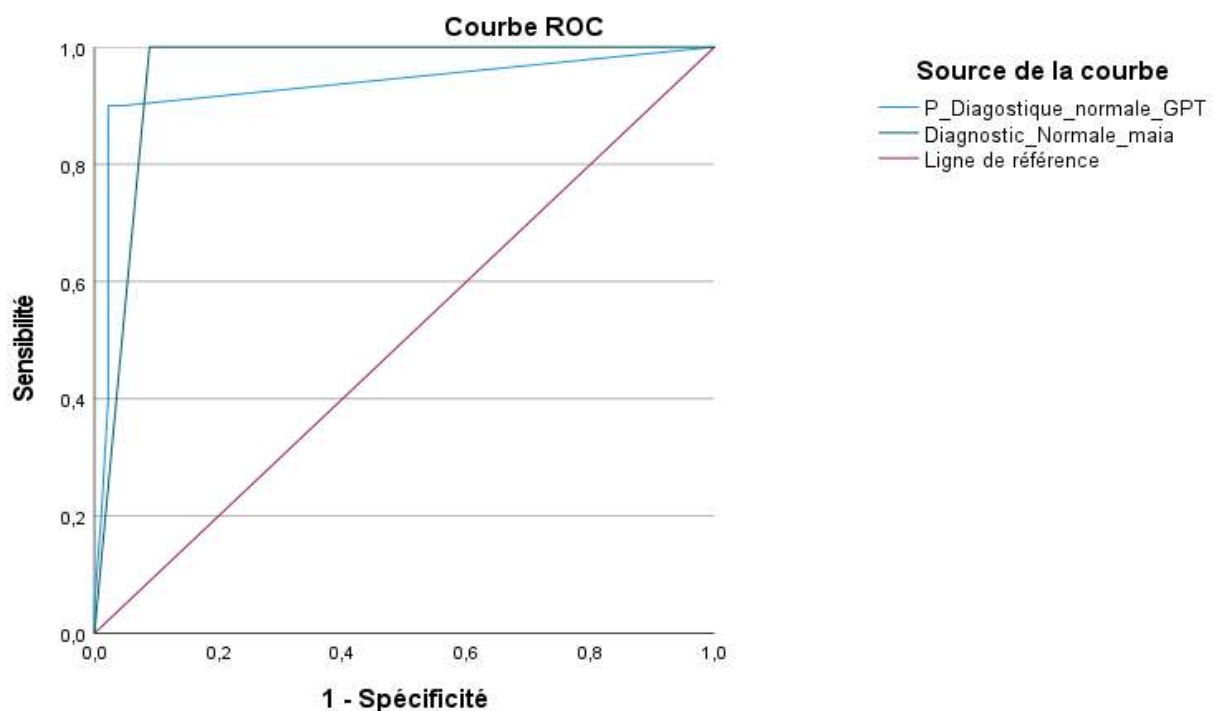


Figure 24. Courbes ROC comparatives (GPT-5.2 vs MAIA AI), Pneumothorax.

4. Aspect radiologique normal

Pour les téléthorax normaux, GPT-5.2 et MAIA AI présentent des sensibilités identiques (85,0 % vs 100,0 %). La spécificité est légèrement meilleure pour GPT-5.2 (97,8 % vs 91,1 %), mais l'exactitude reste comparable (95,5 % vs 92,7 %, $p = 0,022$). Les deux modèles présentent une excellente capacité discriminante pour les clichés normaux, avec une AUC de 0,956 pour MAIA AI et de 0,933 pour GPT-5.2. Pour cette entité diagnostique, MAIA AI fait mieux en termes de sensibilité, avec une différence statistiquement significative confirmée par le test de McNemar ($p < 0,001$), tandis que GPT-5.2 présente une meilleure spécificité et une meilleure exactitude, cette dernière différence étant également statistiquement significative ($p = 0,022$).

Figure 25. Courbes ROC comparatives (GPT-5.2 vs MAIA AI), Radiographie thoracique



Les segments diagonaux sont générés par les liens.

normale.

5. Synthèse générale

De manière globale, MAIA AI a démontré des performances diagnostiques supérieures pour la pneumonie et le pneumothorax, avec des différences statistiquement significatives confirmées par le test de McNemar. GPT-5.2 présente des performances comparables pour l'épanchement pleural et l'aspect normal, mais souffre d'une faiblesse notable dans la détection du pneumothorax, avec une sensibilité insuffisante pour une utilisation clinique fiable. Ces résultats suggèrent que MAIA AI constitue un outil d'aide diagnostique plus performant et plus équilibré dans l'interprétation du téléthorax, bien qu'aucun des deux modèles ne puisse à ce stade se substituer à l'expertise du pneumologue ou du radiologue.

