

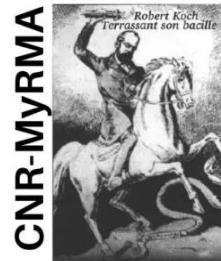
INTELLIGENCE ARTIFICIELLE ET MALADIES INFECTIEUSES PEDIATRIQUES

GPIP : 41ème JOURNEE ANNUELLE - 27 MARS 2026

Applications concrètes en microbiologie

Dr Alexandre Godmer, AHU

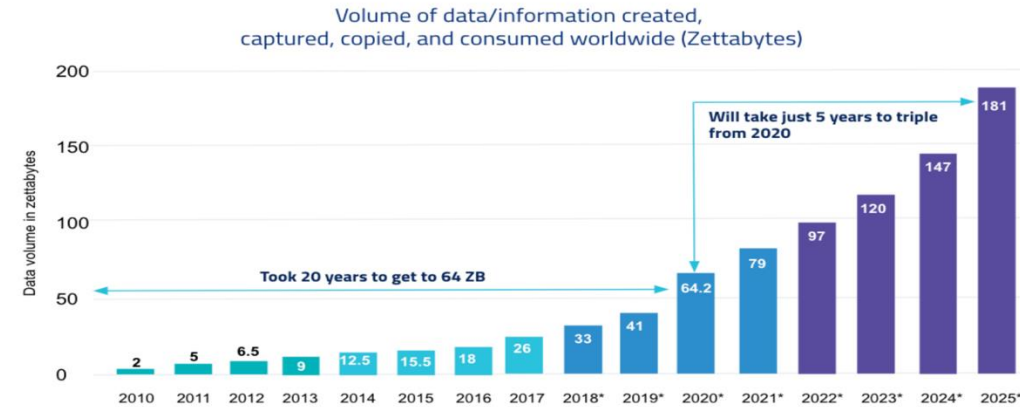
Hôpital Saint-Antoine (AP-HP, Sorbonne-Université)



L'ère du Big data (1)

1

Big Data Grows Ever Bigger

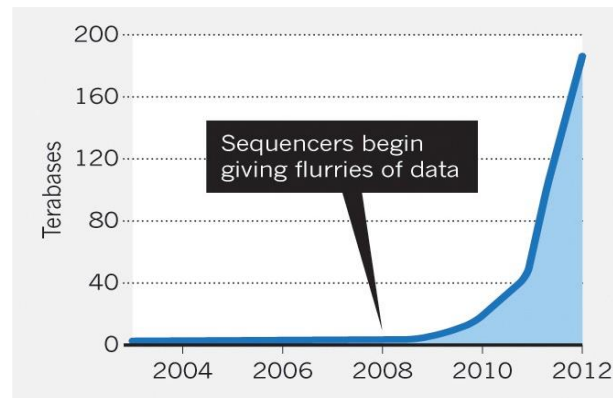


35%
of all data will be life sciences + healthcare by 2025

10 Ways Big Data is Revolutionizing Healthcare - Fusion Chat

2

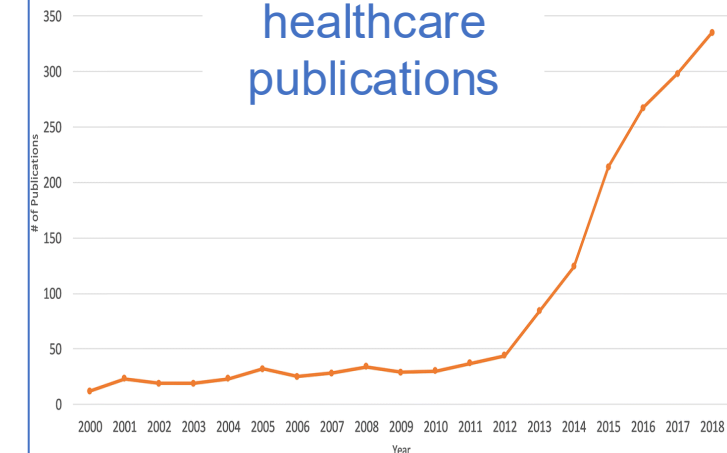
Sequencing data



Max V, *Nature* (2013), DOI: 10.1038/498255a

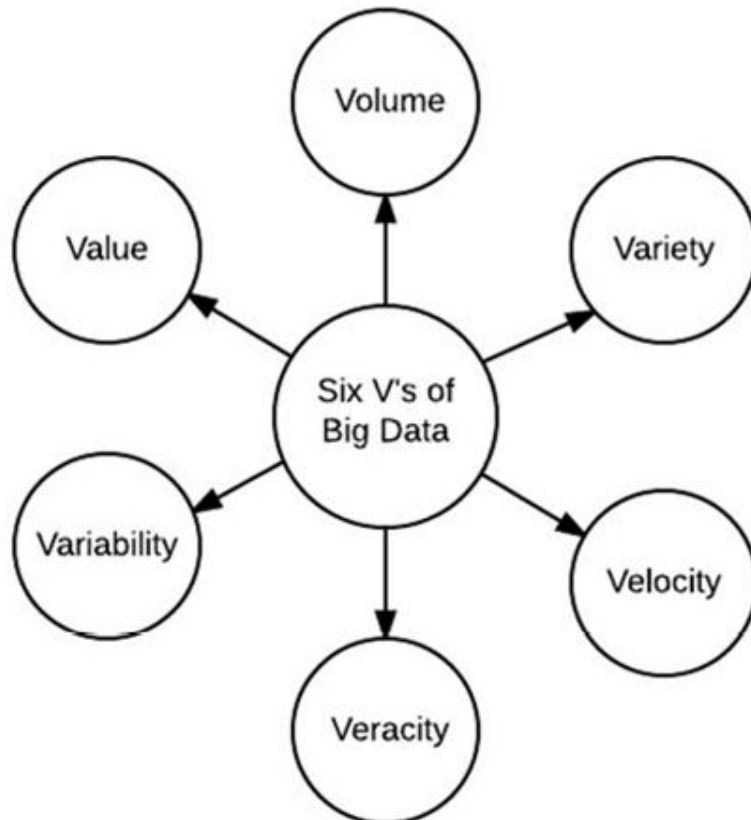
3

Big data healthcare publications



Big data in healthcare: management, analysis and future prospects

Big Data: Données massives et hétérogènes

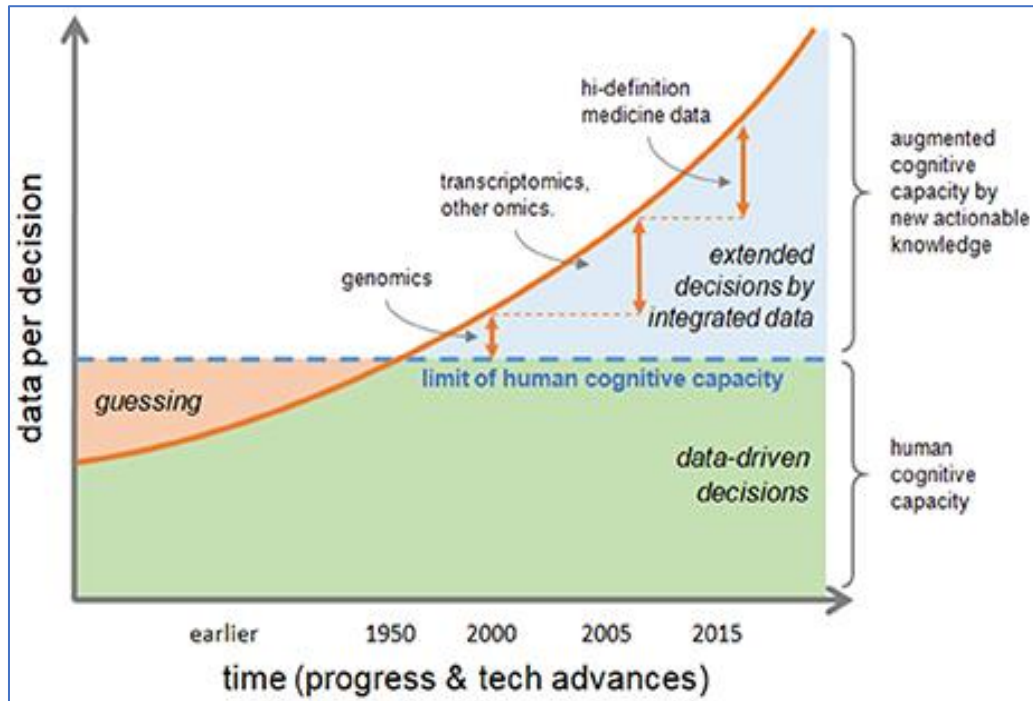


Big Data in Healthcare Management: A Review of Literature

La production de données médicales augmente plus rapidement que notre capacité à les stocker, les traiter et les interpréter

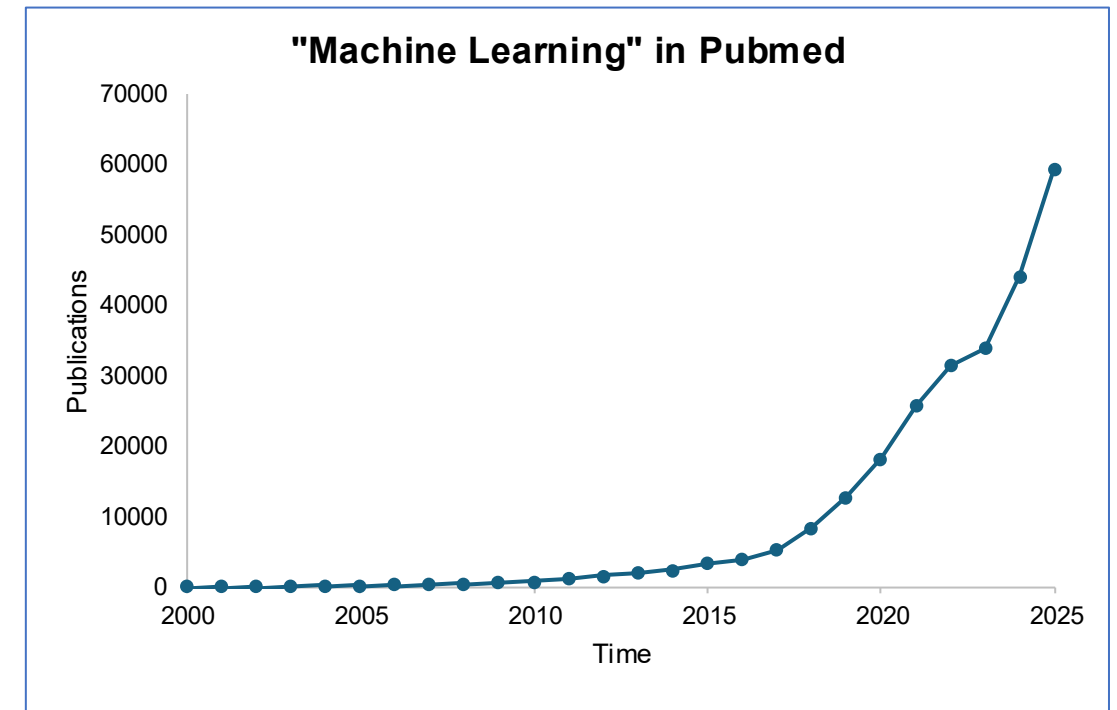
L'ère du Big data (2)

Analyses avancées essentielles pour la recherche translationnelle, clinique et fondamentale



Rossi & Grifantini, Frontiers in Digital Humanities (2018), DOI: 10.3389/fdigh.2018.00013

Démocratisation de l'utilisation des algorithmes d'intelligence artificielle



La recherche moderne passe de l'intuition humaine à la connaissance assistée par des algorithmes, car le volume et la complexité des données dépassent rapidement notre capacité à les interpréter

Les étapes du diagnostic microbiologique classique

J0

≥ 16-24 heures

≥ 48 heures

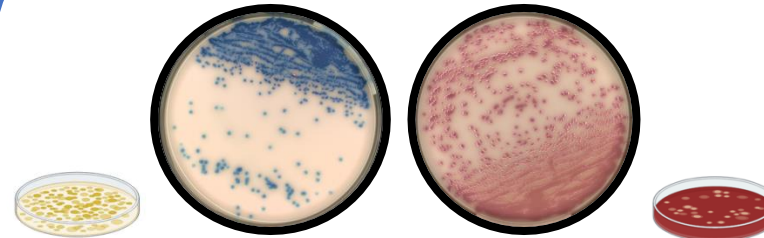
1

Examen microscopique

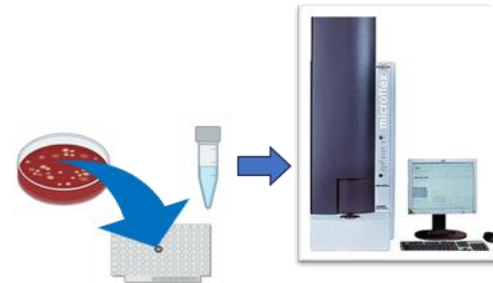


2

Culture



Identification



Spectrométrie de masse de type
MALDI-TOF

3

Détermination de la sensibilité aux antibiotiques

→ antibiogramme



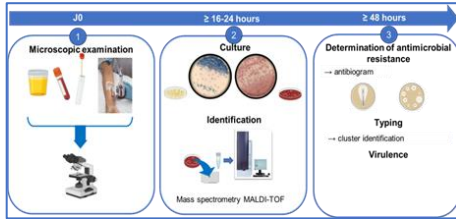
« Typing »

→ Identification de clusters

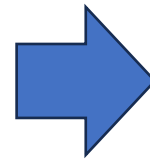
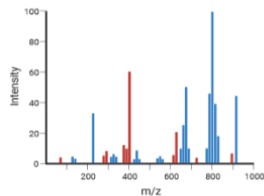
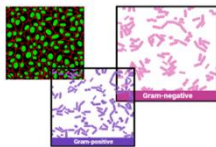
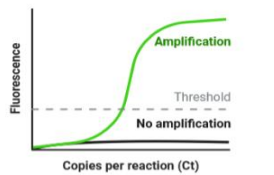
Virulence

Temps total : 48-72 h ⇒ diagnostic classique lent
Données segmentées ⇒ interprétation globale rétrospective

La richesse des données biologiques



Production des données



Méthodes classiques

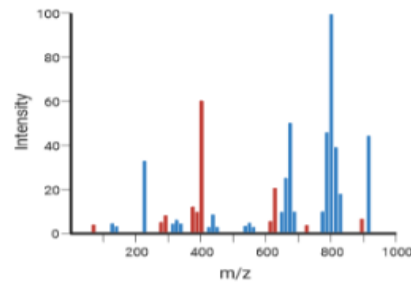
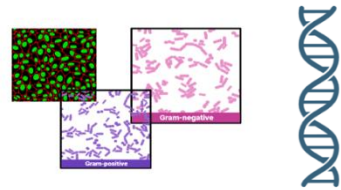
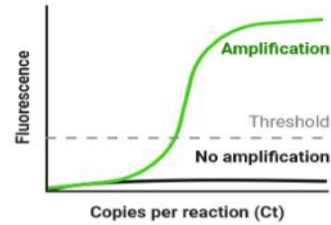
- Données phénotypiques
- Coloration de Gram
- Identification
- Antibiogramme

« Dark raw massive data »

- Séquençage
- Spectres MALDI-TOF
- Métabolomique
- Imagerie
- Données clinico-biologiques

**Production de données massives (omiques, image, etc) à chaque étape
⇒ sous exploitation par les méthodes classiques segmentées**

Les promesses de l'IA en microbiologie



1

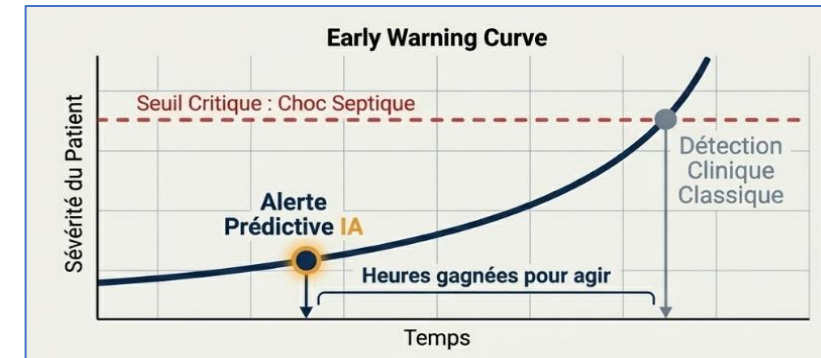
Amélioration de l'exploitation des données segmentées

2

Exploitation des données à grande échelle

3

Système d'aide au diagnostic



Opportunité de mieux exploiter les données du laboratoire
 ⇒ **accélération du diagnostic + tendre vers le prospectif**

Les étapes du diagnostic microbiologique classique

J0

≥ 16-24 heures

≥ 48 heures

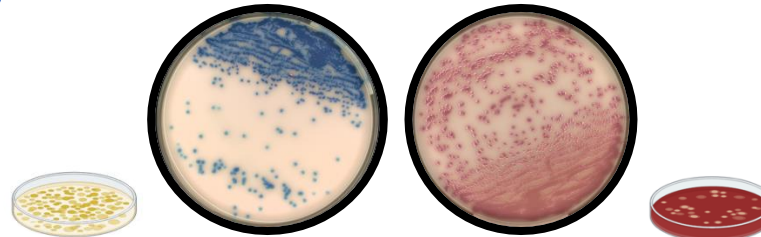
1

Examen microscopique

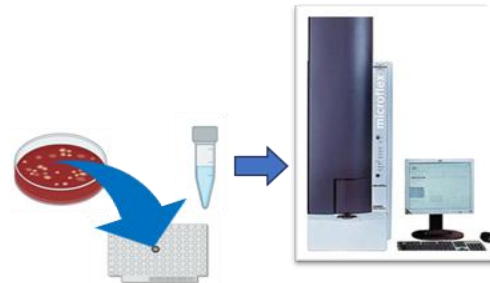


2

Culture



Identification



Spectrométrie de masse de type
MALDI-TOF

3

Détermination de la sensibilité aux antibiotiques

→ antibiogramme



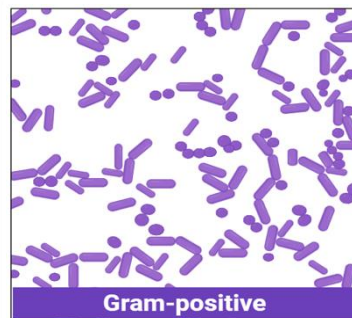
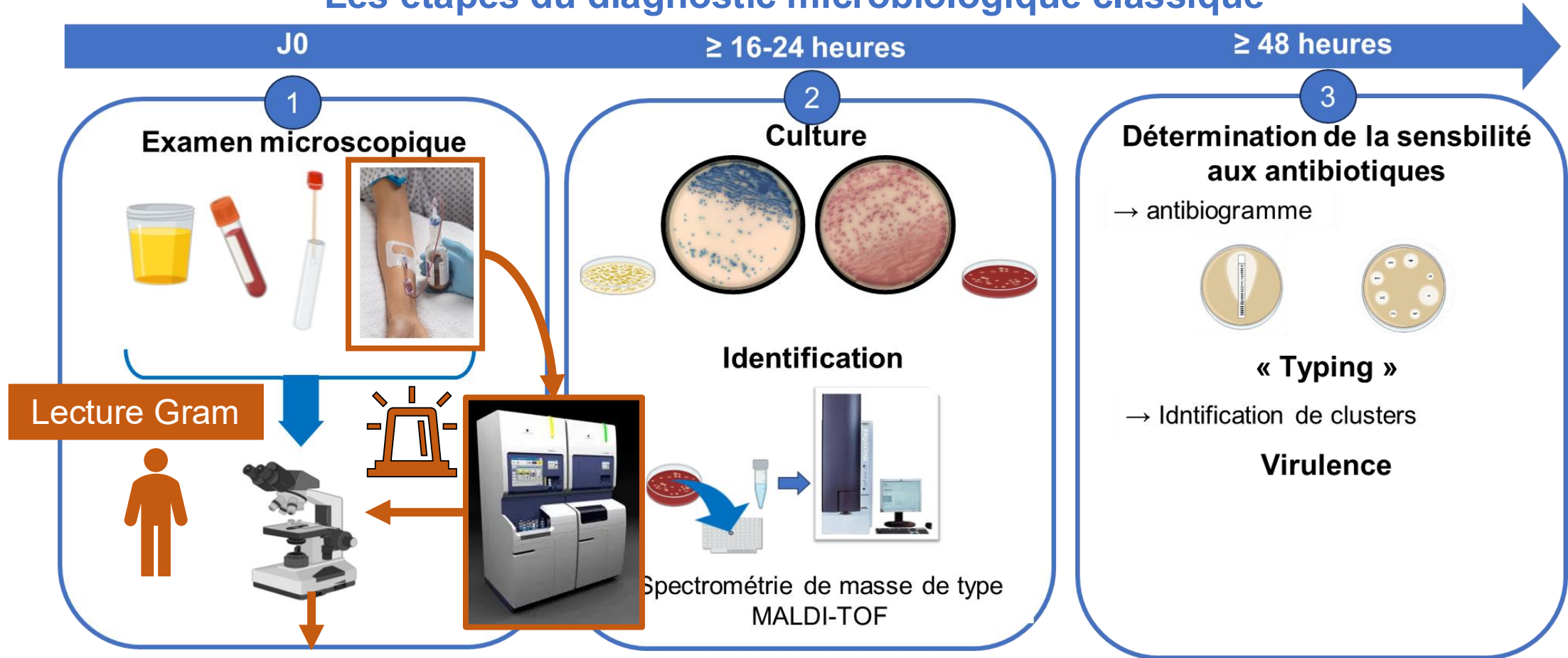
« Typing »

→ Identification de clusters

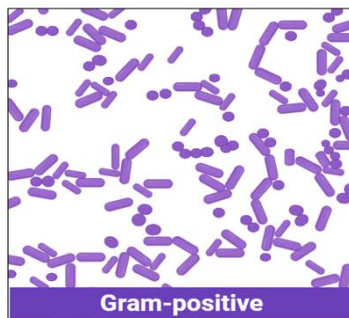
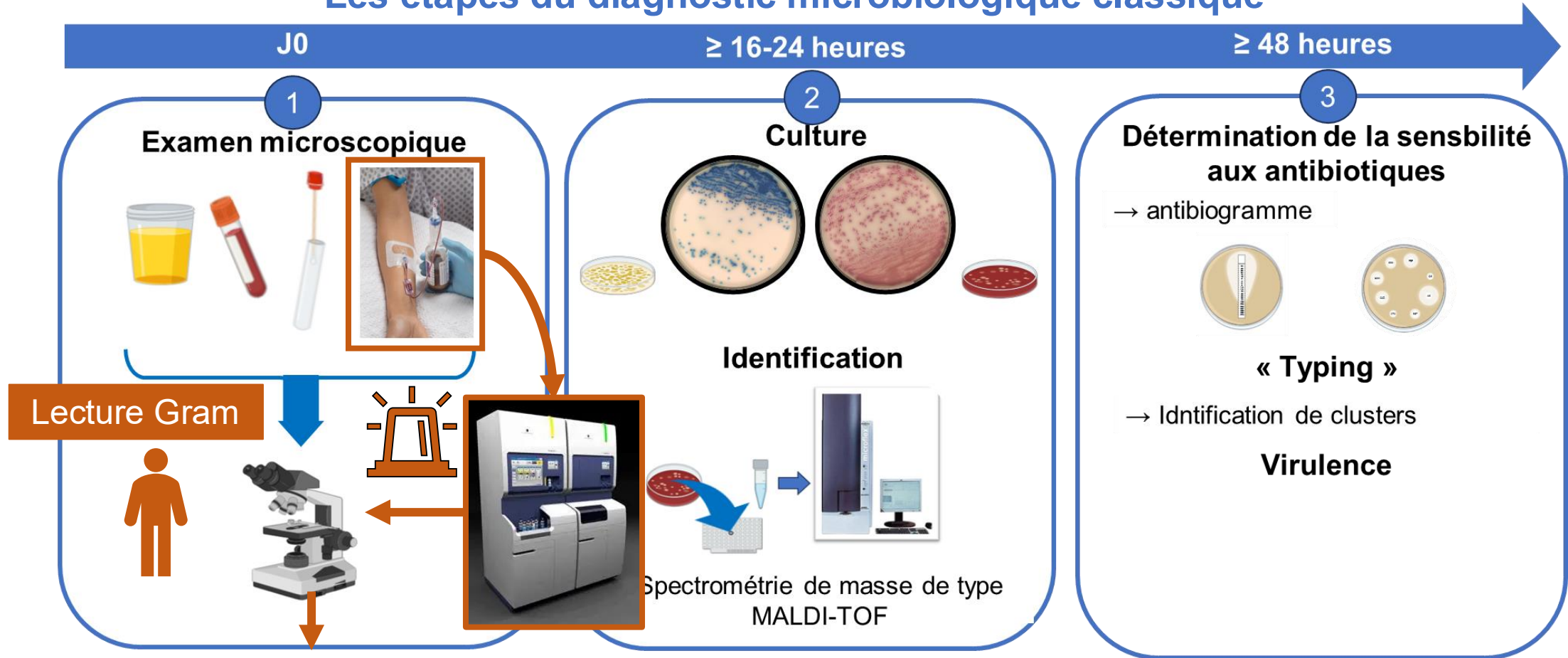
Virulence

Comment exploiter les données segmentées pour accélérer le diagnostic ?