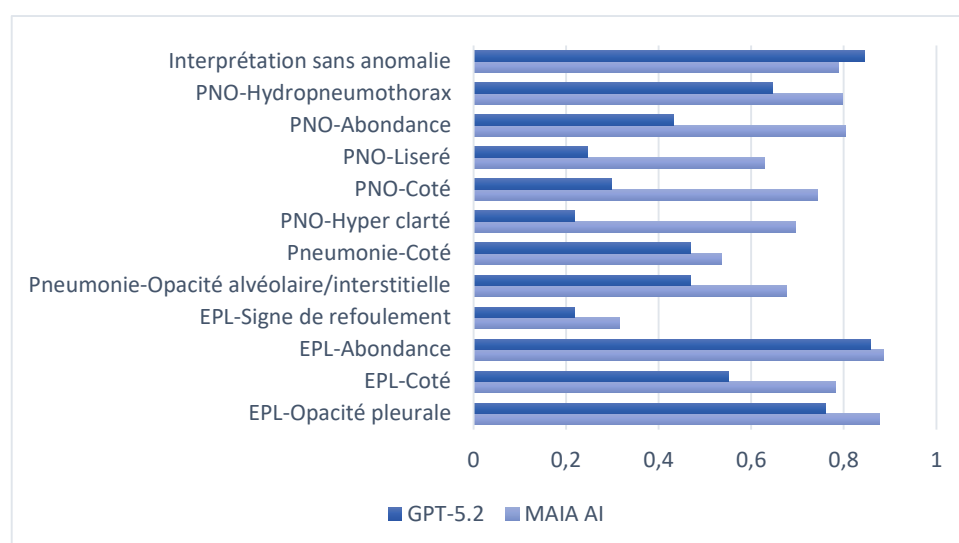
- **COMPARAISON DE CONCORDANCE DIAGNOSTIQUE**

Entité diagnostique	GPT-5.2 : Kappa (IC 95 %)	MAIA AI : Kappa (IC 95 %)
Épanchement pleural	0,760	0,832
Pneumonie	0,496	0,676
Pneumothorax	0,332	0,705
Aspect normal	0,844	0,788

Tableau 28. Coefficient Kappa de Cohen par entité diagnostique pour chaque modèle d'IA.

Le coefficient Kappa de Cohen a montré une concordance variable entre les modèles d'IA et la référence diagnostique selon l'entité étudiée. Pour l'épanchement pleural, les deux modèles ont présenté une bonne concordance, avec une performance supérieure de MAIA AI ($\kappa = 0,832$) par rapport à GPT-5.2 ($\kappa = 0,760$). Concernant la pneumonie, MAIA AI a également obtenu une concordance plus élevée ($\kappa = 0,676$) que GPT-5.2 ($\kappa = 0,496$), traduisant une meilleure reproductibilité diagnostique. La différence la plus marquée a été observée pour le pneumothorax, où MAIA AI a montré une concordance substantielle ($\kappa = 0,705$), alors que GPT-5.2 présentait une concordance faible à modérée ($\kappa = 0,332$). En revanche, pour l'aspect normal, GPT-5.2 a obtenu la meilleure concordance ($\kappa = 0,844$) comparativement à MAIA AI ($\kappa = 0,788$). Globalement, MAIA AI semble offrir une concordance diagnostique plus homogène et plus élevée pour les principales pathologies thoraciques, tandis que GPT-5.2 se distingue davantage dans l'identification des radiographies normales.

• COMPARAISON DE CONCORDANCE SEMIOLOGIQUE



*Figure 26. Le coefficient Kappa de Cohen par signe radiologique
GPT-5.2 vs MAIA AI*

Dans l'ensemble, l'analyse comparative des coefficients Kappa met en évidence une meilleure concordance diagnostique de MAIA AI par rapport à GPT-5.2 pour la majorité des signes radiologiques étudiés. MAIA AI obtenait des accords plus élevés et plus constants dans l'évaluation des anomalies pleurales, des opacités pulmonaires et surtout des signes de pneumothorax, traduisant une meilleure capacité de détection et de caractérisation des lésions thoraciques.

À l'inverse, GPT-5.2 présentait des performances plus variables, avec plusieurs critères classés en accord passable, notamment pour les signes du pneumothorax et le signe de refoulement. Toutefois, GPT-5.2 conservait une excellente concordance pour l'identification des radiographies normales ainsi que pour certains critères quantitatifs comme l'abondance de l'épanchement pleural. Ces résultats suggèrent que MAIA AI offre une performance diagnostique plus stable et plus fiable dans l'analyse globale des radiographies thoraciques standard.

7. ANALYSE DES ERREURS DIAGNOSTIQUES :

Pathologie	GPT-5.2 FP	GPT-5.2 FN	MAIA AI FP	MAIA AI FN
Épanchement pleural	3	8	4	4
Pneumonie	20	8	4	13
Pneumothorax	0	21	1	10
Aspect normal	2	3	8	0
TOTAL	25	40	17	27

Tableau 29. Répartition des faux positifs (FP) et faux négatifs (FN) par entité et par modèle.

Le modèle GPT-5.2 a produit un total de 65 erreurs diagnostiques (25 FP + 40 FN) contre 44 erreurs pour MAIA AI (17 FP + 27 FN). Le pneumothorax représentait la principale source de faux négatifs pour GPT-5.2 (21 FN, soit 75 % des cas manqués) en cohérence avec la faible concordance sémiologique observée pour cette pathologie (Kappa hyperclarté = 0,217 ; Kappa liseré = 0,246), tandis que la pneumonie constituait la principale source d'erreur pour MAIA AI (13 FN).

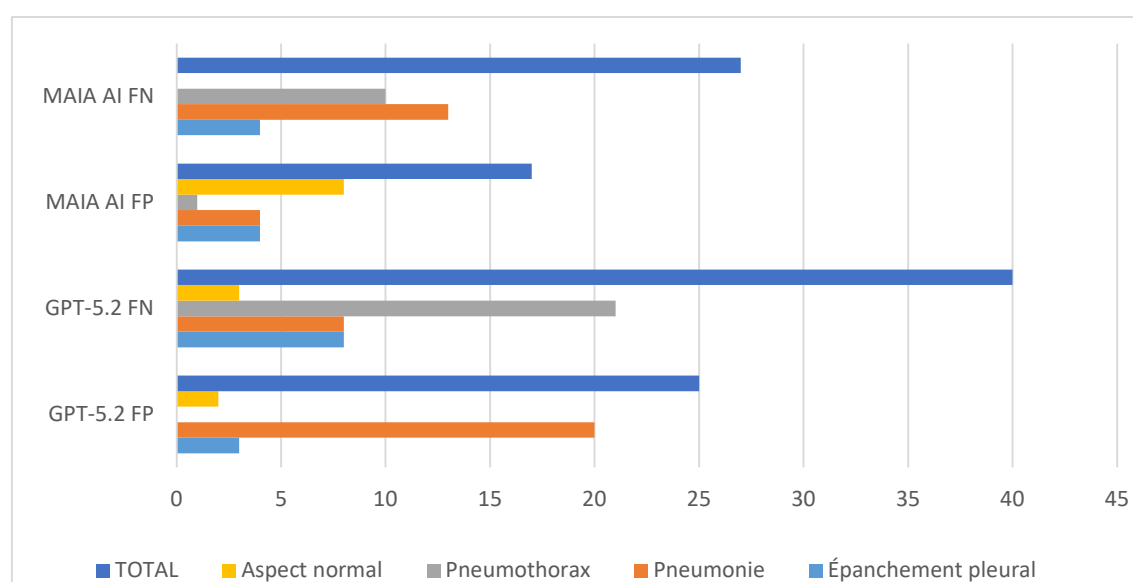


Figure 27. Répartition des erreurs diagnostiques par entité pathologique et par modèle d'IA

III. DISCUSSION

L'intégration de l'intelligence artificielle dans l'interprétation radiologique constitue une avancée prometteuse, mais exige une évaluation méthodologique rigoureuse. Cette étude met en regard deux approches technologiques distinctes : un modèle de langage multimodal généraliste (GPT-5.2) et un système d'IA dédié à l'imagerie médicale (MAIA AI), révélant des disparités notables entre les deux.

➤ Épanchement pleural

Les deux modèles ont démontré des performances satisfaisantes, avec des AUC respectives de 0,881 pour GPT-5.2 et 0,916 pour MAIA AI. La sensibilité supérieure de MAIA AI (88,6 % contre 77,1 %) se traduit par un nombre de faux négatifs plus faible (4 contre 8), ce qui revêt une importance clinique particulière en contexte d'urgence. Toutefois, cette différence n'atteint pas le seuil de significativité statistique ($p = 0,344$, test de McNemar), suggérant une efficacité globalement comparable pour cette pathologie.

L'analyse de la concordance sémiologique révèle néanmoins des différences qualitatives importantes : MAIA AI présente un meilleur accord pour l'opacité pleurale ($\kappa = 0,876$ contre 0,760), la latéralisation ($\kappa = 0,782$ contre 0,551) et l'estimation de l'abondance ($\kappa = 0,885$ contre 0,858). Le signe de refoulement médiastinal demeure difficile à identifier pour les deux modèles ($\kappa < 0,32$), ce qui est cliniquement préoccupant.