Les étapes du diagnostic microbiologique classique

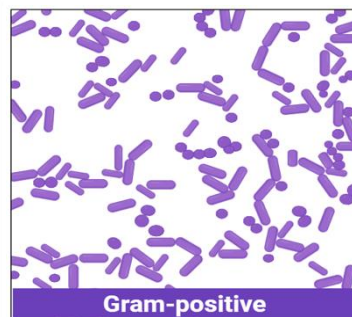
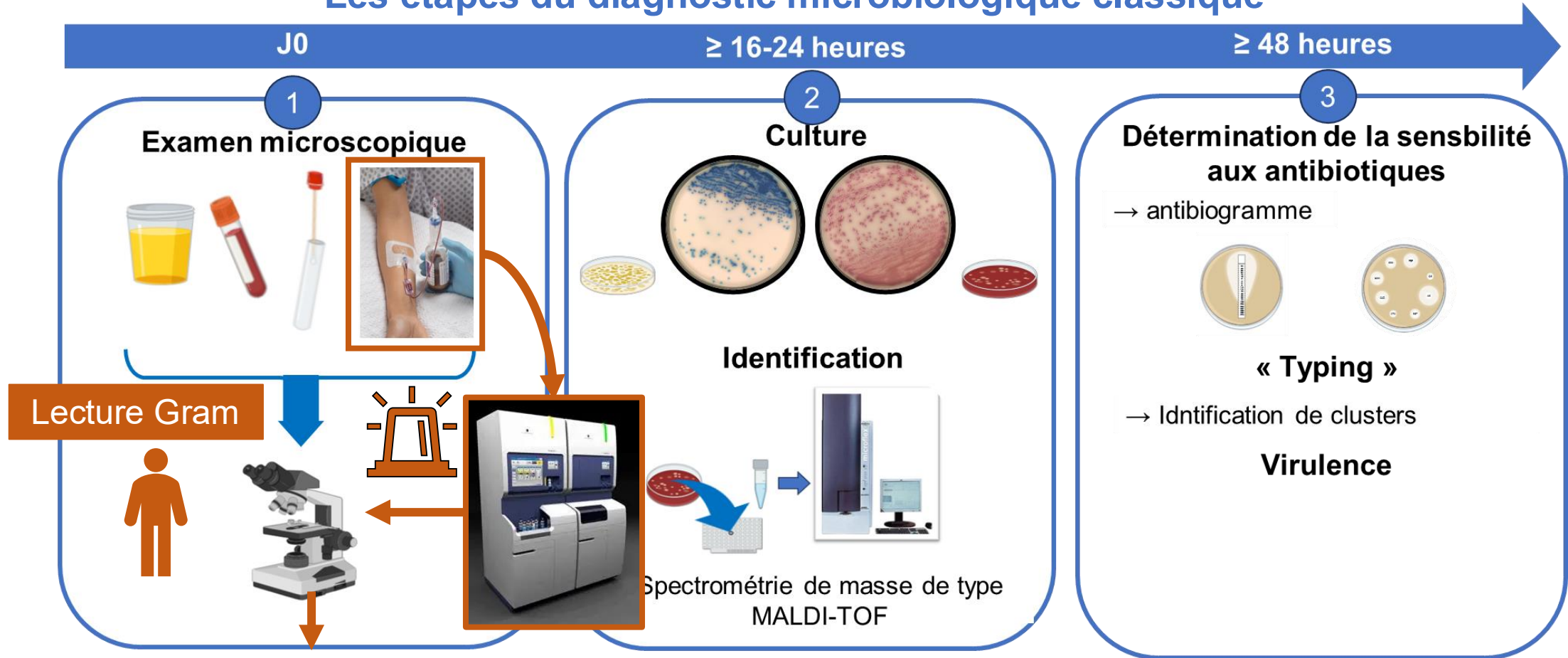


Les étapes du diagnostic microbiologique classique



- Hautement opérateur-dépendant
- Chronophage
- Pénurie de techniciens qualifiés
- Peu d'automatisation possible à ce jour

Les étapes du diagnostic microbiologique classique



- Hautement opérateur-dépendant
- Chronophage
- Pénurie de techniciens qualifiés
- Peu d'automatisation possible à ce jour

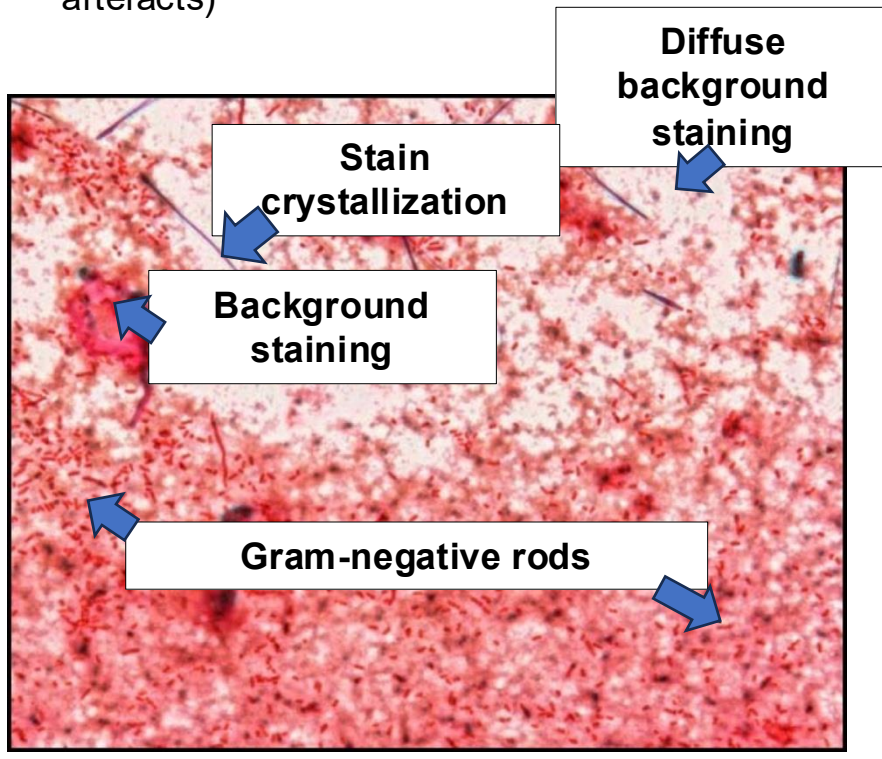
Automatisation+ interprétation des colorations de Gram des hémocultures à l'aide de l'imagerie et du Deep Learning ?

Contexte

Automatisation de la lecture des lames

468 lames d'hémoculture colorées au Gram positives

- Données recueillies dans le cadre de la pratique clinique courante (avril-juillet 2017)
- Aucune présélection (niveaux de bruit élevés dans les diapositives, bruit de fond, cristaux, artefacts)

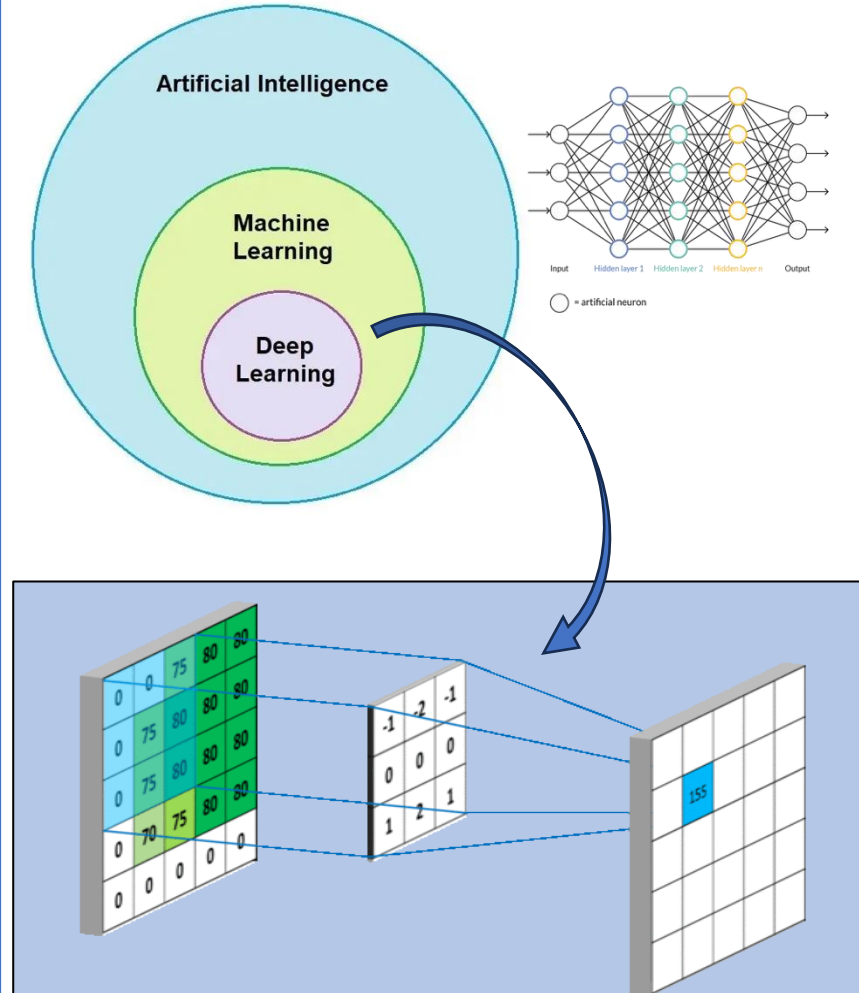
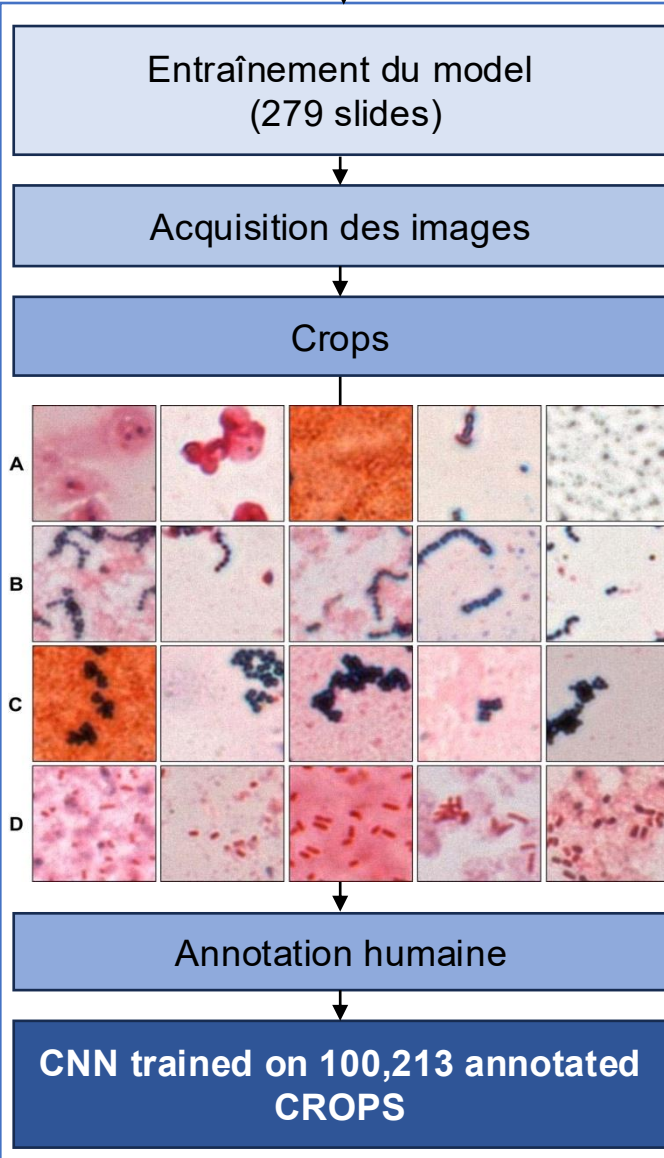
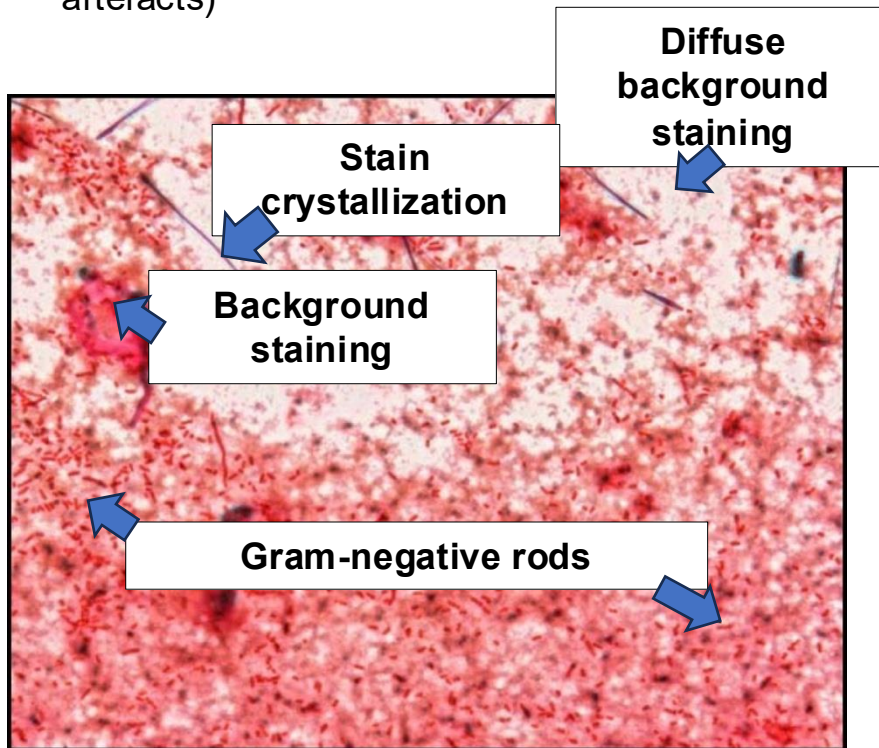


Automatisation de la lecture des lames

Methodes

468 lames d'hémoculture colorées au Gram positives

- Données recueillies dans le cadre de la pratique clinique courante (avril-juillet 2017)
- Aucune présélection (niveaux de bruit élevés dans les diapositives, bruit de fond, cristaux, artefacts)



Smith KPKang AD, Kirby JE 2018. J Clin Microbiol 56: 10.1128/jcm.01521-17.
<https://doi.org/10.1128/jcm.01521-17>

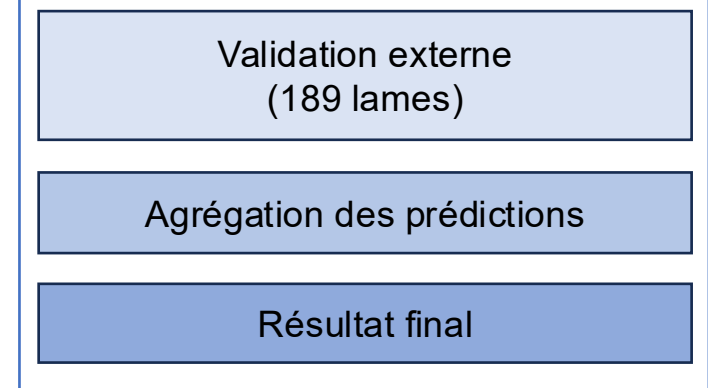
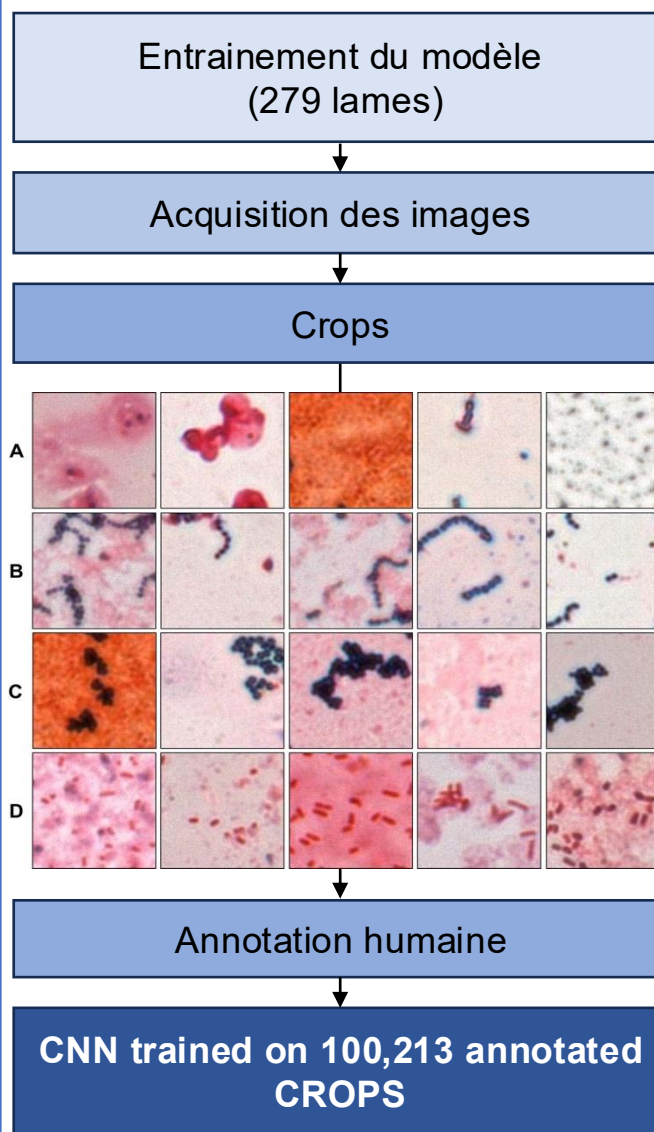
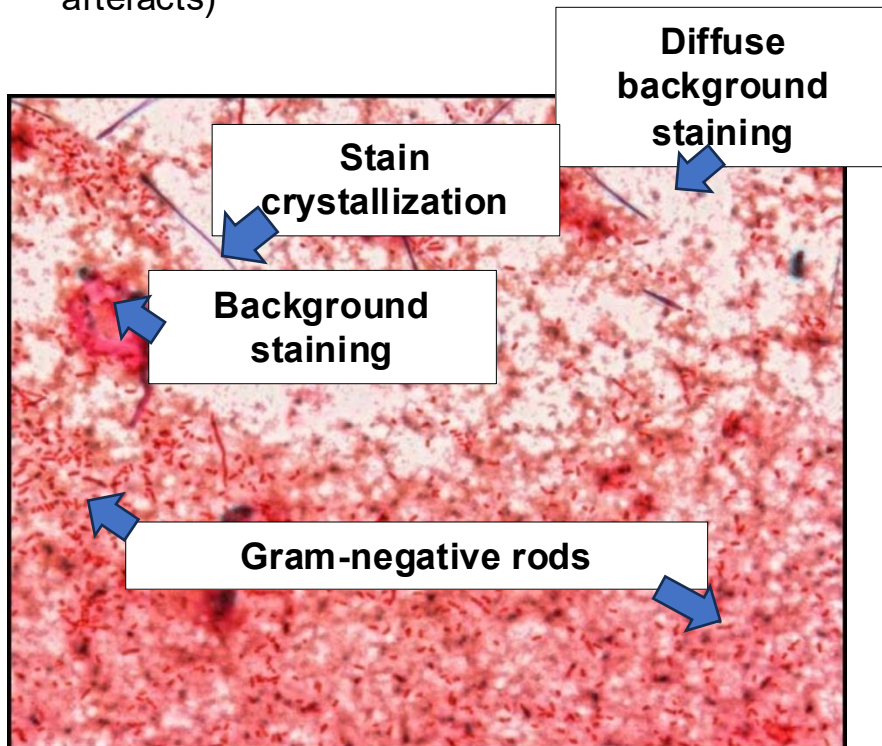
Convolutional neural networks = inspirés du cortex visual des animaux