➤ Pneumonie

La pneumonie illustre le compromis classique entre sensibilité et spécificité. GPT-5.2 affiche une sensibilité plus élevée (83,0 % contre 72,3 %), mais génère 20 faux positifs (spécificité 68,3 %), contre seulement 4 pour MAIA AI (spécificité 93,7 %, $p < 0,001$). Ce nombre élevé de faux positifs est cohérent avec un biais de surgénéralisation documenté pour les modèles généralistes, qui tendent à évoquer une pneumonie devant tout infiltrat ou condensation, quelle qu'en soit l'étiologie. La concordance sémiologique confirme ce mécanisme : GPT-5.2 décrivait une opacité dans 59 cas, contre 47 pour le gold standard ($\kappa = 0,469$).

Sur le plan clinique, GPT-5.2 apparaît plus adapté au dépistage en raison de sa haute sensibilité et de son faible taux de faux négatifs, tandis que MAIA AI convient davantage à la confirmation diagnostique, limitant le risque de prescription antibiotique injustifiée (VPP = 89,5 %).

➤ **Pneumothorax — écart majeur et implications cliniques critiques**

Le pneumothorax est la pathologie pour laquelle l'écart de performance est le plus marqué et le plus lourd de conséquences cliniques. GPT-5.2 présente une sensibilité de seulement 25,0 % (IC 95 % : 12,7–43,4 %), avec 21 faux négatifs sur 28 cas confirmés. MAIA AI affiche des performances nettement supérieures : sensibilité 64,3 %, spécificité 98,8 %, AUC 0,815 ($p < 0,001$). La concordance sémiologique de GPT-5.2 est particulièrement déficiente (κ hyperclarté = 0,217 ; κ liséré = 0,246), tandis que MAIA AI obtient des accords bons à très bons (κ = 0,628 à 0,803). La sensibilité de 25,0 % constitue une contre-indication formelle à l'utilisation isolée de GPT-5.2 pour cette urgence potentiellement létale.

➤ **Aspect radiologique normal**

La reconnaissance de la normalité représente le domaine de meilleure performance de GPT-5.2 (AUC = 0,933, exactitude 95,5 %), tandis que MAIA AI y atteint sa sensibilité maximale (100 %, sans aucun faux négatif). GPT-5.2 conserve un avantage en spécificité (97,8 % contre 91,1 %) et en exactitude globale ($p = 0,022$). Cette supériorité de GPT-5.2 pour la reconnaissance de la normalité peut s'expliquer par son entraînement sur des corpus massifs et diversifiés incluant une grande proportion de radiographies normales issues de populations variées, lui conférant une représentation interne plus robuste de l'anatomie thoracique standard. À l'inverse, MAIA AI, bien qu'atteignant une sensibilité parfaite (100 %), présente une spécificité plus basse (91,1 %), témoignant d'une tendance à surdiagnostiquer des anomalies sur des clichés normaux.

➤ **Comparaison avec la littérature**

Les performances de MAIA AI (AUC : 0,815–0,956) s'inscrivent dans la fourchette haute des données publiées concernant les modèles spécialisés en imagerie thoracique. Rajpurkar et al. (2017) ont développé CheXNet, un modèle de deep learning spécialisé dans la détection de pathologies thoraciques sur radiographie standard, démontrant des performances supérieures ou comparables à celles de radiologues experts pour plusieurs entités pathologiques ; les AUC obtenues par MAIA AI dans notre étude s'inscrivent dans une fourchette similaire, suggérant que les modèles spécialisés maintiennent des niveaux de performance élevés même dans des contextes cliniques différents de leurs données d'entraînement. De même, Irvin et al. (2019), à travers la base CheXpert regroupant plus de 220 000 radiographies thoraciques, ont confirmé la robustesse des approches spécialisées pour plusieurs entités pathologiques dont l'épanchement pleural et la pneumonie. Concernant

GPT-5.2, les AUC obtenues (0,755–0,933) sont globalement concordantes avec celles décrites pour des modèles multimodaux généralistes de génération comparable, tels que GPT-4V évalué dans des tâches d'imagerie médicale (Huh et al., 2023 ; Kim et al., 2024). Néanmoins, sa sensibilité particulièrement faible pour le pneumothorax (25,0 %) confirme que la reconnaissance de patterns visuels fins et spécifiques représente la limite structurelle inhérente aux grands modèles de langage multimodaux généralistes appliqués à l'imagerie médicale, rejoignant ainsi les conclusions de Singhal et al. (2023), selon lesquelles les modèles généralistes demeurent inférieurs aux modèles spécialisés dès lors que la perception visuelle fine constitue un prérequis diagnostique indispensable.

➤ **Les points forts**

- L'utilisation des standards **STARD-AI** garantit la reproductibilité de notre évaluation.
- L'inclusion d'une analyse de concordance sémiologique (Objectif secondaire) dépasse la simple mesure de précision pour toucher à la qualité du raisonnement médical.

➤ **Limites principales**

- Étude mono-centrique (Hôpital Mixte de Laghouat) sur n = 110 radiographies ; intervalles de confiance larges pour le pneumothorax (n = 28), limitant la précision des estimations.
- Exclusion de la tuberculose pulmonaire faute d'effectif suffisant, limitant la représentativité de l'étude dans le contexte épidémiologique algérien.
- Gold standard unique (Un pneumologue) ; En effet, bien que le pneumologue senior retenu présente une expérience reconnue, le recours à un lecteur unique expose à un biais d'interprétation individuelle et ne permet pas d'estimer la variabilité inter-observateurs. L'idéal aurait été de constituer un comité de lecture composé d'au moins deux pneumologues ou radiologues.
- Export JPEG : la compression d'image est susceptible de dégrader la détection de quelques signes tels que le liséré de décollement pleural ou le petit pneumothorax (fichiers DICOM non disponibles).

➤ **Implications cliniques**

Sur la base des résultats de cette étude, MAIA AI présente un profil de performance qui pourrait justifier un rôle d'aide au diagnostic pour le médecin (urgentiste, médecin généraliste, interne) en l'absence d'accès immédiat à un radiologue. Sa très haute spécificité pour le pneumothorax (98,8 %, VPP = 94,7 %) en fait un outil de confirmation fiable ; en revanche, sa sensibilité de 64,3 % interdit tout usage en triage automatisé sans vérification humaine systématique.

GPT-5.2 ne peut être recommandé pour aucun usage clinique isolé en urgence respiratoire. Sa sensibilité de 25,0 % pour le pneumothorax, avec 21 cas sur 28 non détectés, représente un risque patient inacceptable. Il pourrait être envisagé exclusivement pour des radiographies programmées normales ou comme outil d'aide à la rédaction du compte rendu, sous supervision médicale stricte.

➤ **Recommandations pratiques**

- Aucun des deux modèles ne doit remplacer l'expertise radiologique humaine dans l'état actuel de leurs performances.
- Tout résultat d'IA doit être considéré comme une aide à la décision et validé médicalement avant toute action thérapeutique.
- Une formation des utilisateurs aux limites spécifiques de chaque modèle par pathologie est indispensable avant toute implémentation.
- Un essai prospectif contrôlé évaluant l'impact de MAIA AI sur la performance diagnostique des médecins non spécialisés est recommandé, conduit selon les standards STARD-AI.
- Des études multicentriques incluant la tuberculose pulmonaire et des échantillons stratifiés par sexe et âge sont nécessaires pour établir la généralisabilité de ces résultats.

IV CONCLUSION :

L'intelligence artificielle appliquée à l'imagerie médicale connaît un essor considérable, portée par des avancées technologiques rapides et une demande croissante d'outils d'aide au diagnostic. Pour autant, toute intégration clinique de ces outils demeure conditionnée à une évaluation rigoureuse, indépendante et méthodologiquement fondée.

Ce travail démontre que les performances diagnostiques des systèmes d'intelligence artificielle varient considérablement selon la nature de la pathologie considérée et l'architecture du modèle utilisé. La distinction entre approches généralistes et modèles spécialisés revêt ainsi une importance capitale, et le choix d'un outil doit impérativement être guidé par son profil de performance spécifique à chaque entité pathologique, et non par ses seules capacités globales. En contexte d'urgence respiratoire, où le retard ou l'erreur diagnostique peut engager le pronostic vital, cette exigence de précision prend une dimension particulièrement critique.

Ces résultats confirment que si les modèles généralistes peuvent constituer des outils d'aide utiles pour certaines tâches de raisonnement médical, ils restent insuffisants dès lors que la détection de patterns visuels fins est requise. À l'inverse, les modèles spécialisés comme MAIA AI présentent un profil de performance susceptible de soutenir le médecin dans sa démarche diagnostique, sans toutefois pouvoir se substituer à l'expertise du radiologue.

Au-delà de la détection primaire, l'intelligence artificielle peut également jouer un rôle de seconde lecture, renforçant et confirmant certaines hypothèses diagnostiques émises par le clinicien, et réduisant ainsi le risque d'omission. À mesure que ces systèmes continuent d'évoluer, il est permis d'envisager qu'ils atteignent une résolution analytique suffisante pour identifier des lésions de petite taille ou de faible contraste, potentiellement en deçà du seuil de détection visuelle humaine, ouvrant ainsi de nouvelles perspectives dans le dépistage précoce des pathologies thoraciques.