Automatisation de la lecture des lames

Résultats

468 lames d'hémoculture colorées au Gram positives

- Données recueillies dans le cadre de la pratique clinique courante (avril-juillet 2017)
- Aucune présélection (niveaux de bruit élevés dans les diapositives, bruit de fond, cristaux, artefacts)



Comparison with expert human reading

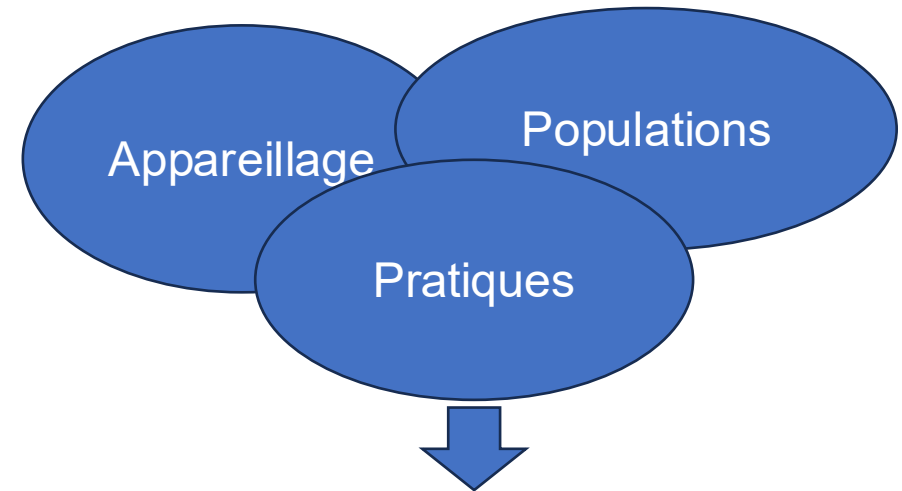
Metric	Result
Bactéries détectées	84.7% (160 / 189)
Concordance (Gram +/ Gram-)	92.5%

Smith KPKang AD, Kirby JE 2018. J Clin Microbiol 56: 10.1128/jcm.01521-17.
<https://doi.org/10.1128/jcm.01521-17>

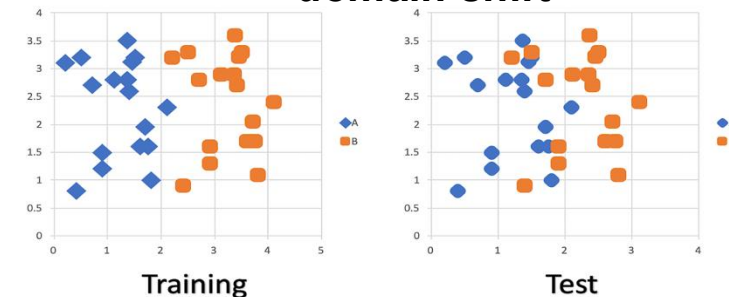
Automatisation de la lecture des lames, littérature

Référence	Application de l'IA	Performance
Droussiveau et al., 2024	Frottis respiratoires positifs ou négatifs	Accuracy = 91.3%
Panahovitz et al., 2021	Détection de bacilles acido-alcool-résistants dans biopsies	Se = 80% , Spé = 99.9%
Wang et al., 2021	Automatisation du score de Nugent (vaginose bactérienne)	Accuracy = 75.1%
Mathison et al., 2020	Détection automatique de protozoaires intestinaux	Accord positif = 98.9%
Torres et al., 2018	Détection de parasites (paludismes) sur frottis Giemsa	Se = 72% , Spé = 85%

Le gouffre de l'implémentation



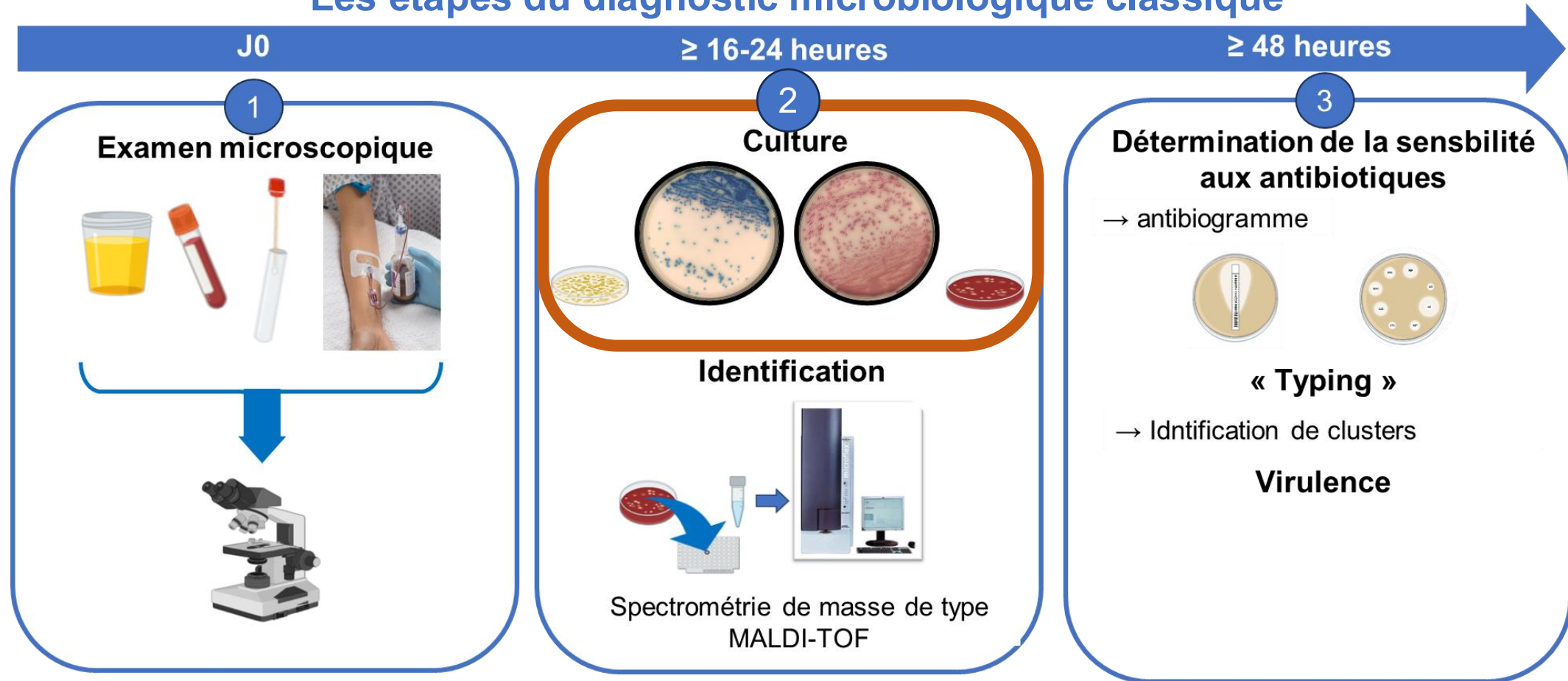
Problématique :
« domain shift »



Robustesse dans des laboratoires différents ?

Le Gram automatisé est prometteur pour de gagner du temps sur des tâches répétitives, mais on manque encore de validation multicentrique et d'études montrant un bénéfice clinique direct

Les étapes du diagnostic microbiologique classique



Lecture des cultures



- Tâche visuelle complexe
- Chronophage
- Dépendante de l'expertise humaine

Utilisation du Deep Learning pour l'interprétation des cultures ?

Automatisation de la lecture des cultures

Méthode

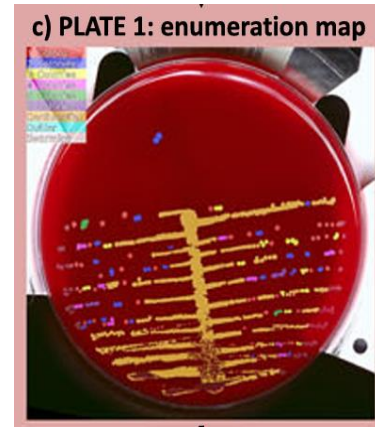
Entraînement

1 351 images de culture
(26 213 images de colonies individuelles,
32 espèces)

Ensemencement et incubation
(WaspLab™)
Numérisation haute définition

Architecture « DeepColony »

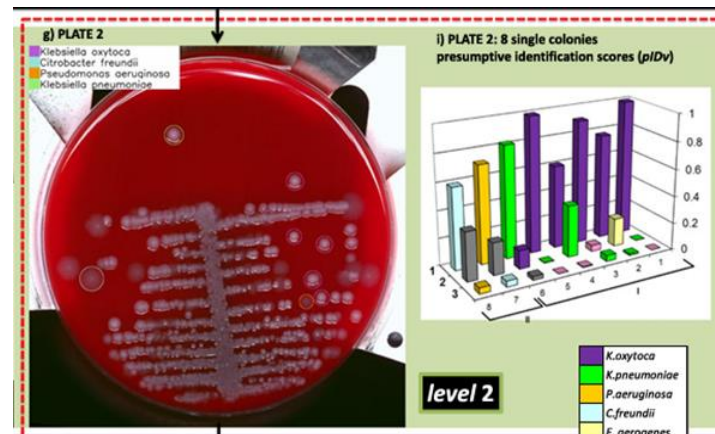
1. Détection des colonies sur la boîte



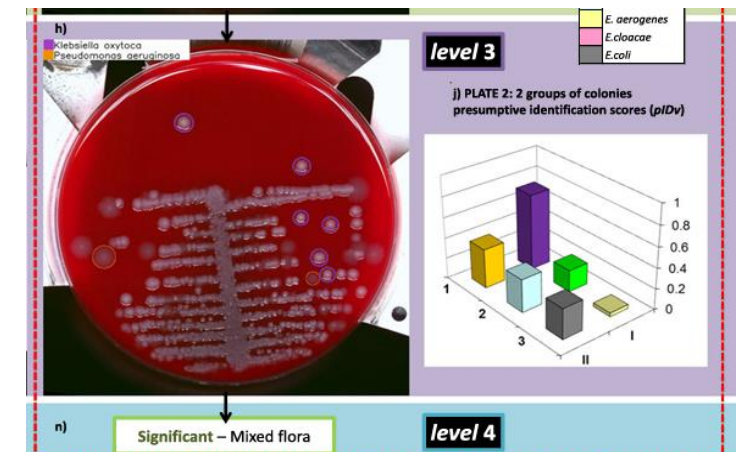
2. Sélection des colonies pertinentes



3. Identification présumptive



4. contexte de la boîte



Signoroni, A et al. Hierarchical. *Nat Commun*, (2023).
<https://doi.org/10.1038/s41467-023-42563-1>

« Plate-level interpretation » system ⇒ identification, nombre et distribution des colonies

Automatisation de la lecture des cultures

Résultats

Entraînement

1 351 images de culture
(26 213 images de colonies individuelles,
32 espèces)

Ensemencement et incubation
(WaspLab™)
Numérisation haute définition

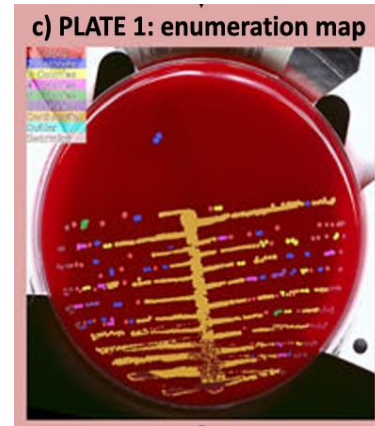
Architecture « DeepColony »

Validation externe (5 051 images)

- Accord (culture significative/flore) = 95%
- Concordance Id= 95% (vs. MALDI-TOF)

Signoroni, A et al. Hierarchical. *Nat Commun*, (2023).
<https://doi.org/10.1038/s41467-023-42563-1>

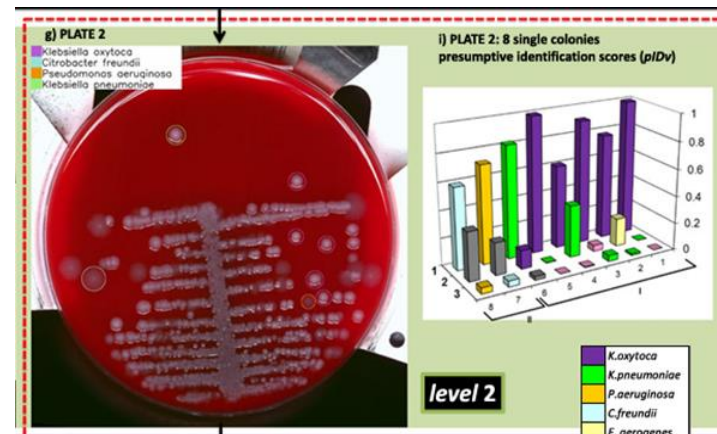
1. Détection des colonies sur la boîte



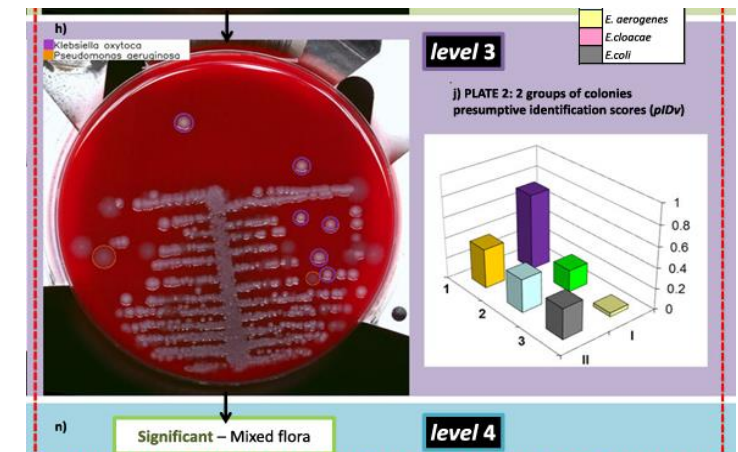
2. Sélection des colonies pertinentes



3. Identification présumptive



4. contexte de la boîte



Priorisation du rendu des cultures non contributives = accélération du tri des cultures

Automatisation de la lecture des cultures, littérature

Application IA	Référence	Performance
Détection VRE sur milieu chromogène	Faron ML et al., J Clin Microbiol, 2016	Se =100% ; Spé = 89.5%
Détection MRSA sur milieu chromogène	Faron ML et al., J Clin Microbiol, 2016	Se =100% ; Spé = 90.7%
Détection Streptococcus pyogenes	Van TT et al., J Clin Microbiol, 2019	Se = 90.6% ; Spé = 94%
Dépistage CPE	Foschi C et al., J Glob Antimicrob Resist, 2020	Se = 100% ; Spé = 77-79%
Culture urinaire (positive / négative)	Faron ML et al., J Clin Microbiol, 2019	Se = 99.8% ; Spé = 68.5%
Détection croissance bactérienne	Jacot D et al., J Clin Microbiol, 2024	Sen = 99.6% ; Spé = 91.9%
Catégorisation cultures urinaires	Alouani DJ et al., Clin Chem, 2022	AUC ≥0.98
Quantification colonies urinaires	Uwamino Y et al., Diagn Microbiol Infect Dis, 2021	Accord 95%

Se=> peu de cultures positives manquées
Spé=> validation humaine nécessaire

Systèmes actuels disponibles :

PhenoMATRIX™ (Copan) (Faron ML et al., JCM, 2019)

- Se >99% et Spé 68,5-72% (culture urinaire)
- Performances détectées Strepto A et B sur milieux chromogènes)
- gain de temps ≈ 4 h pour rendre les cultures négatives

Kiestra TLA (BD) (Theparee T et al. JCM. 2018)

- réduction d'environ 24 % du temps technicien
- + détection de pathogènes et meilleure isolation des colonies

APAS (clever Culture System) (Chiu K et al., Spectrum, 2022)

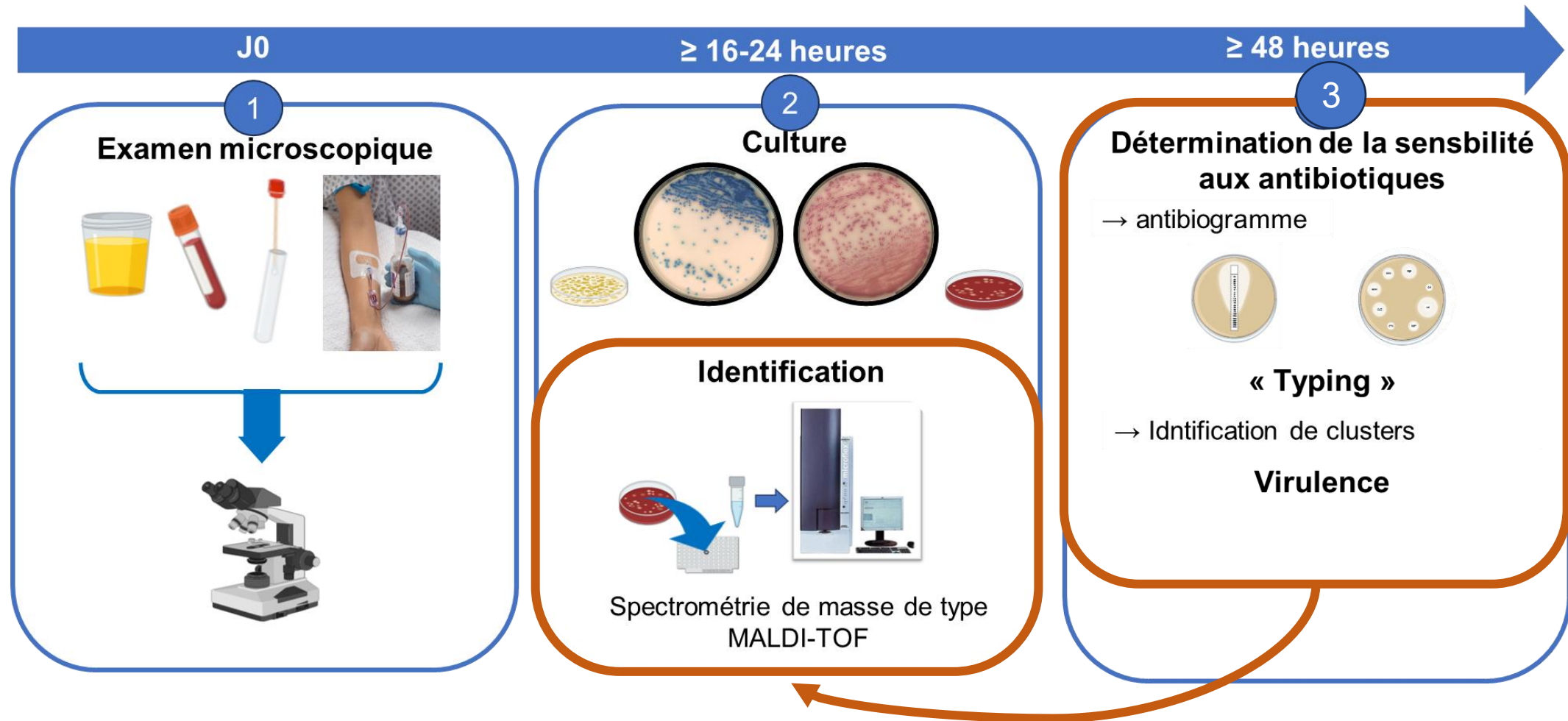
- Se = 99% et Spé = 84,5% (culture urinaire)
- 96% d'accord global
- 45 % de cultures automatiquement classées comme négative

Les limites :

- systèmes ne sont pas entièrement autonomes / coût +++
- résultats à confirmer par le microbiologiste
- erreurs : artéfact, culture polymicrobienne, standardisation (milieux)
- problème récurrent du « domain shift »
- performances difficilement comparables entres les études (biais de métrique), **Quid de la VPN et VPP ?**

Aujourd'hui, la principale valeur de ces systèmes n'est pas de décider, mais de prioriser

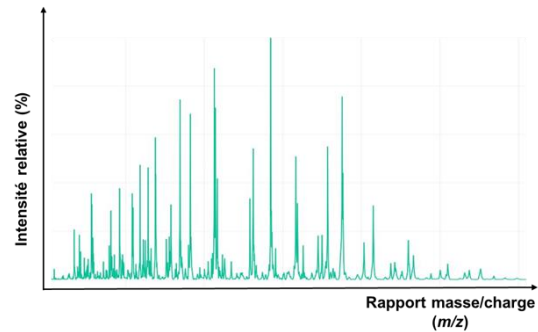
Amélioration des performances de la spectrométrie de masse



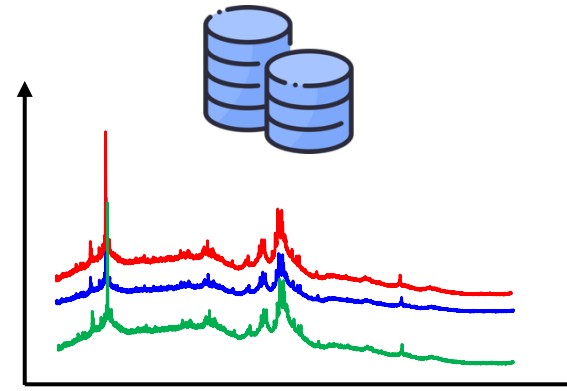
Les données de la spectrométrie de masse peuvent répondre à d'autre tâche que l'identification ?