En définitive, l'intelligence artificielle en imagerie thoracique ne doit pas être envisagée comme un substitut au praticien, mais comme un outil d'aide à la décision, dont l'intégration clinique exige une validation approfondie, une formation adaptée des utilisateurs et une supervision médicale constante. Des études multicentriques prospectives, intégrant des populations plus larges et des pathologies plus diversifiées, demeurent indispensables pour confirmer la généralisabilité de ces résultats et encadrer une implémentation responsable de ces technologies dans la pratique radiologique.

V. BIBLIOGRAPHIE :

1. HAIDAR AHMAD H, HAMID SALIH MA, LUANGRATH A, SEDE T, ZUGAJ L. Evaluation des Applications d'intelligence artificielle au service de l'imagerie médicale [PDF,PowerPoint]. Universite de Technologie de Compiegne; 2022 [cité 4 mai 2026]. Disponible sur: <https://travaux.master.utc.fr/formations-master/ingenierie-de-la-sante/ids122> doi:10.34746/KBKC-CD39
2. Waymel Q, Badr S, Demondion X, Cotten A, Jacques T. Impact of the rise of artificial intelligence in radiology: What do radiologists think? *Diagn Interv Imaging*. juin 2019;100(6):327-36. doi:10.1016/j.diii.2019.03.015
3. Laredo JD, Morvan G, Adamsbaum C, Argyropoulou M, Badoual C, Hauet T, et al. Rapport 26-01. Apport de l'intelligence artificielle en imagerie médicale. *Bull Académie Natl Médecine*. mai 2026;210(4):367-75. doi:10.1016/j.banm.2026.03.004
4. Bettinger H, Lenczner G, Guigui J, Rotenberg L, Zerbib E, Attia A, et al. Evaluation of the Performance of an Artificial Intelligence (AI) Algorithm in Detecting Thoracic Pathologies on Chest Radiographs. *Diagnostics*. 4 juin 2024;14(11):1183. doi:10.3390/diagnostics14111183
5. McCarthy J. WHAT IS ARTIFICIAL INTELLIGENCE?
6. intelligence artificielle - LAROUSSE [Internet]. [cité 3 mai 2026]. Disponible sur: https://www.larousse.fr/encycopedie/divers/intelligence_artificielle/187257
7. Marvin L. Minsky - Semantic Information Processing-The MIT Press (1968) | PDF [Internet]. [cité 3 mai 2026]. Disponible sur: <https://www.scribd.com/document/638038757/Marvin-L-Minsky-Semantic-Information-Processing-The-MIT-Press-1968>
8. Haenlein M, Kaplan A. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *Calif Manage Rev*. août 2019;61(4):5-14. doi:10.1177/0008125619864925
9. Cardon D, Cointet JP, Mazières A, Carey-Libbrecht L. Neurons spike back: The invention of inductive machines and the artificial intelligence controversy. *Réseaux*. 2 nov 2018;No 211(5):173-220. doi:10.3917/res.211.0173
10. Mayor A. Gods and robots: myths, machines, and ancient dreams of technology. Princeton: Princeton university press; 2018.
11. Russell SJ, Norvig P. Artificial intelligence: a modern approach. Fourth edition, global edition. Boston: Pearson; 2022. 1 p. (Prentice Hall series in artificial intelligence).
12. McCulloch WS, Pitts W. A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY.
13. Samuel AL. Some Studies in Machine Learning Using the Game of Checkers. 1959.

14. Bhavsar P, Safro I, Bouaynaya N, Polikar R, Dera D. Machine Learning in Transportation Data Analytics. In: Data Analytics for Intelligent Transportation Systems [Internet]. Elsevier; 2017 [cité 2 mai 2026]. p. 283-307. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/B9780128097151000122> doi:10.1016/B978-0-12-809715-1.00012-2
15. Mohri M, Rostamizadeh A, Talwalkar A. Foundations of machine learning. Second edition. Cambridge, Massachusetts: The MIT Press; 2018. 486 p. (Adaptive computation and machine learning).
16. Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016. 801 p.
17. Azizi S, Mustafa B, Ryan F, Beaver Z, Freyberg J, Deaton J, et al. Big Self-Supervised Models Advance Medical Image Classification. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV) [Internet]. Montreal, QC, Canada: IEEE; 2021 [cité 2 mai 2026]. p. 3458-68. Disponible sur: <https://ieeexplore.ieee.org/document/9710396/> doi:10.1109/ICCV48922.2021.00346
18. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A Survey on Deep Learning in Medical Image Analysis. Med Image Anal. déc 2017;42:60-88. doi:10.1016/j.media.2017.07.005
19. Amazon S the authors FCMPMW. Deep Learning with Python, Third Edition [Internet]. [cité 3 mai 2026]. Disponible sur: <https://deeplearningwithpython.io/>
20. Macukow B. Neural Networks – State of Art, Brief History, Basic Models and Architecture. In: Saeed K, Homenda W, éditeurs. Computer Information Systems and Industrial Management [Internet]. Cham: Springer International Publishing; 2016 [cité 28 avr 2026]. p. 3-14. (Lecture Notes in Computer Science). Disponible sur: https://link.springer.com/10.1007/978-3-319-45378-1_1 doi:10.1007/978-3-319-45378-1_1
21. Wira P. Réseaux de neurones artificiels : architectures et applications [Internet]. Université de Haute-Alsace; 2009. Disponible sur: <https://www.scribd.com/document/546022635/Reseaux-de-Neurones-Artificiels>
22. Haykin SS, Haykin SS. Neural networks and learning machines. 3rd ed. New York: Prentice Hall; 2009. 906 p.
23. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging. août 2018;9(4):611-29. doi:10.1007/s13244-018-0639-9
24. Mohajer-Bastami A, Moin S, Ahmad S, Ahmed AR, Pouwels S, Hajibandeh S, et al. Artificial intelligence in healthcare: applications, challenges, and future directions. A narrative review informed by international, multidisciplinary expertise. Front Digit Health. 6 nov 2025;7:1644041. doi:10.3389/fdgth.2025.1644041
25. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. janv 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7

26. Fahad N, Rabbi RI, Benta Hasan S, Sultana Prity F, Ahmed R, Ahmed F, et al. Generative AI in clinical (2020–2025): a mini-review of applications, emerging trends, and clinical challenges. *Front Digit Health*. 3 nov 2025;7:1653369. doi:10.3389/fdgth.2025.1653369
27. Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ*. 22 sept 2023;23(1):689. doi:10.1186/s12909-023-04698-z
28. Briganti G, Le Moine O. Artificial Intelligence in Medicine: Today and Tomorrow. *Front Med*. 5 févr 2020;7:27. doi:10.3389/fmed.2020.00027
29. Kaul V. history of artificiel intelligence. 2020.
30. Vela V, Sonay AY, Limani P, Graf L, Sabani B, Gjermani D, et al. Toward Artificial Intelligence in Oncology and Cardiology: A Narrative Review of Systems, Challenges, and Opportunities. *J Clin Med*. 24 oct 2025;14(21):7555. doi:10.3390/jcm14217555
31. Trasca DM, Dorin PI, Carmen S, Varut RM, Singer CE, Radivojevic K, et al. Artificial Intelligence in Biomedicine: A Systematic Review from Nanomedicine to Neurology and Hepatology. *Pharmaceutics*. déc 2025;17(12):1564. doi:10.3390/pharmaceutics17121564
32. Mentzou A, Rogers A, Carvalho E, Daly A, Malone M, Kerasidou X. Artificial Intelligence in Digital Self-Diagnosis Tools: A Narrative Overview of Reviews. *Mayo Clin Proc Digit Health*. sept 2025;3(3):100242. doi:10.1016/j.mcpdig.2025.100242
33. Ogut E. Artificial Intelligence in Clinical Medicine: Challenges Across Diagnostic Imaging, Clinical Decision Support, Surgery, Pathology, and Drug Discovery. *Clin Pract*. 16 sept 2025;15(9):169. doi:10.3390/clinpract15090169
34. Wang TW, Hong JS, Chiu HY, Chao HS, Chen YM, Wu YT. Standalone deep learning versus experts for diagnosis lung cancer on chest computed tomography: a systematic review. *Eur Radiol*. 22 mai 2024;34(11):7397-407. doi:10.1007/s00330-024-10804-6
35. ohana m, ferritti g. Les fondamentaux – Chapitre 25 – Imagerie thoracique. In. Disponible sur: <https://www.cnp-mn.fr/les-fondamentaux-chapitre-25-imagerie-thoracique/>
36. Stephen m, Christopher. The WHO manual of diagnostic imaging. 2006.
37. Hodler J, Kubik-Huch RA, Von Schulthess GK, éditeurs. Diseases of the Chest, Breast, Heart and Vessels 2019-2022: Diagnostic and Interventional Imaging [Internet]. Cham: Springer International Publishing; 2019 [cité 6 mai 2026]. (IDKD Springer Series). Disponible sur: <http://link.springer.com/10.1007/978-3-030-11149-6> doi:10.1007/978-3-030-11149-6
38. Klein JS, Rosado-de-Christenson ML. A Systematic Approach to Chest Radiographic Analysis. In: Hodler J, Kubik-Huch RA, Von Schulthess GK, éditeurs. Diseases of the Chest, Breast, Heart and Vessels 2019-2022 [Internet]. Cham: Springer International Publishing; 2019 [cité 6 mai 2026]. p. 1-16. (IDKD Springer Series). Disponible sur:

http://link.springer.com/10.1007/978-3-030-11149-6_1 doi:10.1007/978-3-030-11149-6_1