Amélioration des performances de la spectrométrie de masse

Spectre de masse



Base de données



Identification



Exemples :
K. pneumoniae
C. albicans

Algorithme

Comparaison

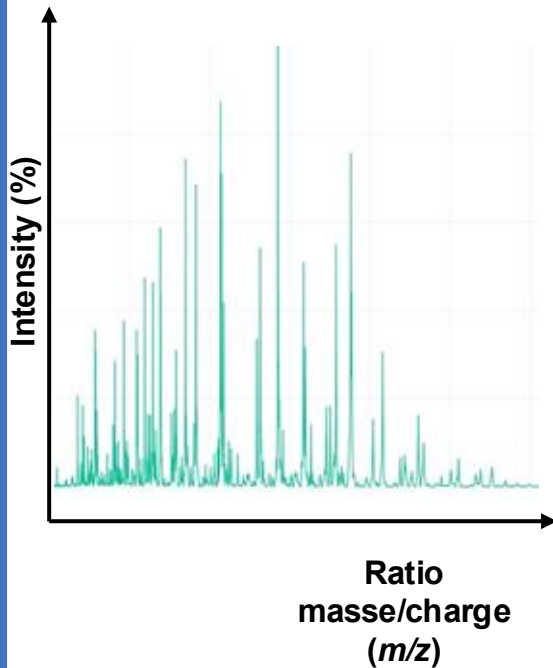


Algorithme : élément important pour extraire les informations spectrales

Amélioration des performances de la spectrométrie de masse

1

Spectre de masse



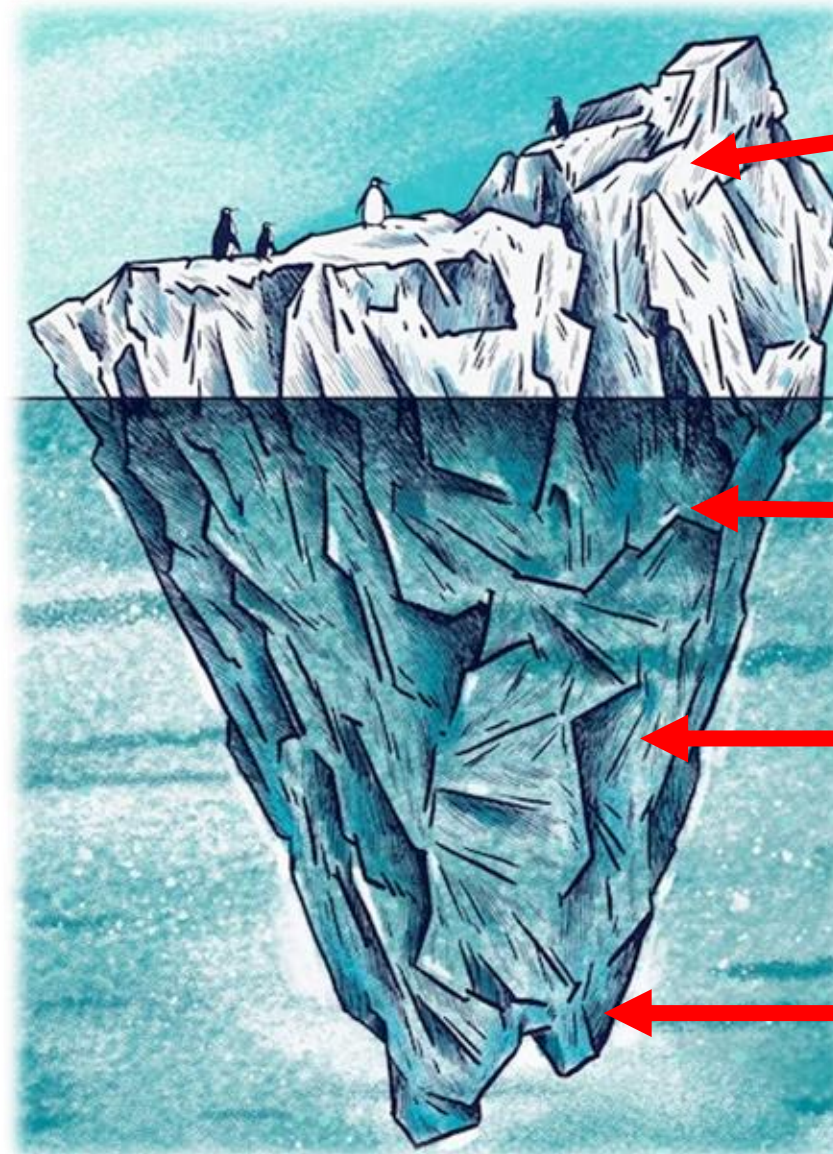
2

Base de données



3

Algorithme

Identification de
micro-organismes

Resistance

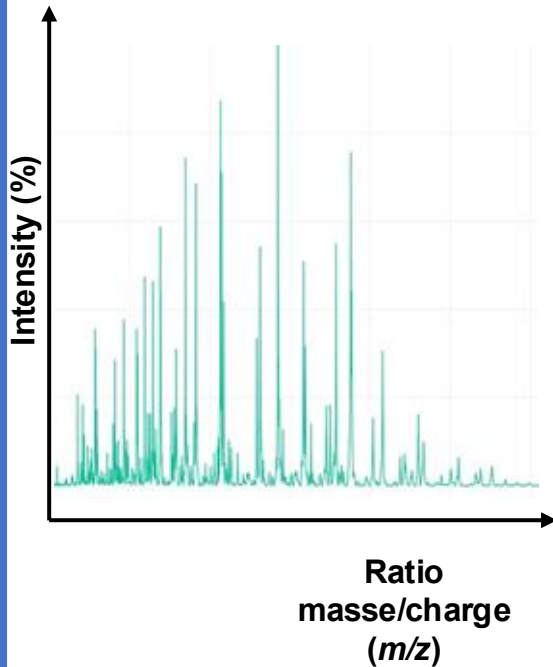
Virulence

« Typing »

Amélioration des performances de la spectrométrie de masse, limites des systems commerciaux

1

Spectre de masse



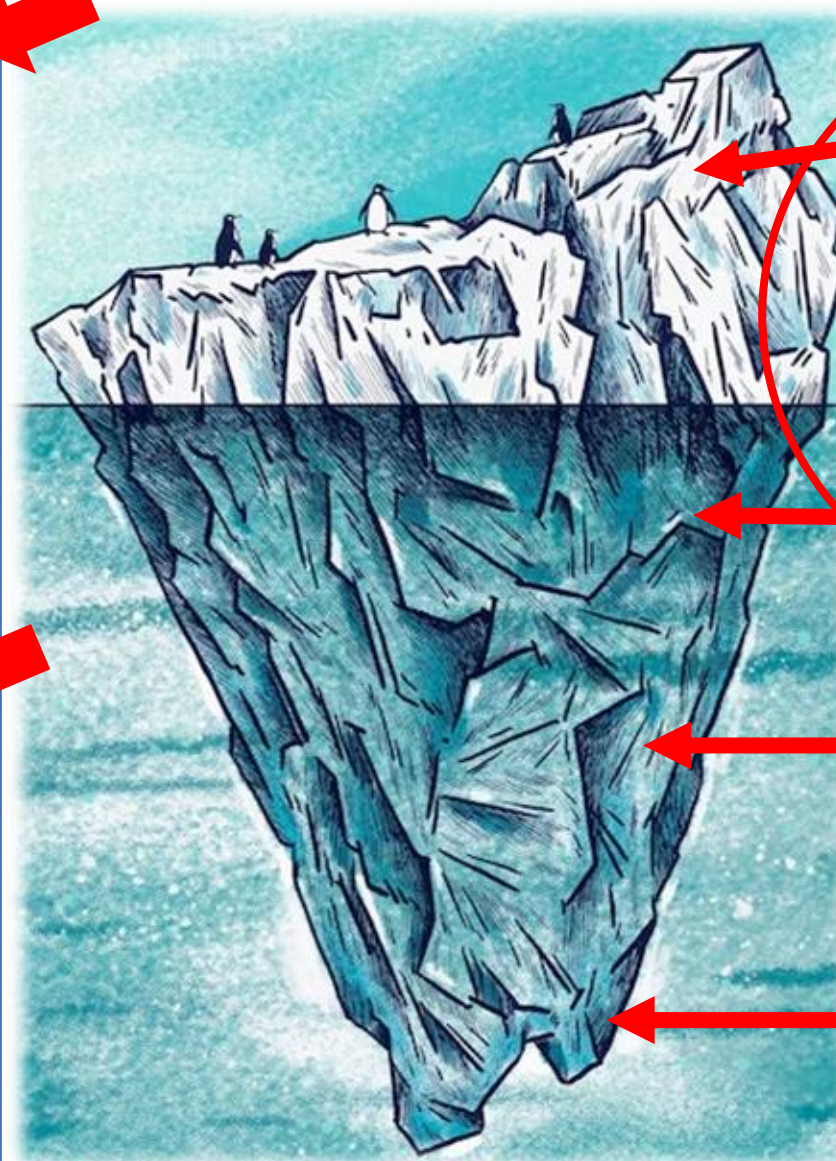
2

Base de données



3

Algorithme



Identification de micro-organismes

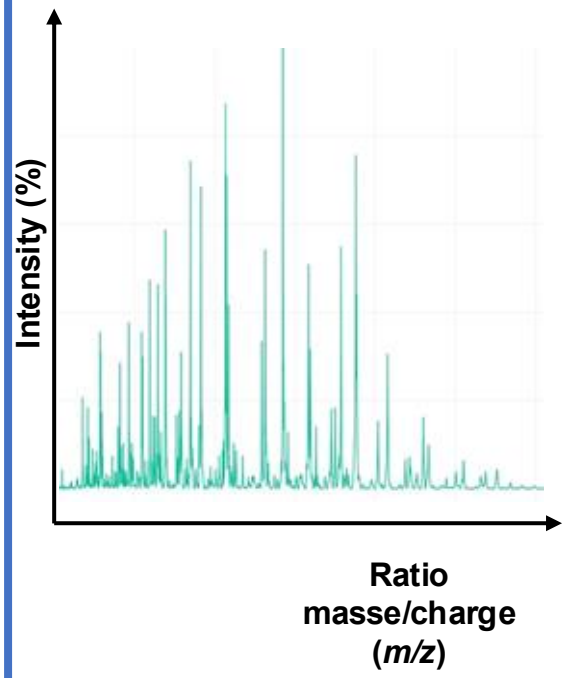
Resistance

Virulence

« Typing »

Amélioration des performances de la spectrométrie de masse

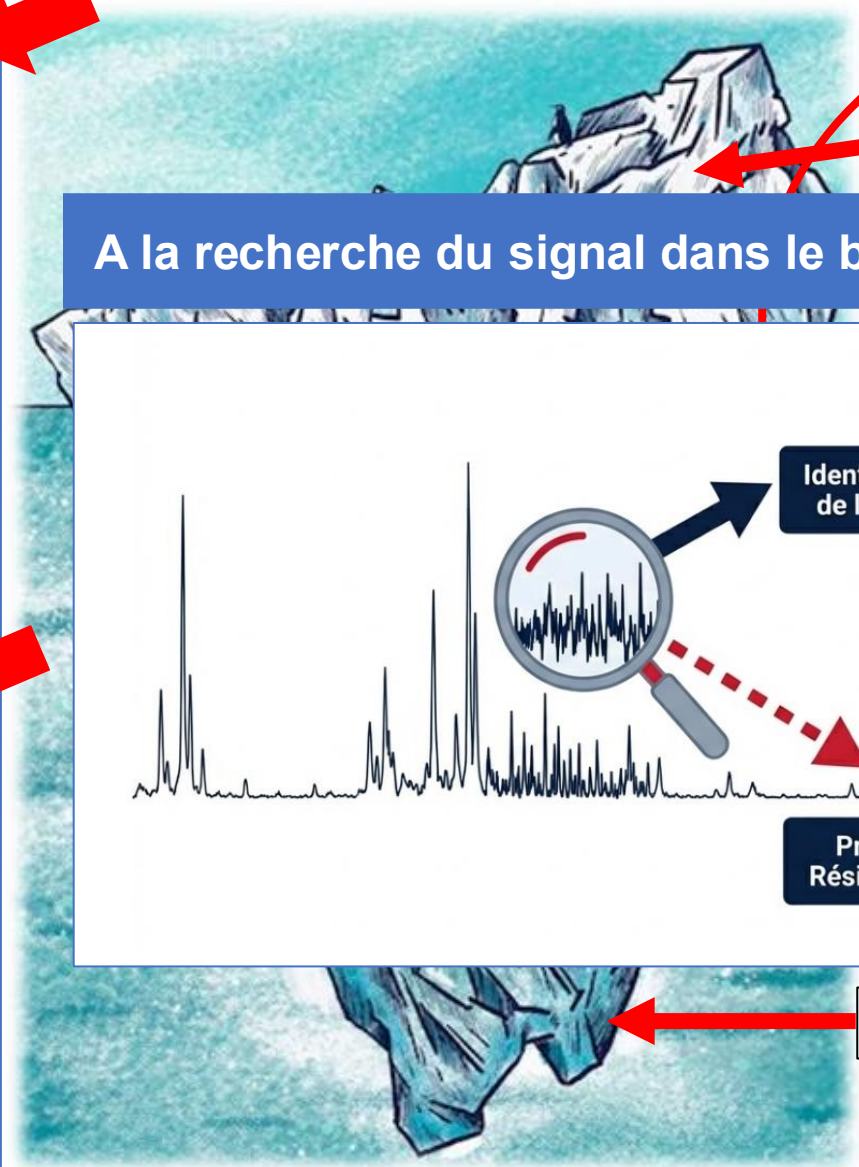
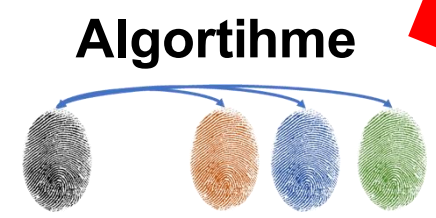
1 Spectre de masse



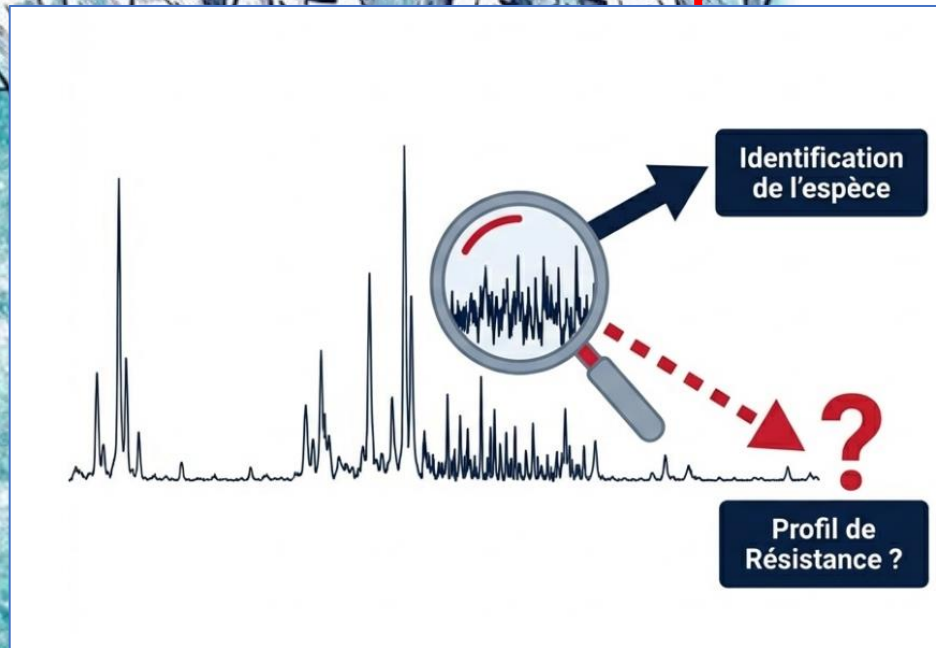
2 Base de données



3 Algorithme



A la recherche du signal dans le bruit !



Identification de micro-organismes

Resistance

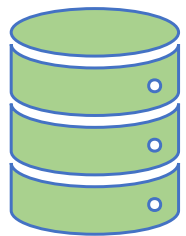
Virulence

« Typing »

Amélioration des performances de la spectrométrie de masse

Base de données

4 laboratoires de routine



DRIAMS

- > 300 000 spectres cliniques
- <750 000 phénotypes de résistance

Traitement du signal

+

Entraînement des algorithmes



Régression Logistique



LightGBM



Réseau de Neurones Profond (MLP)

20%
Test (Validation)80%
Entraînement

Performances prédictives

S. aureus

(Oxacilline / Risque SARM)



AUROC (Excellente précision)

E. coli

(Ceftriaxone / Risque BLSE)

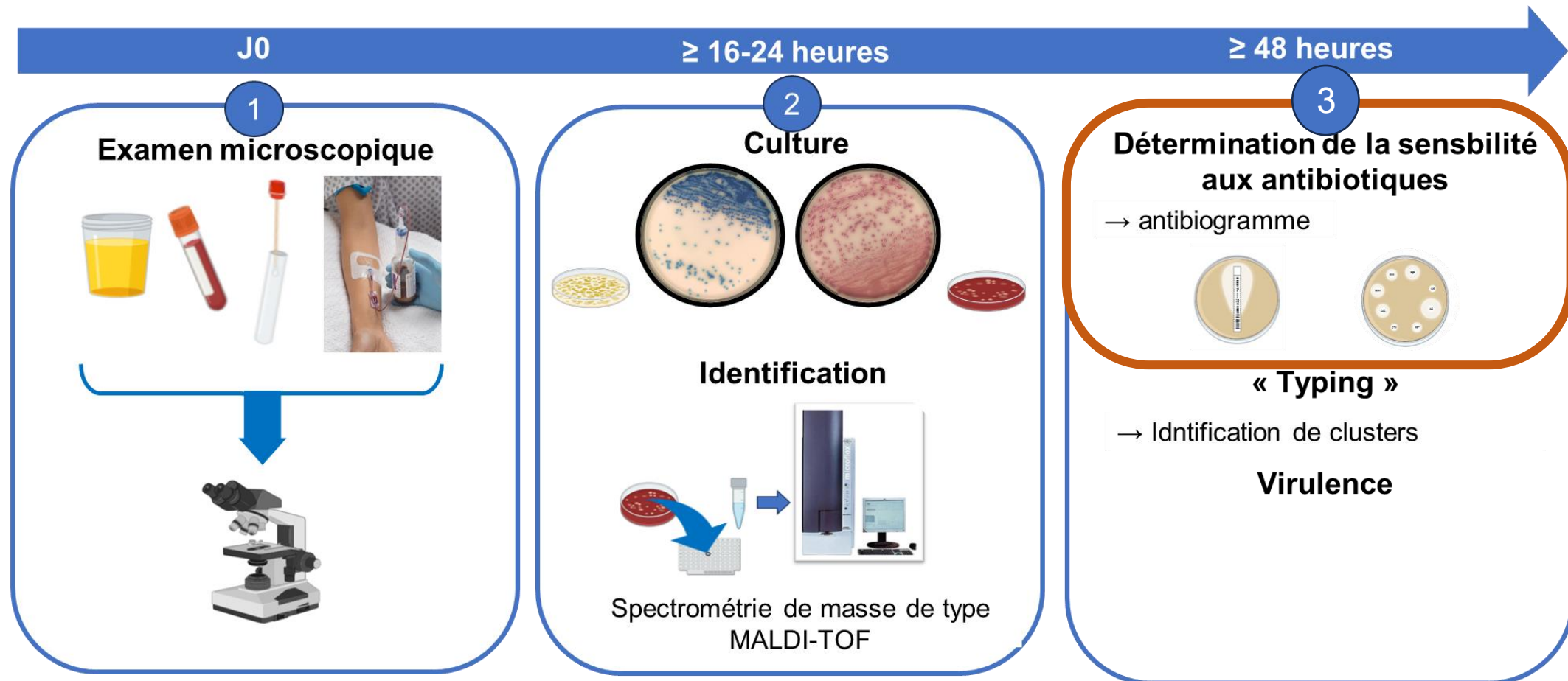


AUROC

Sur 42 antibiotiques étudiés :
31/42 ont une AUC-ROC > 0.80

Performances bonnes mais AUC-ROC difficilement interprétable, quid du modèle en prospectif ?