39. Savoir quand demander une radiographie thoracique en urgence | Société d’Imagerie Thoracique [Internet]. [cité 6 mai 2026]. Disponible sur: <https://www.sit-radiologie.fr/juniors-0/2eme-cycle-fiche-lisa/savoir-quand-demander-une-radiographie-thoracique-en-urgence>
40. Pham HH, Nguyen HQ, Nguyen HT, Le LT, Khanh L. An Accurate and Explainable Deep Learning System Improves Interobserver Agreement in the Interpretation of Chest Radiograph [Internet]. arXiv; 2022 [cité 6 mai 2026]. Disponible sur: <http://arxiv.org/abs/2208.03545> doi:10.48550/arXiv.2208.03545
41. Adam Essa ME. Diagnostic accuracy of AI in chest radiography for pneumonia and lung cancer: A meta-analysis. *Eur J Radiol Open*. déc 2025;15:100701. doi:10.1016/j.ejro.2025.100701
42. Monti CB, Bianchi LMG, Rizzetto F, Carbonaro LA, Vanzulli A. Diagnostic performance of an artificial intelligence model for the detection of pneumothorax at chest X-ray. *Clin Imaging*. janv 2025;117:110355. doi:10.1016/j.clinimag.2024.110355
43. ZAFAR A, NOOR U. X-rays in the AI age: revolutionizing pulmonary tuberculosis diagnosis—a systematic review & meta-analysis [Internet]. Vol. 8. 2025. Disponible sur: <https://jmai.amegroups.org/article/view/9822>
44. Ahn JS, Ebrahimian S, McDermott S, Lee S, Naccarato L, Di Capua JF, et al. Association of Artificial Intelligence–Aided Chest Radiograph Interpretation With Reader Performance and Efficiency. *JAMA Netw Open*. 31 août 2022;5(8):e2229289. doi:10.1001/jamanetworkopen.2022.29289
45. Anderson PG, Tarder-Stoll H, Alpaslan M, Keathley N, Levin DL, Venkatesh S, et al. Deep learning improves physician accuracy in the comprehensive detection of abnormalities on chest X-rays. *Sci Rep*. 24 oct 2024;14(1):25151. doi:10.1038/s41598-024-76608-2
46. Yanagawa M, Tomiyama N. Clinical Performance of Current-Generation AI Tools for Chest Radiographs. *Radiology*. 1 sept 2023;308(3):e232139. doi:10.1148/radiol.232139
47. Kocak B, Klontzas ME, Stanzione A, Meddeb A, Demircioğlu A, Bluethgen C, et al. Evaluation metrics in medical imaging AI: fundamentals, pitfalls, misapplications, and recommendations. *Eur J Radiol Artif Intell*. sept 2025;3:100030. doi:10.1016/j.ejrai.2025.100030

VI. LES ANNEXES :

FICHES DE COLLECT DE DONNEES :

EPL :

	COTE ATTENINT	OPACITE	ABONDANCE	SIGNES DE REFOULEMENT
TLT N				
TLT N				
TLT N				
TLT N				
TLT N				
TLT N				
TLT N				
TLT N				
TLT N				
TLT N				
TLT N				
TLT N				
TLT N				
TLT N				
TLT N				
TLT N				
TLT N				
TLT N				
TLT N				
TLT N				

PNEUMONIE :

	LE COTE ATTEINT	OPACITE ALVEOLAIRE
TLT N		
TLT N		
TLT N		
TLT N		
TLT N		
TLT N		
TLT N		
TLT N		
TLT N		
TLT N		
TLT N		
TLT N		
TLT N		
TLT N		
TLT N		
TLT N		
TLT N		
TLT N		
TLT N		
TLT N		

PNEUMOTHORAX :

[illegible]

TLT NORMALE :

[illegible]

LES MODELES D'IA TESTEES :

MAIA.AI :

<https://maiahub.my/login>



GPT-5.2 :

<https://openai.com/index/introducing-gpt-5-2/>





ECG mobile assisté par IA : Kardia



Les capteurs Wearables



Le modèle d'AI Sugar IQ