Les étapes du diagnostic microbiologique classique



MALCA : « MACHine Learning Carbapenemase » : détection des carbapénémases

1

Contexte

- Les CPE sont une menace mondiale, responsables d'infections difficiles à traiter
- Le typage des carbapénémases (KPC, NDM, OXA-48-like) est essentiel pour adapter l'antibiothérapie

2

Problématique

- Méthodes actuelles de détection restent coûteuses, chronophages et limitées (PCR, test immunochromatographiques)
- Nombreux tests complémentaires

3

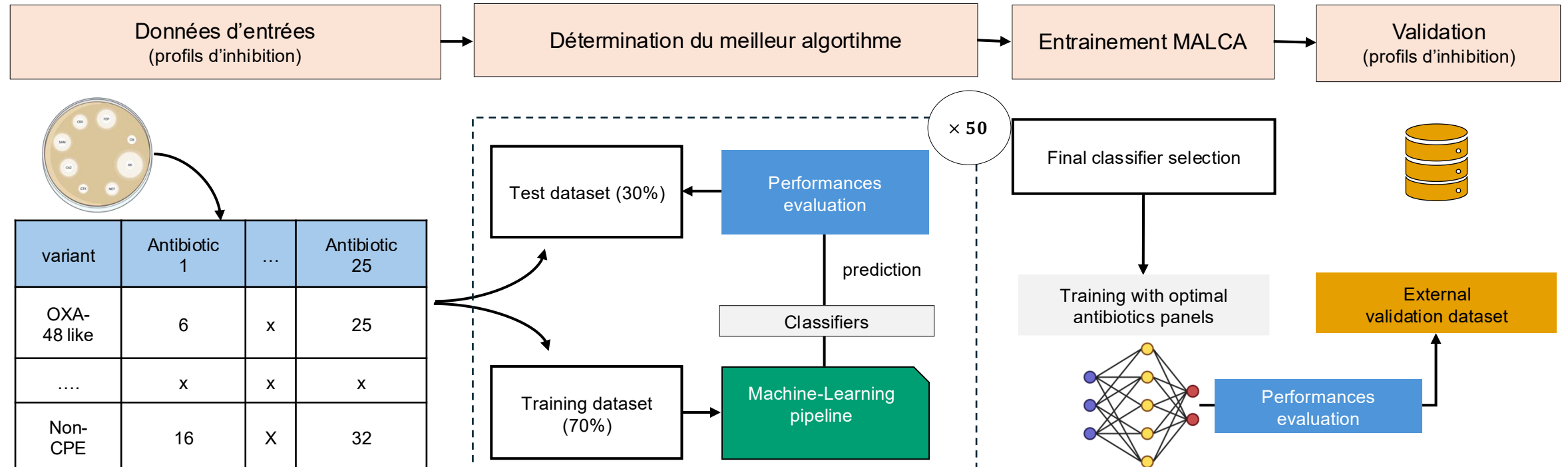


Méthodes

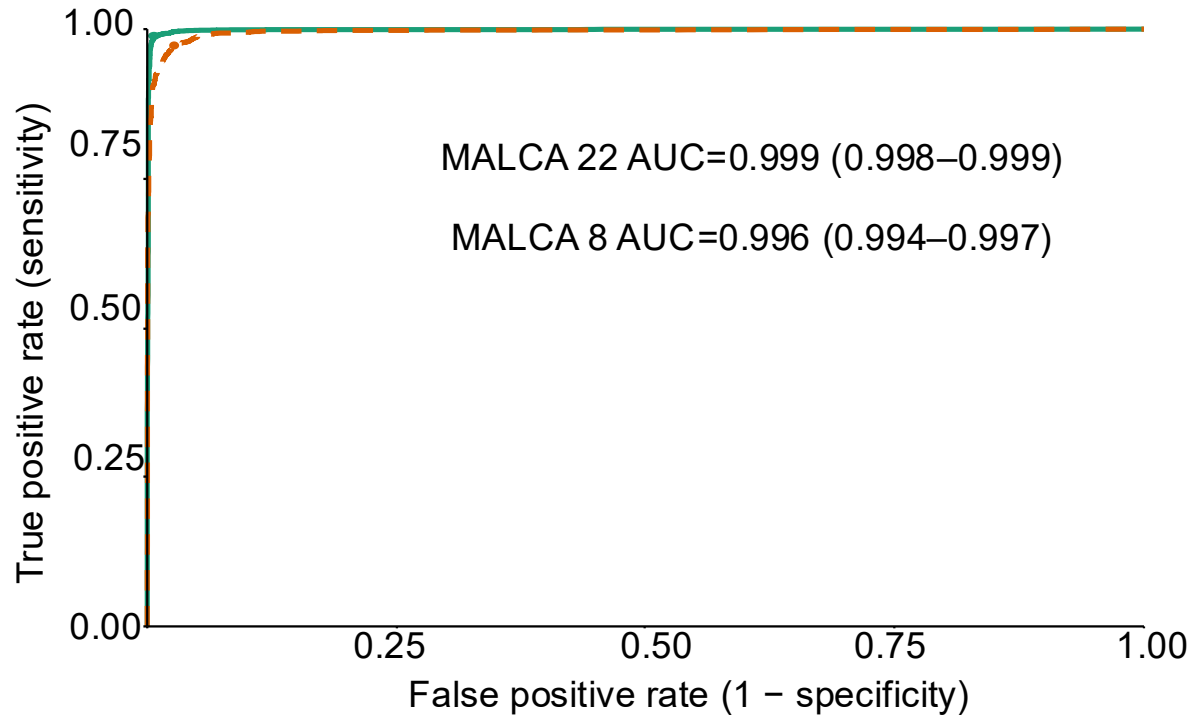


11 992 isolats (CNR et routine)

8 514 isolats (CNR et routine)



MALCA : « MACHine Learning Carbapenemase » : détection des carbapénémases



AUC-ROC élevée pour discrimination carbapénémase / non carbapénémases

Metric	MALCA-22	MALCA-8	CA-SFM	EUCAST
Sensitivity	99.4%	96.5%	97.3%	88.3%
Specificity	98.8%	97.8%	88.4%	87.7%
Tests de confirmation	NA	NA	4787	4511
VPP	98.5%	97.3%	87.0%	85.2%
VPN	99.6%	97.2%	97.6%	90.3%
Accuracy	99.1%	97.3%	92.3%	87.9%



Dataset	Prevalence	MALCA	Se/Sp	VPP/VN
Routine	4.30%	22	100%/99.5%	89.6%/100%
		8	96.3%/98.7%	77.4%/99.0%
CNR	88%	22	99.4%/93.1%	99.1%/95.0%
		8	96.5%/90.1%	98.6%/78.0%

Validation temporelle / pas de biais de métriques (2 prévalences réelles analysées)

Typage possible = Se > 97% and spé > 98%, pour OXA-48-like, NDM, et KPC
⇒ Analyse prospective des données d'antibiogrammes, pas de coût additionnel

Conclusion (partie laboratoire)

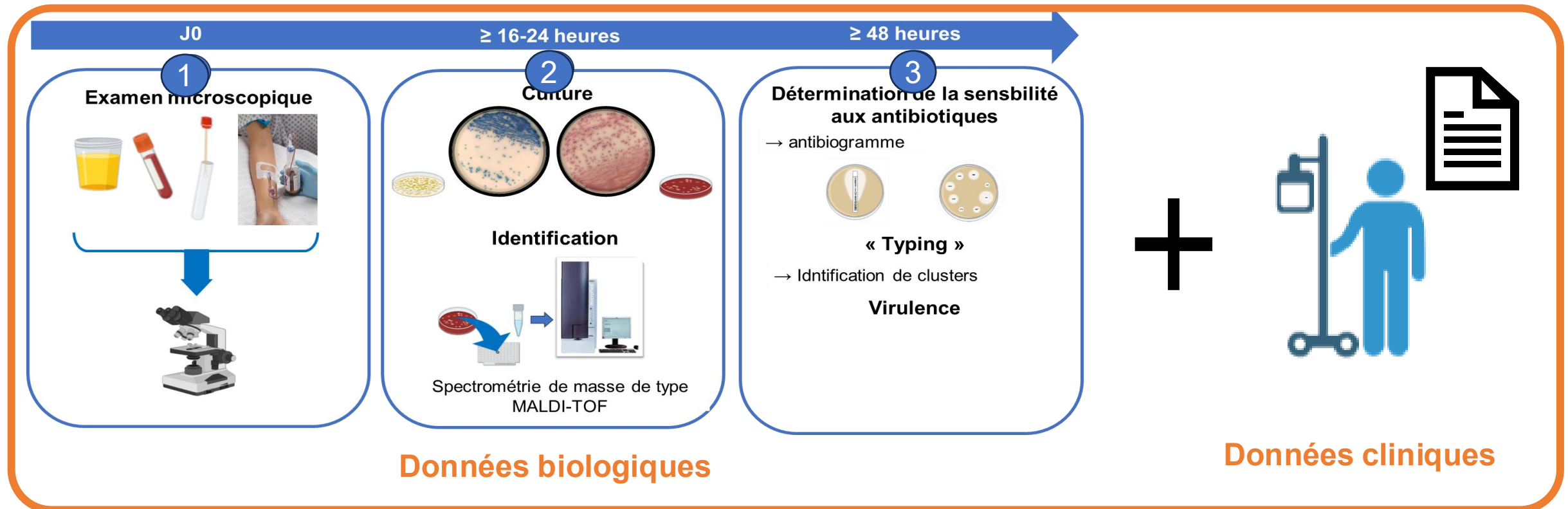
« Le vrai gain aujourd'hui n'est pas de remplacer le microbiologiste, mais de lui faire gagner du temps là où il n'apporte pas de valeur »
Peu de systèmes disponibles en routine au laboratoire

Conclusion (partie laboratoire)

« Le vrai gain aujourd'hui n'est pas de remplacer le microbiologiste, mais de lui faire gagner du temps là où il n'apporte pas de valeur »
Peu de systèmes disponibles en routine au laboratoire



Et les modèles intégratifs ?
⇒ Impact clinique : le vrai enjeu



Particularités du diagnostic microbiologique chez l'enfant : exemple du sepsis néonatal

1

Contexte

- > 1 million de décès dans le monde (Valiescu et al., 2025)
- 3^{ème} cause de mortalité néonatale (Zea-Vera et Ochoa, 2015)
- Système immunitaire immature (production de cytokines altérées)

Vulnérabilité

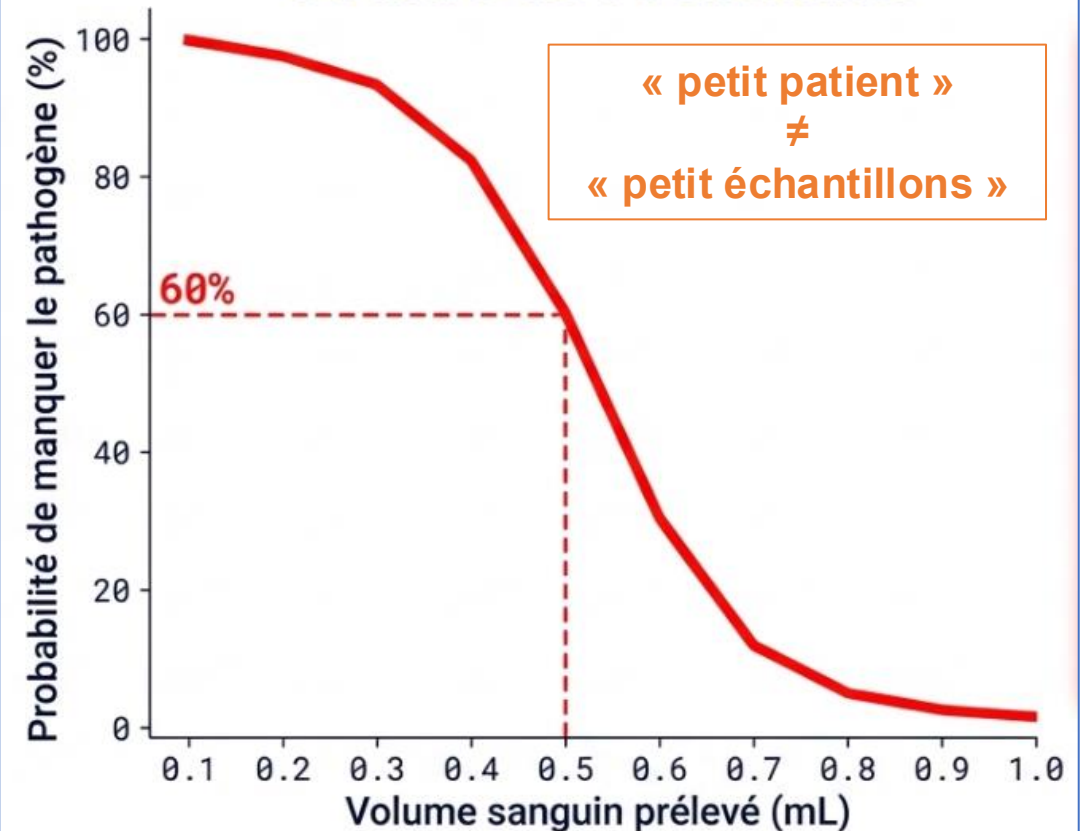
2

Problématique

- Sensibilité CRP et PCT (Se < 65%)
- Sensibilité médiocre des hémoculture (Se = 35-60%)
- 3 contraintes en pédiatrie
 - volume de sang limité
 - densité bactérienne (68% avec ≤ 10 CFU/mL)
 - incompatibilité entre temps d'incubation et détérioration fulgurante du patient

**Traitement empirique (ration 16/1)
sans preuve microbiologique +++**

Courbe de Probabilité

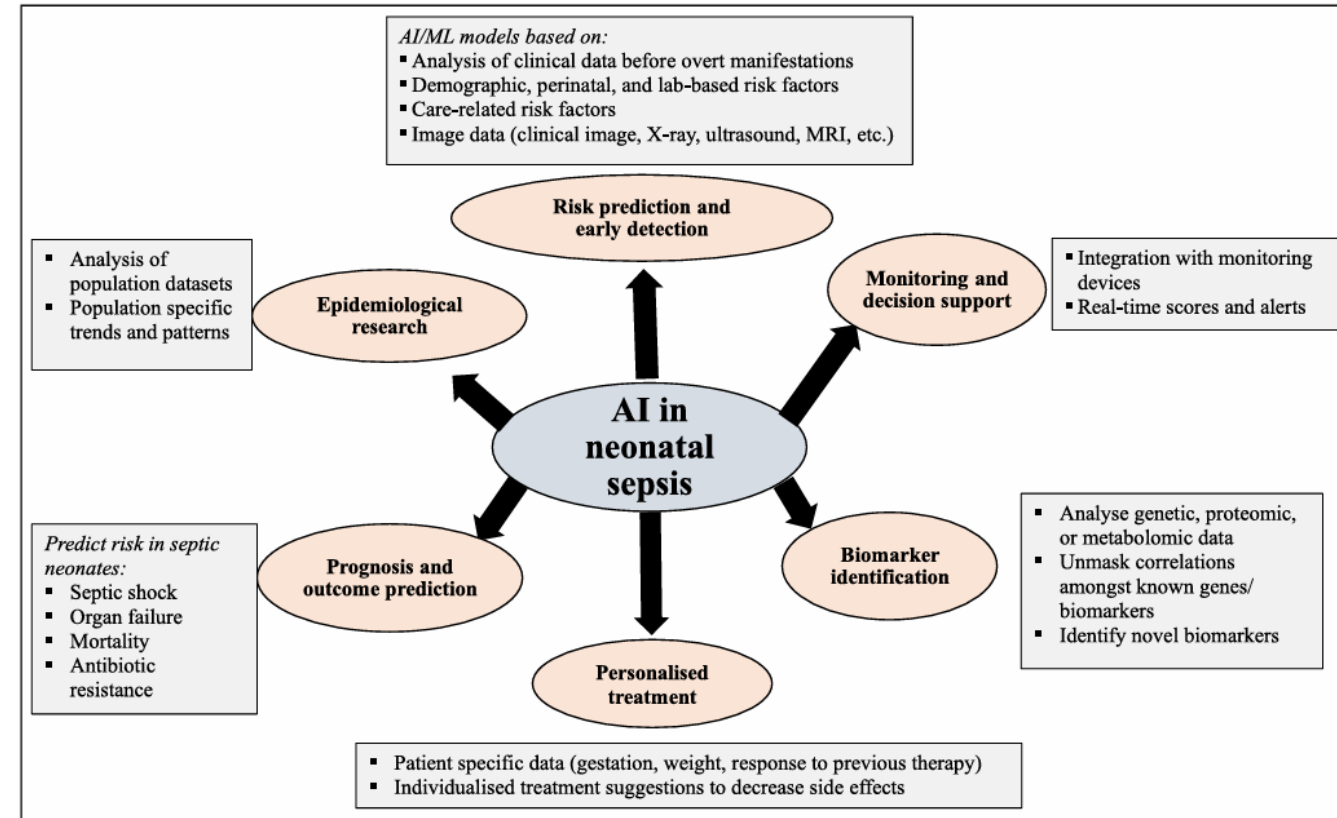


Chez l'enfant : équation diagnostic biologiquement et physiquement inadaptée à la physiologie pédiatrique avec les techniques diagnostic actuelle actuelle

La promesse numérique (exemple des du sepsis néonatal)

La promesse:

- Intégration multi-données :
 - Clinique (signes non spécifiques)
 - Biologie (CRP, PCT, NFS)
 - Microbiologie (hémocultures, PCR, etc)
 - Automatisation = surveillance continue
- ⇒ Détection de patterns invisibles individuellement



Kainth D et al., Artificial intelligence in neonatal sepsis: Scope, challenges, and potential solutions! Semin Fetal Neonatal Med. 2026. doi: 10.1016/j.siny.2025.101687

Le diagnostic du sepsis repose sur des données fragmentées et imparfaites
 ⇒ IA = intégration et interprétation globale pour divers objectifs

La promesse numérique (exemple des du sepsis néonatal)

Zone d'Intervention de l'IA



Est-ce qu'il y a sepsis ?

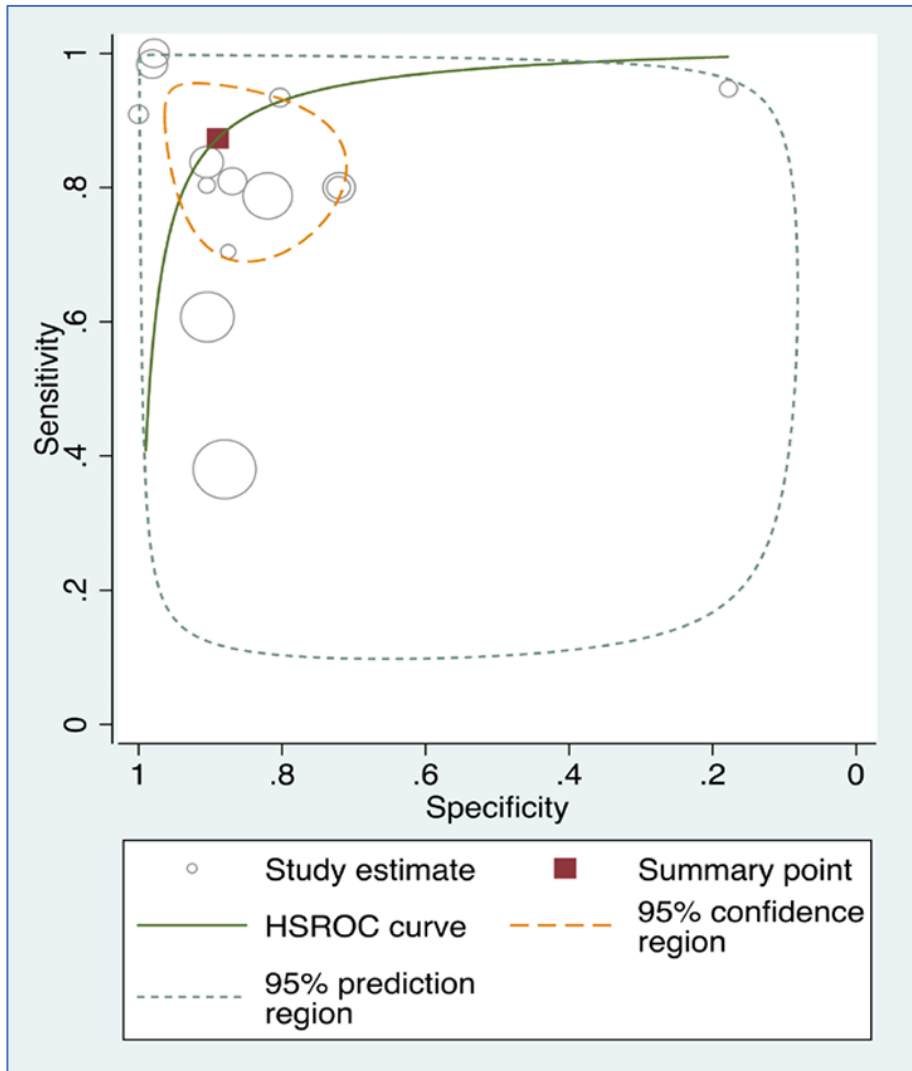
- Microbiologie :
 - hémocultures = gold standard mais : délai 24-48 h
 - sensibilité limitée (faible volume sanguin)
 - PCR : rapide mais ciblée, coûteuse
- Biologie :
 - CRP / PCT : cinétique retardée
 - performances variables selon les études
- Clinique : signes peu spécifiques



**Ne répond pas à la question :
qui va faire une infection ?**

**Données disponibles = tardives, incomplètes, peu spécifiques
⇒ décision empirique + surconsommation d'antibiotiques**

Résultats (exemple des du sepsis néonatal)



Performances de la méta-analyse (n=19 études) (15,984 participants, 76 algorithmes de ML)

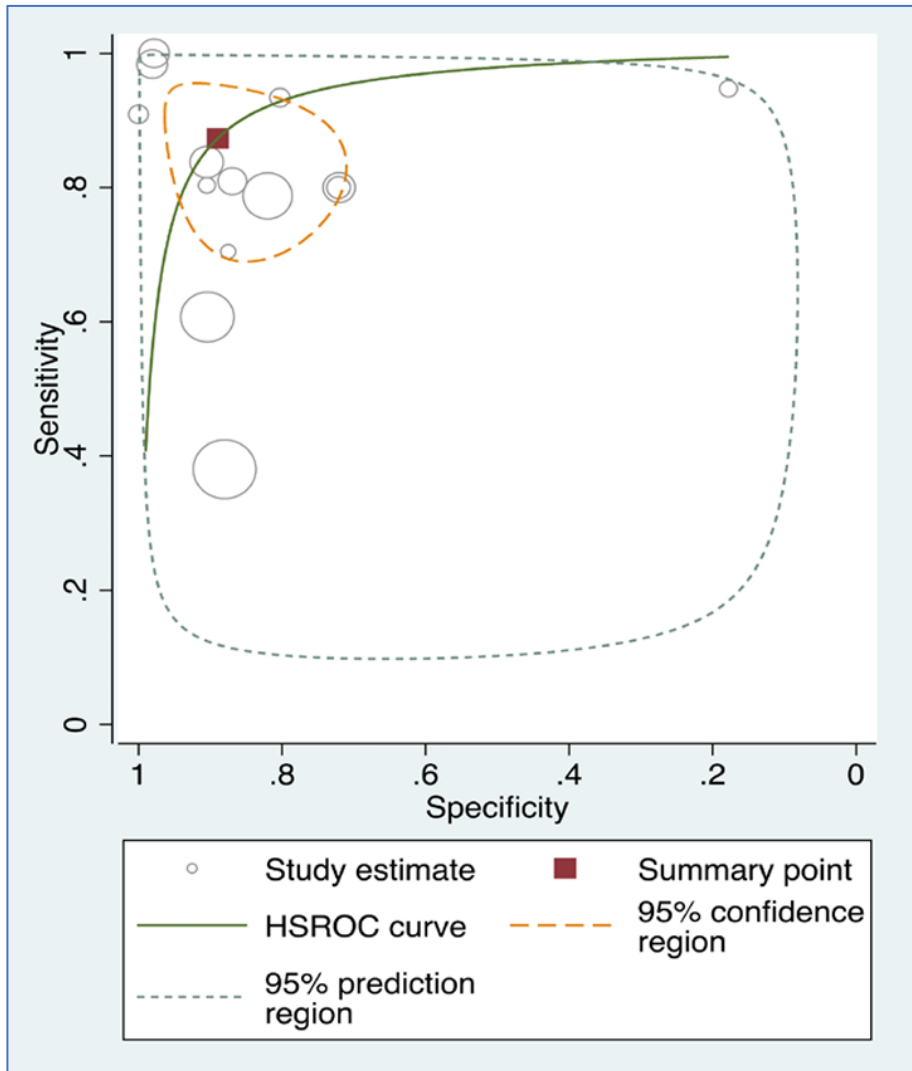
- AUC : 0,92-0,94
- Sensibilité : 0,87
- Spécificité : 0,89

$I^2 = 95,5$: performances élevées MAIS très variables

Kainth et al. Diagnostic Performance of Machine Learning-based Models in Neonatal Sepsis: A Systematic Review. *Pediatr Infect Dis J.* 2024. doi: 10.1097/INF.0000000000004409

**Performances des modèles sont globalement élevées, mais extrêmement hétérogènes
= les modèles privilégient la discrimination, pas l'aide à la décision**

Résultats (exemple des du sepsis néonatal)



Performances de la méta-analyse (n=19 études) (15,984 participants, 76 algorithmes de ML)

- AUC : 0,92-0,94
- Sensibilité : 0,87
- Spécificité : 0,89

$I^2 = 95,5$: performances élevées MAIS très variables

Un bon ROC ne garantit pas une bonne valeur prédictive positive !

Kainth et al. Diagnostic Performance of Machine Learning-based Models in Neonatal Sepsis: A Systematic Review. *Pediatr Infect Dis J.* 2024. doi: 10.1097/INF.0000000000004409

**Performances des modèles sont globalement élevées, mais extrêmement hétérogènes
= les modèles privilégient la discrimination, pas l'aide à la décision**

Sur le terrain !

Vous êtes infectiologue dans un hôpital. On vous propose un nouvel outil d'IA pour la surveillance pour prédire un sepsis. C'est un excellent outil, son AUC-ROC est de 0,86. Sa sensibilité est de 0,83 et sa spécificité est de 0,92.

La prévalence des sepsis dans votre hôpital est de 10 % (réanimation néonatale) calculer la VPP et la VPN sur 1000 patient hospitalisé ?

	Sepsis réel	Pas de sepsis	Total
Alerte IA +	83 (TP)	72 (FP)	155
Alerte IA -	17 (FN)	828 (TN)	845
Total	100	900	1000

$$VPP = \frac{\text{Vrai positif}}{\text{Vrai positif} + \text{Faux négatif}} = \frac{83}{155}$$

54%

$$VPN = \frac{\text{Vrai négatif}}{\text{Vrai négatif} + \text{Faux positif}} = \frac{828}{845}$$

98%

**AUC haute mais le système génère 1 alerte fautive sur 2 ⇒ réduction de la charge de travail ?
Tout dépend de la prévalence ! L'IA exige des solutions locales**

Conclusion Générale (1)

- **Aujourd'hui : preuves surtout techniques (Se, Spé, AUC)**
- **Ce qu'on attend :**
 - ↓ délai diagnostic
 - ↓ antibiothérapie inutile
 - ↓ mortalité (non démontré)
 - ↓ charge de travail
- **Très peu d'études interventionnelles**

1

La technologie est prête même si design lourd

2

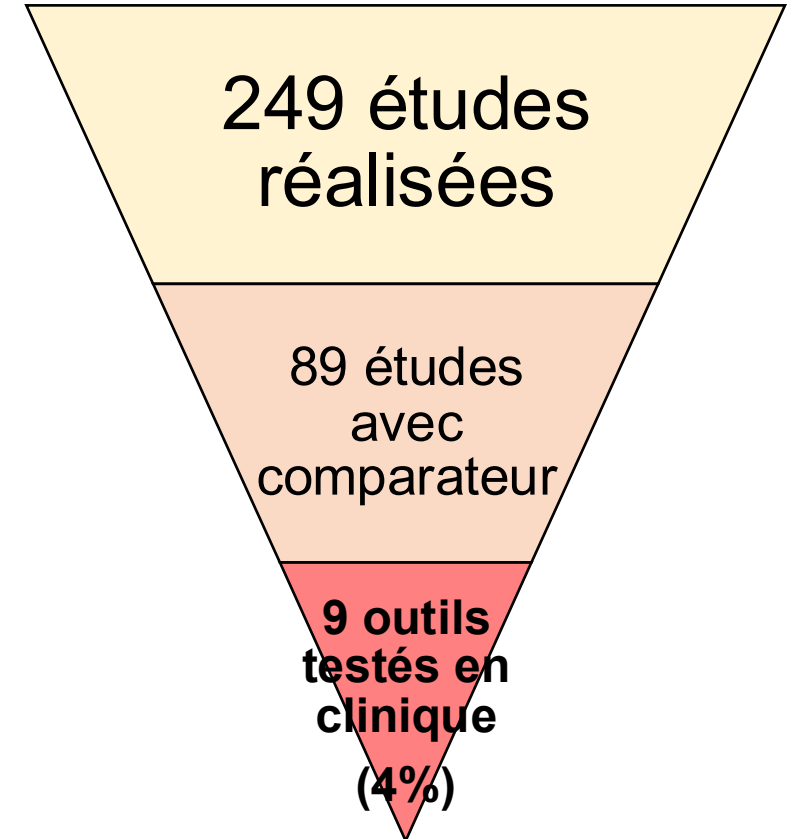
Le défi = implémentation + certification + adhésion

3

Validation multicentrique et calibrage indispensable

4

L'outil ne remplace pas l'expert, ne décide pas = aide à la priorisation



12% d'étude avec validation externe
12% intégré dans un logiciel

Conclusion générale (2)

1 Biais de sélection des populations :

- monocentrique
- services à haut risque
- patients déjà très sévères

Généralisation faible :

performances sur jeux de données externe ?

3 Biais de métriques :

- Surreprésentation des métriques globales (AUC)
- VPP rarement rapportée

Décorrélacion entre performance statistique et utilité clinique

2 Biais de temporalité :

- données rétrospectives
- performances surestimées par rapport au temps réel
- reproductibilité des codes

Surperformance artificielle

4 Biais de comparateur faible :

- souvent comparé à : règles simples
- absence de comparaison humaine
- rarement comparé à : clinicien expert ou workflow réel

L'IA "gagne", mais contre un adversaire faible

Cap de l'implémentation ?

Quel impact ?

Réduction de mortalité ?

Réduction de charge de travail ?

Des performances prometteuses, mais globalement obtenues dans des conditions éloignées de la pratique réelle

Merci !

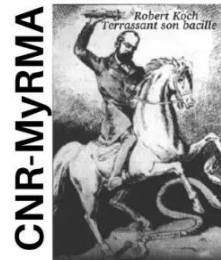
INTELLIGENCE ARTIFICIELLE ET MALADIES INFECTIEUSES PEDIATRIQUES

GPIP : 41ème JOURNEE ANNUELLE - 27 MARS 2026

Dr Alexandre Godmer, AHU, PharmD, PhD

alexandre.godmer@aphp.fr

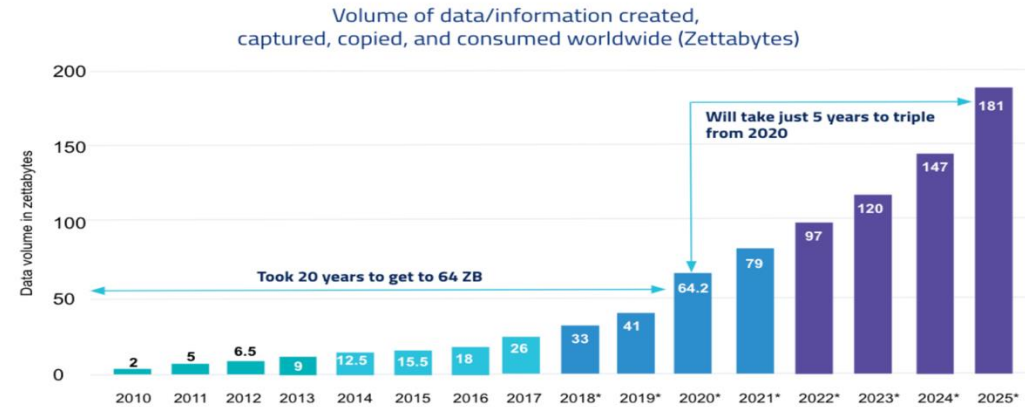
Hôpital Saint-Antoine (AP-HP, Sorbonne-Université)



L'ère du Big data (1)

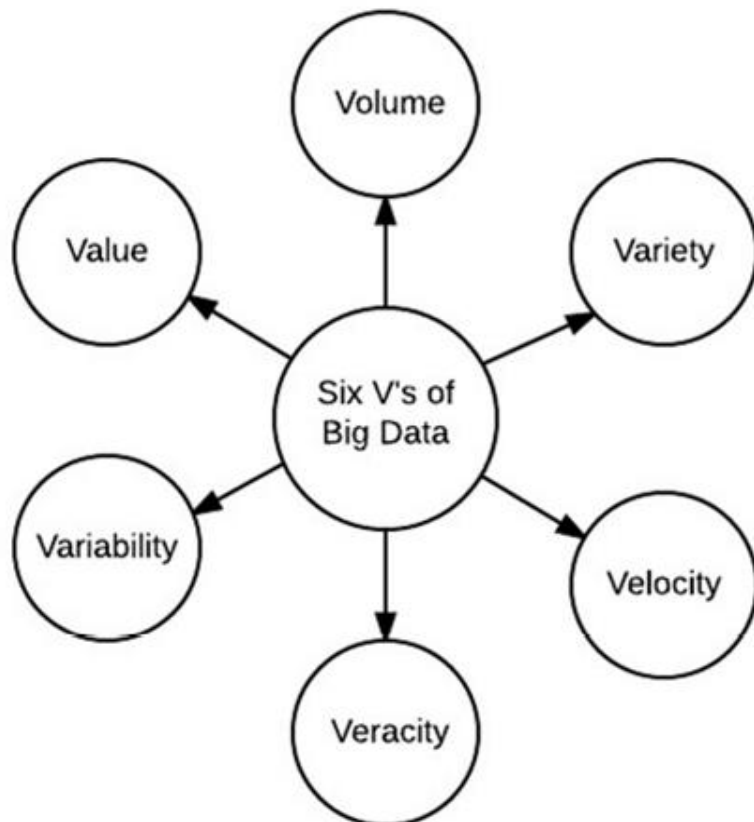
1

Big Data Grows Ever Bigger



35%
of all data will be life sciences + healthcare by 2025

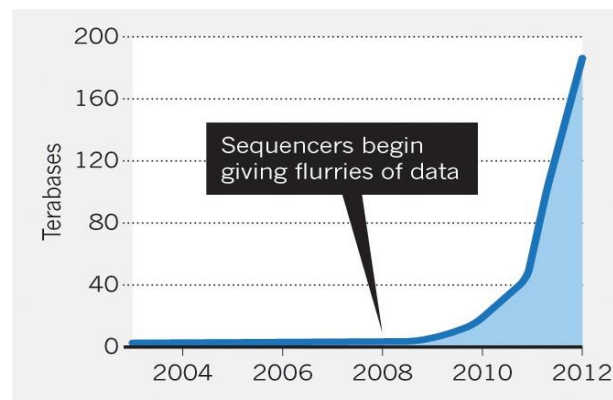
Big Data: Données massives et hétérogènes



Big Data in Healthcare Management: A Review of Literature

2

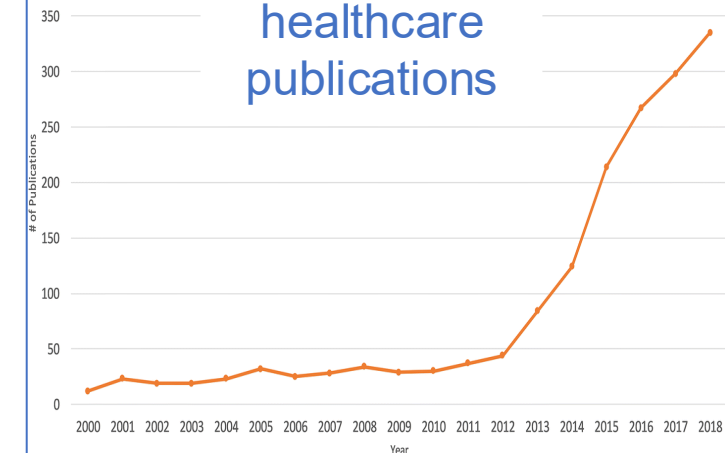
Sequencing data



Max V, *Nature* (2013), DOI: 10.1038/498255a

3

Big data healthcare publications

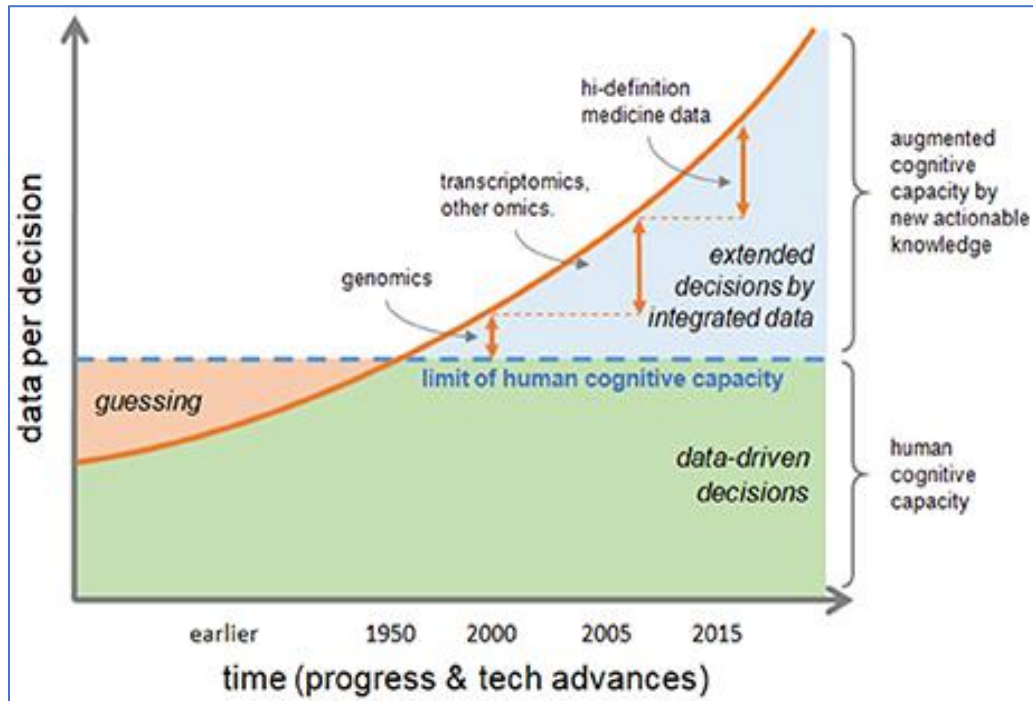


Big data in healthcare: management, analysis and future prospects

La production de données médicales augmente plus rapidement que notre capacité à les stocker, les traiter et les interpréter.

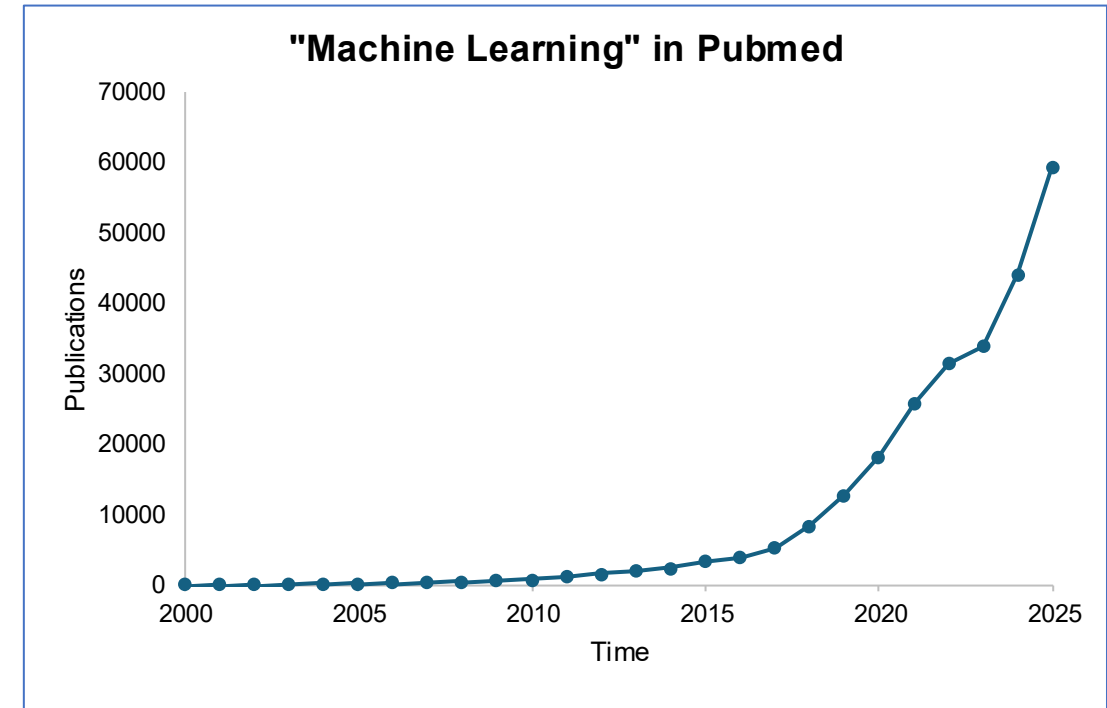
L'ère du Big data (2)

Analyses avancées essentielles pour la recherche translationnelle, clinique et fondamentale



Rossi & Grifantini, Frontiers in Digital Humanities (2018), DOI: 10.3389/fdigh.2018.00013

Démocratisation de l'utilisation des algorithmes d'intelligence artificielle

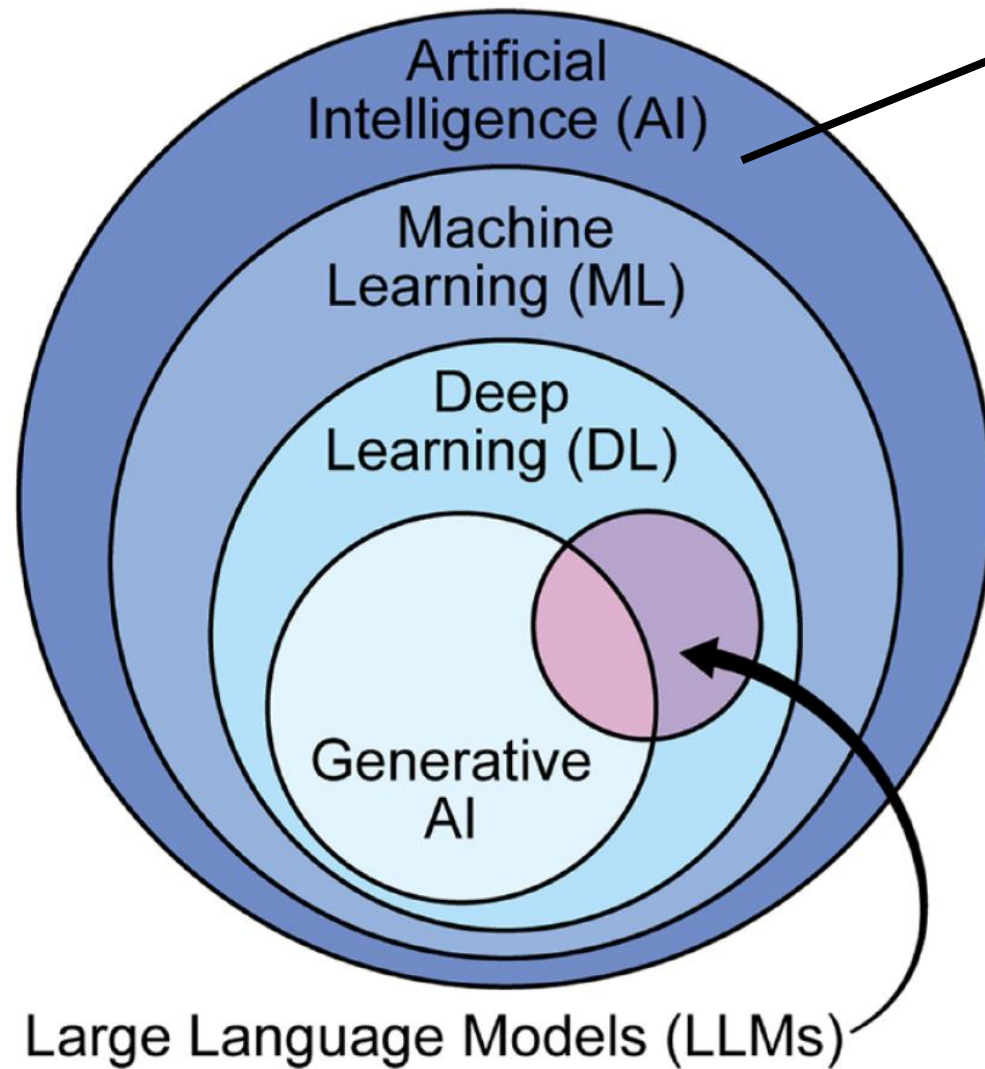


La recherche moderne passe de l'intuition humaine à la connaissance assistée par des algorithmes, car le volume et la complexité des données dépassent rapidement notre capacité à les interpréter

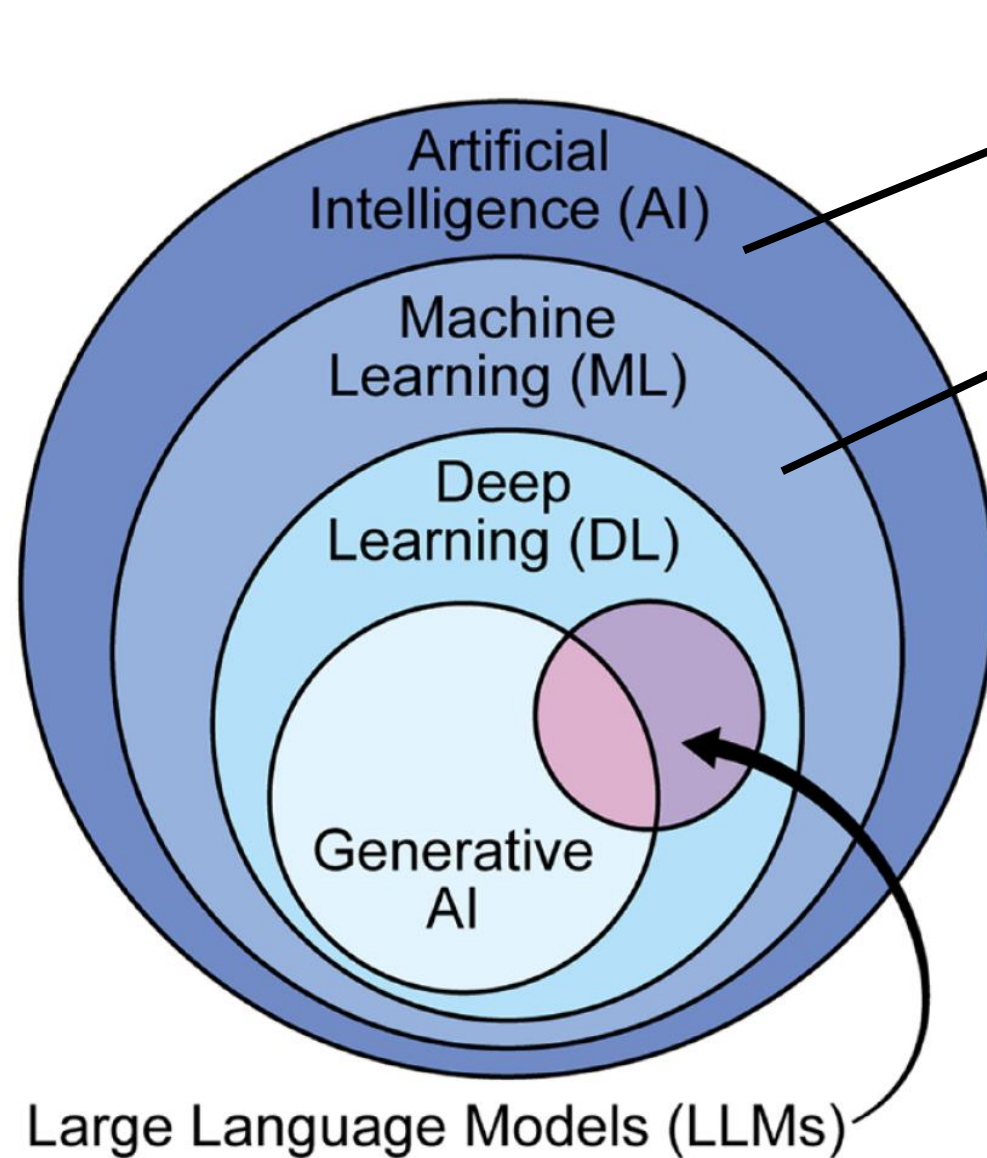
Question Wooclap :
Que vous évoque le mot « intelligence artificielle » ?

Artificial intelligence and its domains (1)

Using computers to perform tasks similar to those performed by humans, such as learning and problem solving



Artificial intelligence and its domains (3)

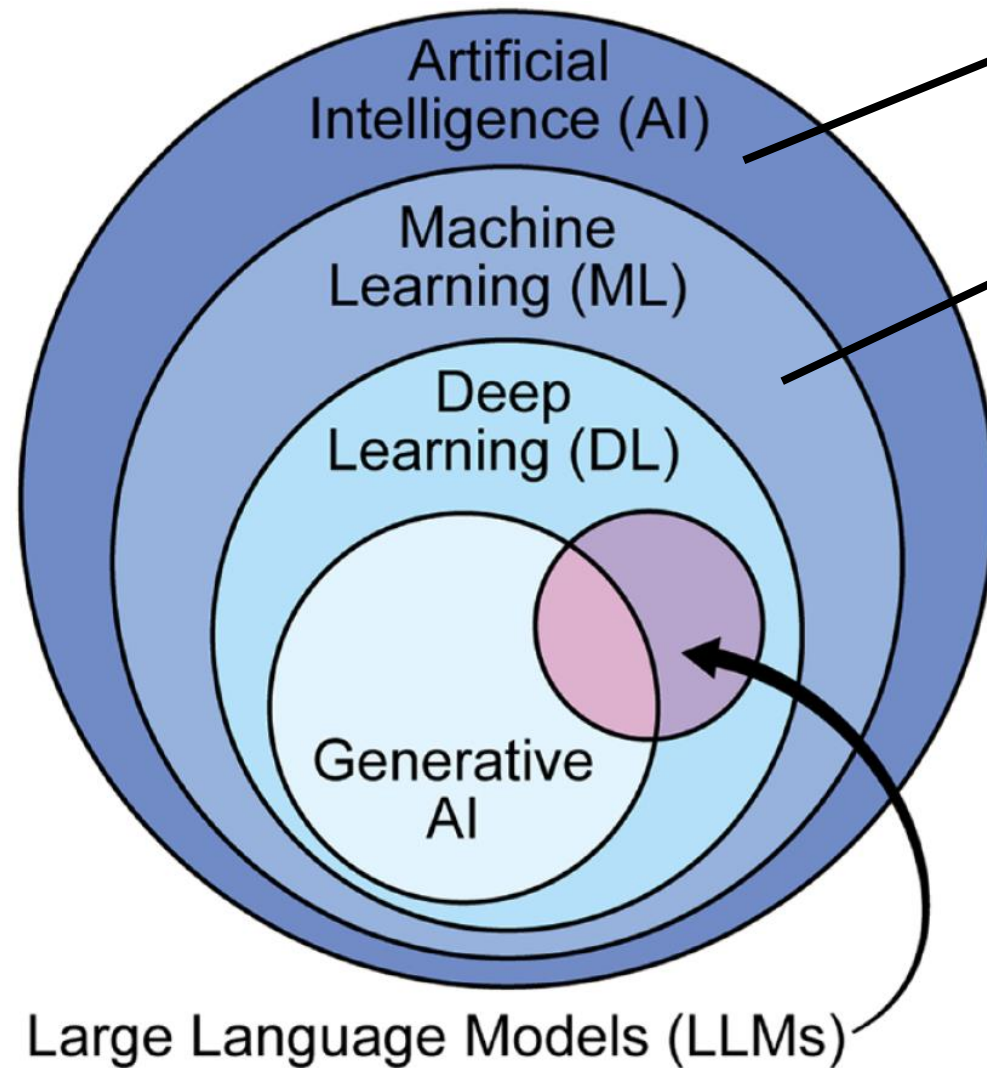


Using computers to perform tasks similar to those performed by humans, such as learning and problem solving

« Field of study that gives computers the ability to learn without being explicitly programmed. »

Arthur L. Samuel, 1959

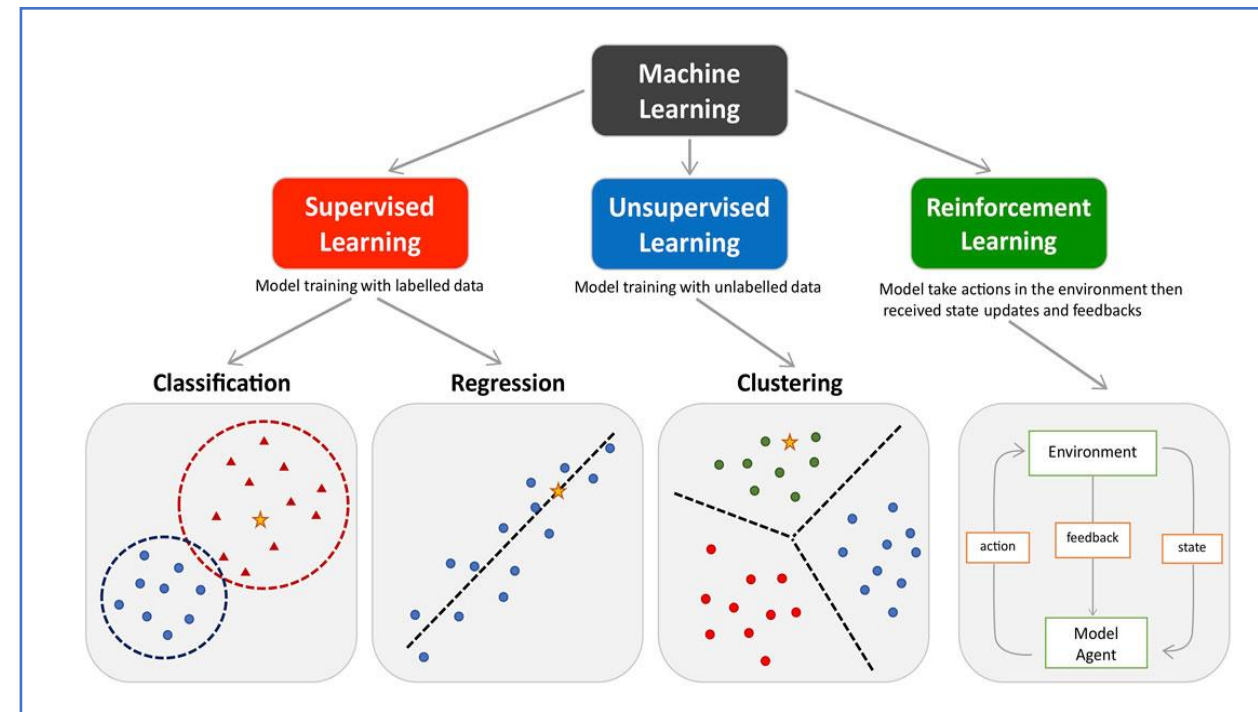
Artificial intelligence and its domains (4)



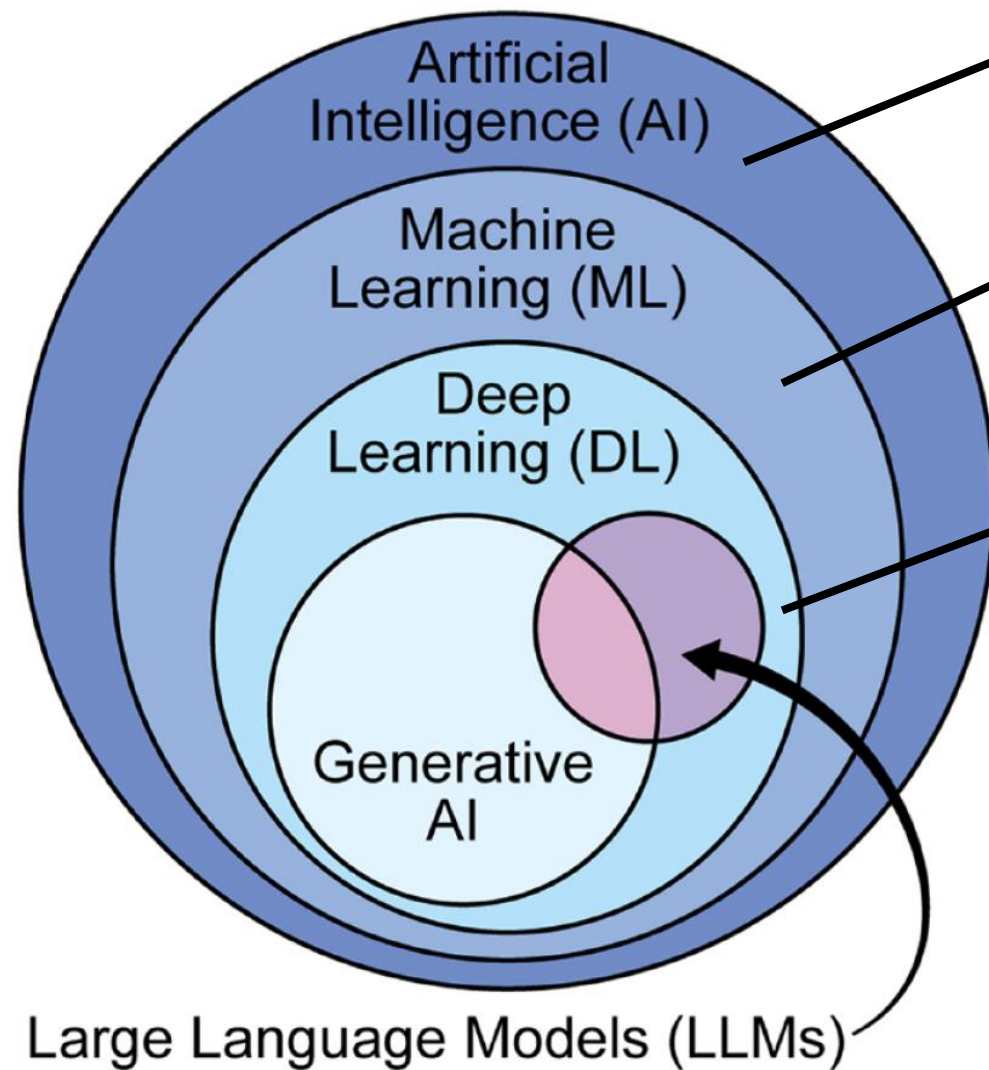
Using computers to perform tasks similar to those performed by humans, such as learning and problem solving

« Field of study that gives computers the ability to learn without being explicitly programmed. »

Arthur L. Samuel, 1959



Artificial intelligence and its domains (5)

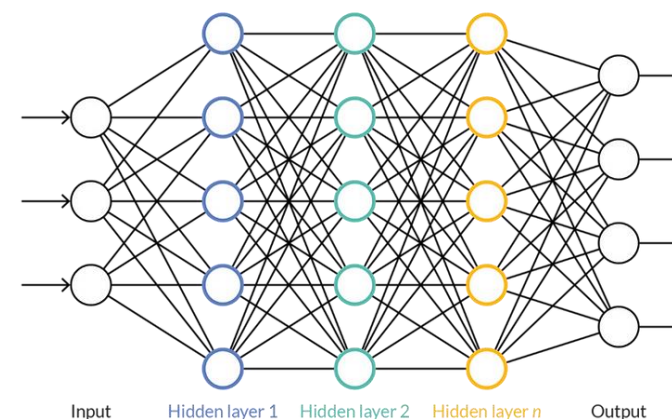


Using computers to perform tasks similar to those performed by humans, such as learning and problem solving

« Field of study that gives computers the ability to learn without being explicitly programmed. »

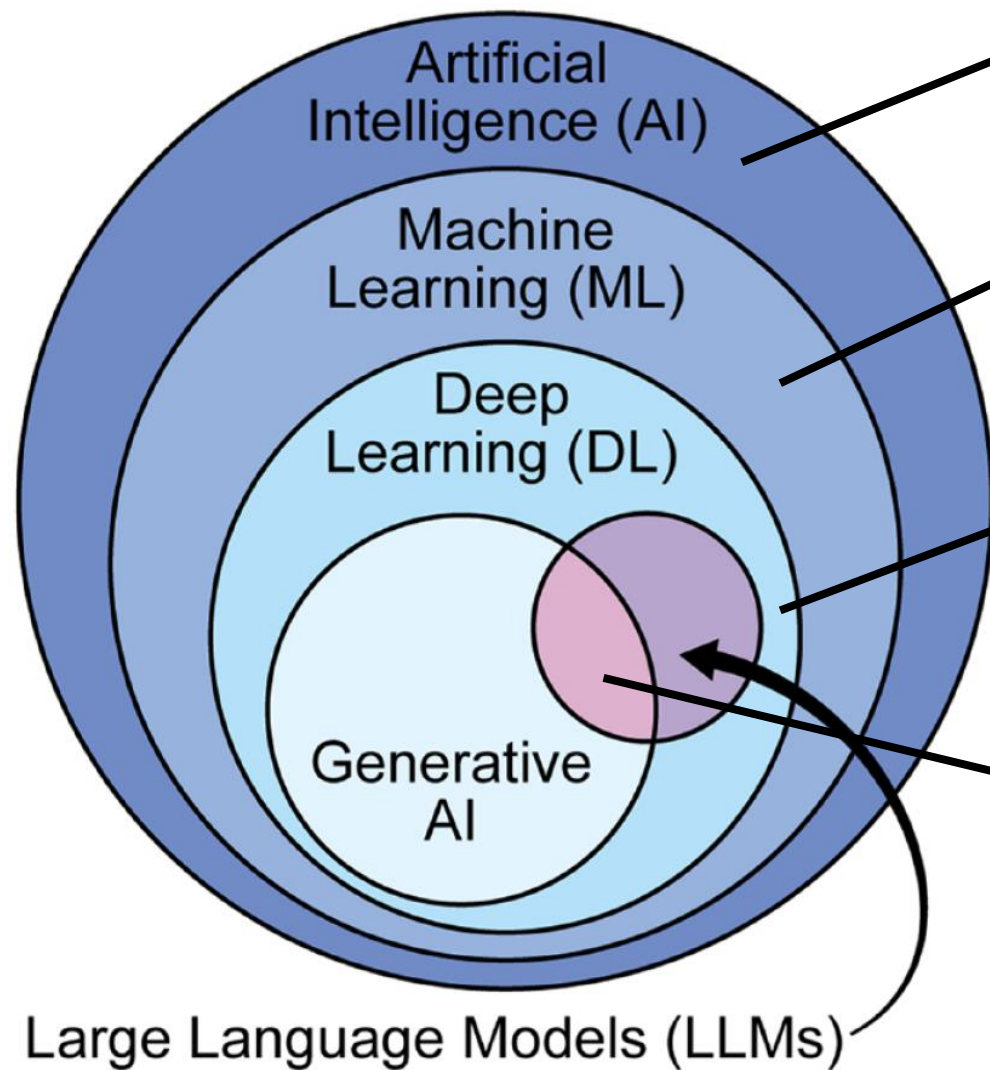
Arthur L. Samuel, 1959

Extension of machine learning that uses an artificial neural network



○ = artificial neuron

Artificial intelligence and its domains (6)



Using computers to perform tasks similar to those performed by humans, such as learning and problem solving

« Field of study that gives computers the ability to learn without being explicitly programmed. »

Arthur L. Samuel, 1959

Extension of machine learning that uses an artificial neural network

Model based on neural networks that is trained on massive text datasets to understand and generate natural language



ChatGPT



Claude

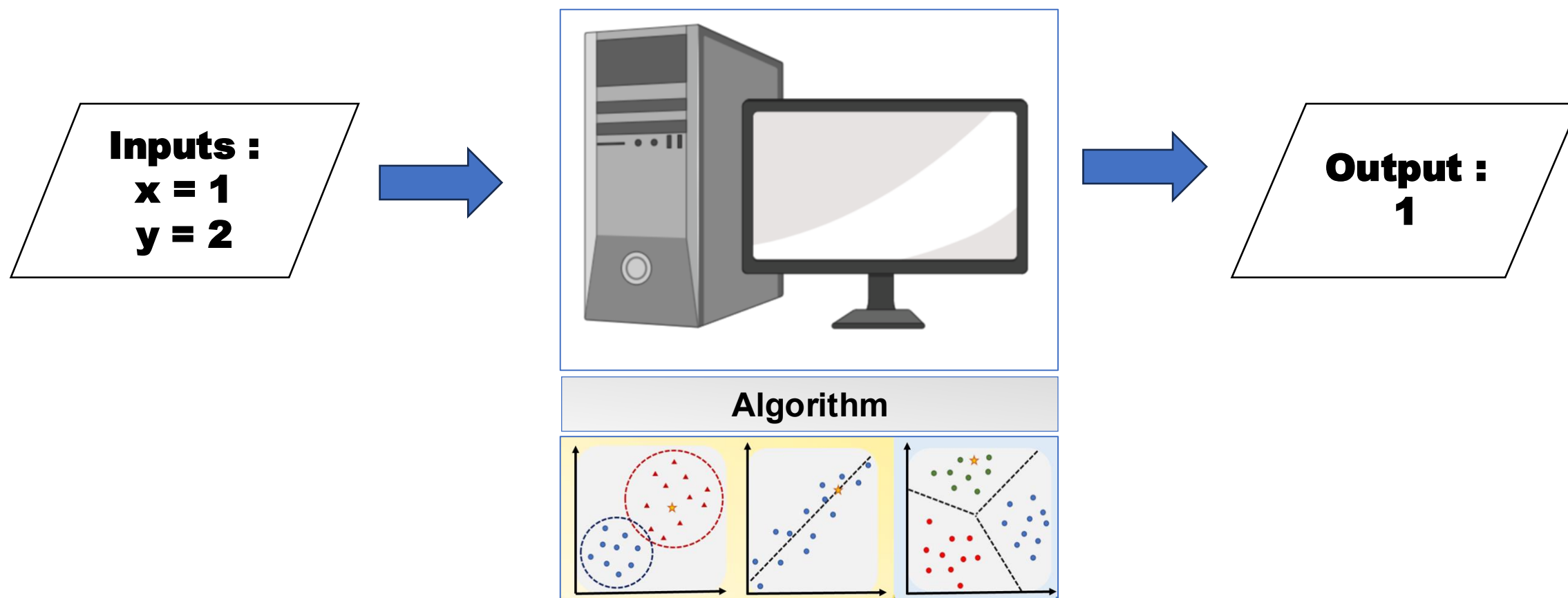


Gemini



MISTRAL
AI_

Models or algorithms of Machine Learning (1)

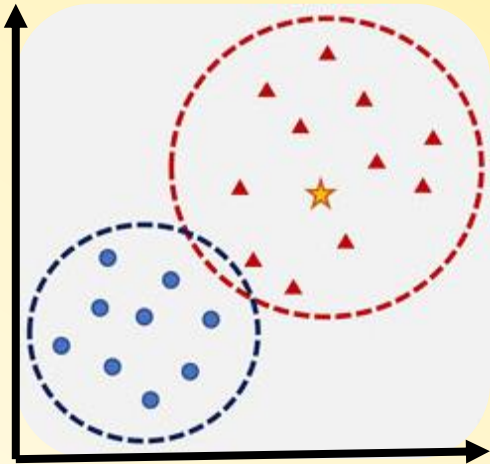


« A set of operational rules whose application allows a stated problem to be solved using a finite number of operations. An algorithm can be translated, using a programming language, into a program that can be executed by a computer.»,

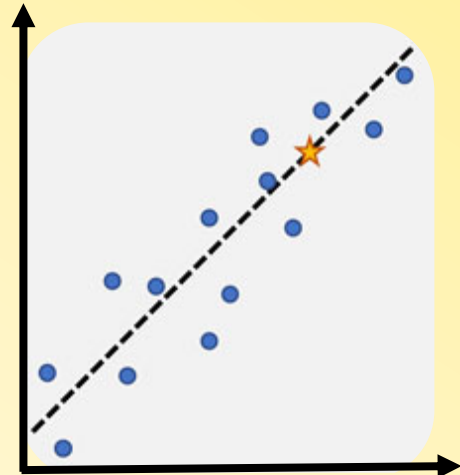
Larousse, 2020

Models or algorithms of Machine Learning (2) an overview

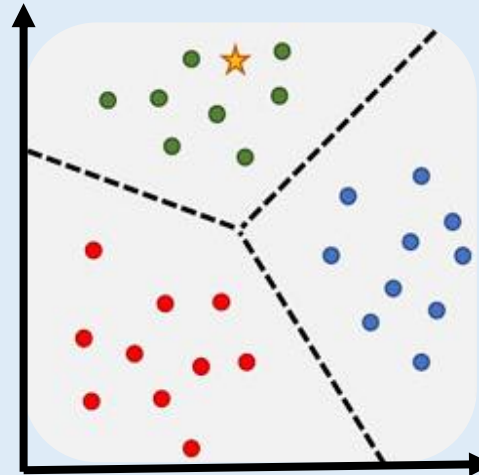
Classification:
Variable discrète



Regression:
Variable continue



Clustering:
Variable discrète



Supervised learning:

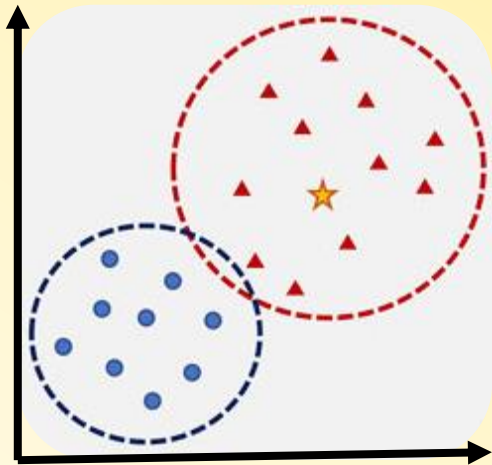
- learning from labeled examples
- example: neural network, support vector machine (SVM)

Unsupervised learning:

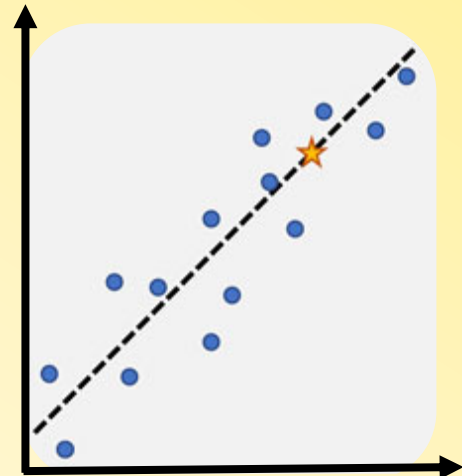
- unlabeled data recognition of structures specific to the data
- example: principal component analysis (PCA), clustering

Models or algorithms of Machine Learning (2) an overview

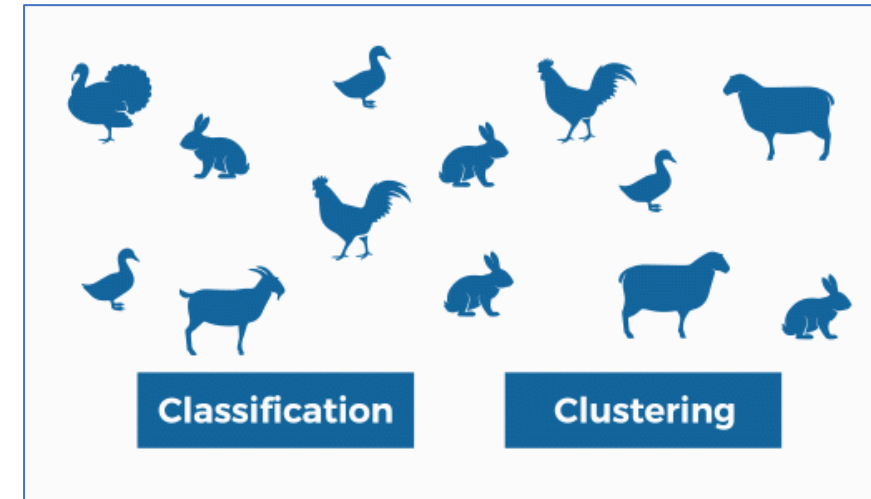
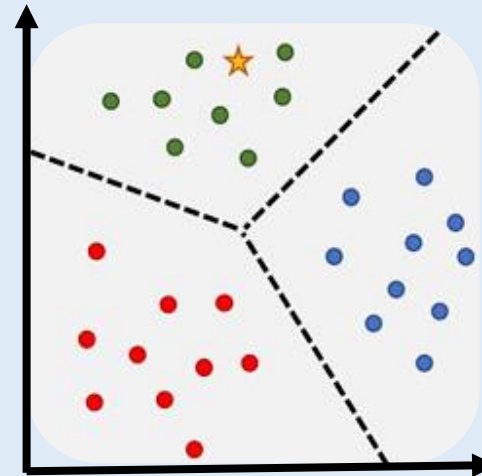
Classification:
Discrete variable



Regression:
Continuous variable



Clustering:
Discrete variable



Supervised learning:

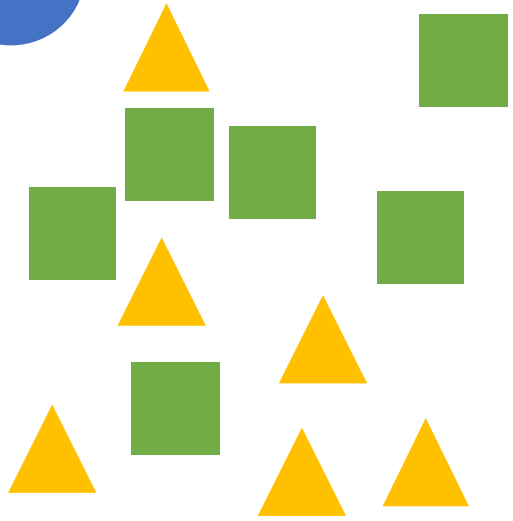
- learning from labeled examples
- example: neural network, support vector machine (SVM)

Unsupervised learning:

- unlabeled data recognition of structures specific to the data
- example: principal component analysis (PCA), clustering

Les algorithmes (modèles) de *Machine Learning*

80%

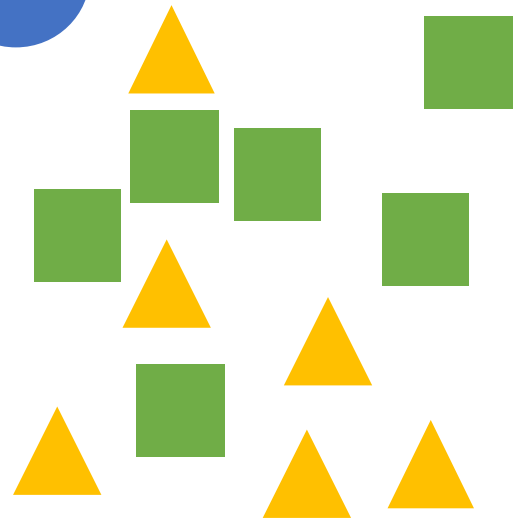


Données labellisées

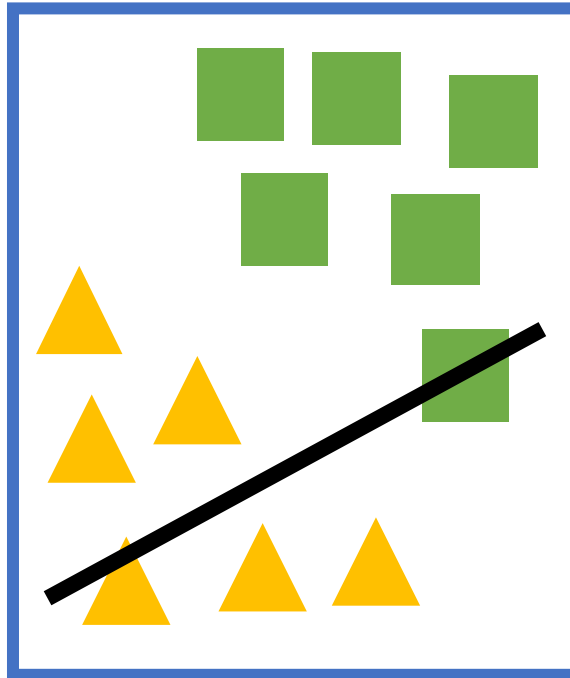
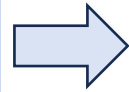
 Jeux d'entraînement

Les algorithmes (modèles) de *Machine Learning*

80%



Données labellisées

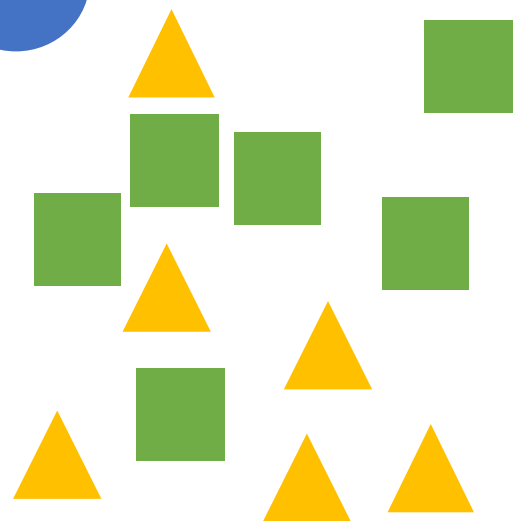


Entraînement

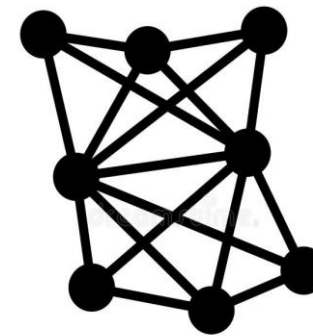
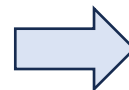
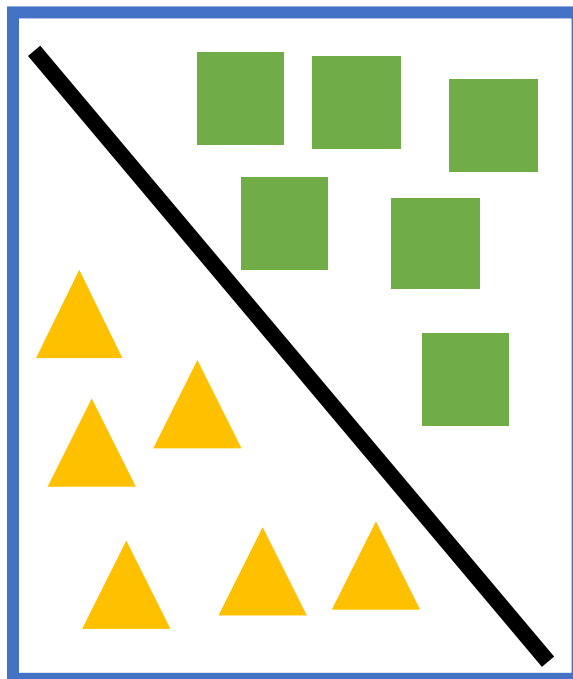
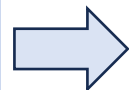
Jeux d'entraînement

Les algorithmes (modèles) de *Machine Learning*

80%



Données labellisées



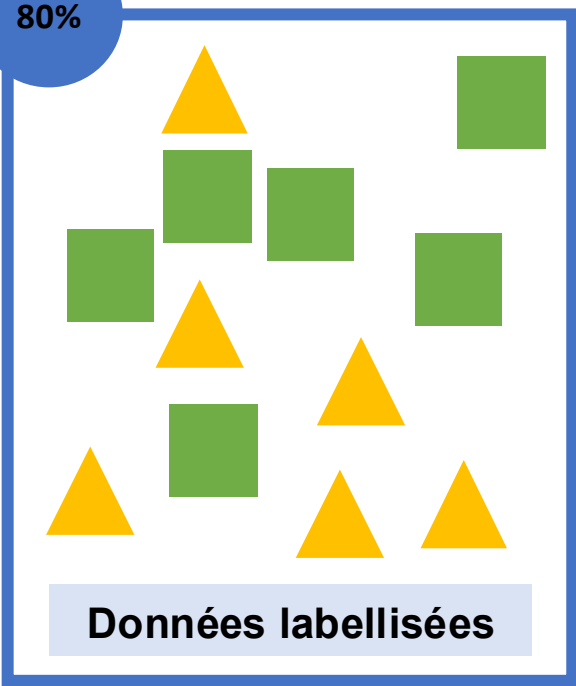
 Jeux d'entraînement

 Entraînement

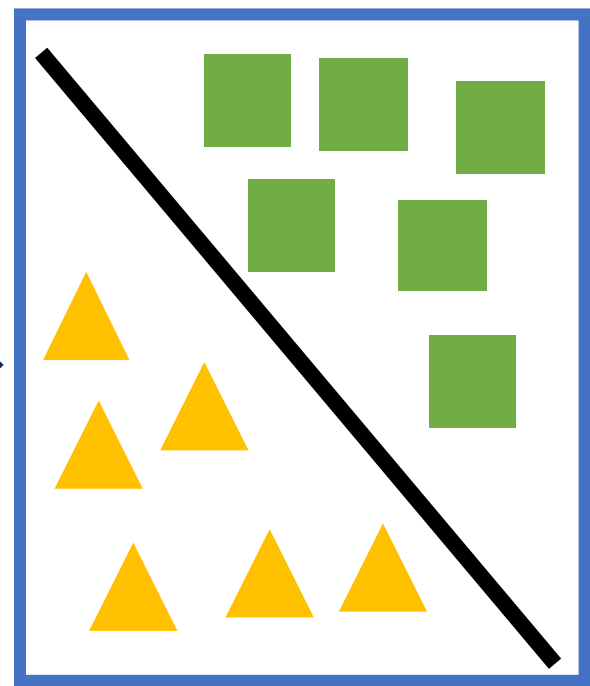
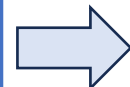
Modèle

Les algorithmes (modèles) de *Machine Learning*

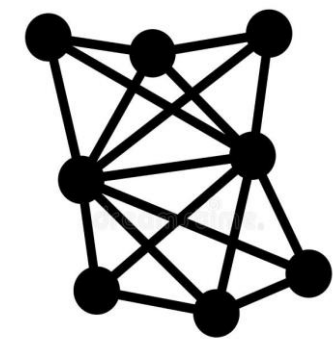
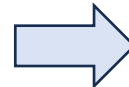
80%



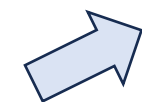
Jeux d'entraînement



Entraînement

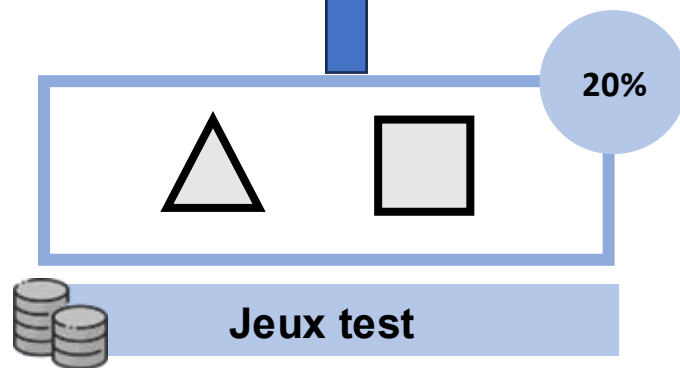


Modèle

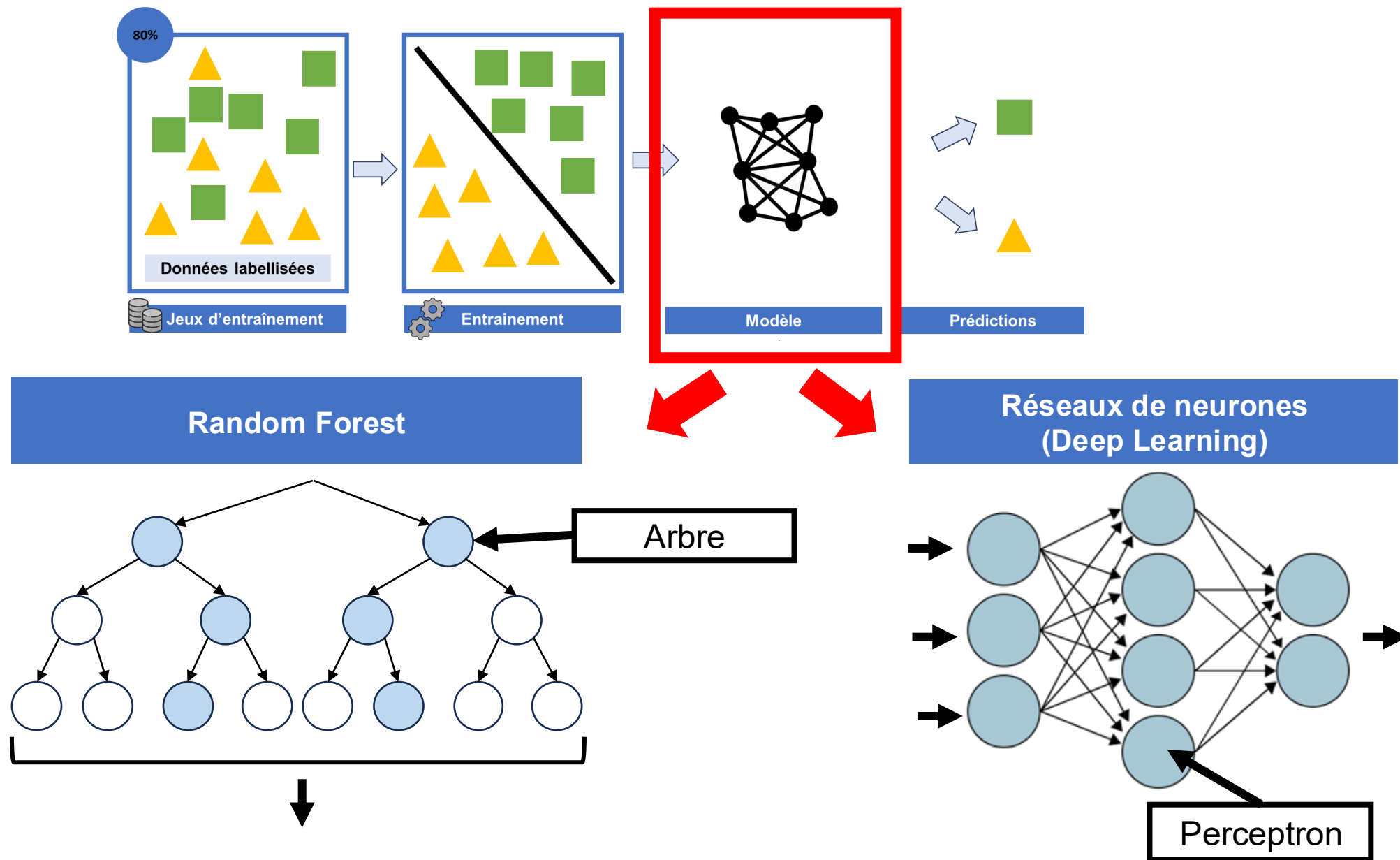


Prédictions

20%



Les algorithmes (modèles) de *Machine Learning*



Question Wooclap :
Selon vous quel est l'intérêt de ces outils pour votre pratique quotidienne ?

La promesse numérique (exemple des IAS)

Le fardeau actuel:

- **Prévalence élevée :**
 - 7,1 % des patients hospitalisés en Europe
 - 4,3 millions de cas/an (ECDC)
- **Impact économique majeur :**
 - Etats-Unis : 7,2 à 14,9 milliards \$/an
 - Europe : ≈ 7 milliards € / an
- **Charge humaine :**
 - Surveillance manuelle chronophage
 - Forte variabilité inter-observateur

Sous-détection structurelle des IAS

La promesse numérique (exemple des IAS)

Le fardeau actuel:

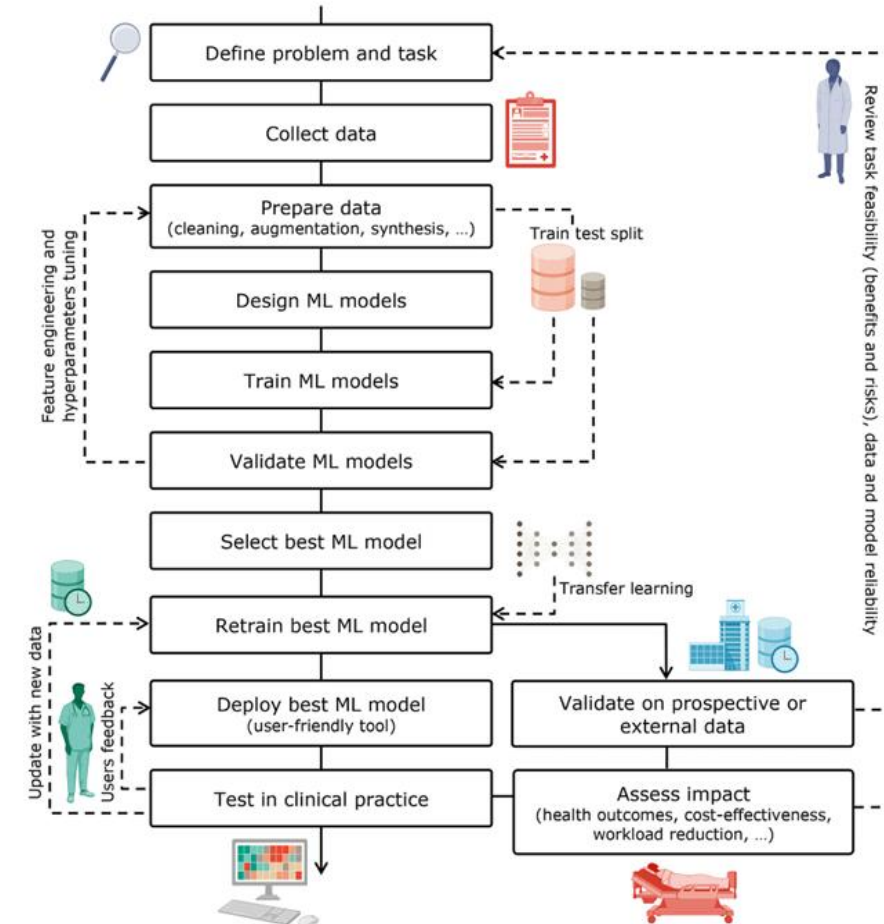
- **Prévalence élevée :**
 - 7,1 % des patients hospitalisés en Europe
 - 4,3 millions de cas/an (ECDC)
- **Impact économique majeur :**
 - Etats-Unis : 7,2 à 14,9 milliards \$/an
 - Europe : ≈ 7 milliards € / an
- **Charge humaine :**
 - Surveillance manuelle chronophage
 - Forte variabilité inter-observateur

Sous-détection structurelle des IAS

La promesse:

- Digitalisation
- Automatisation = surveillance continue
- Précision = détection de signaux faibles

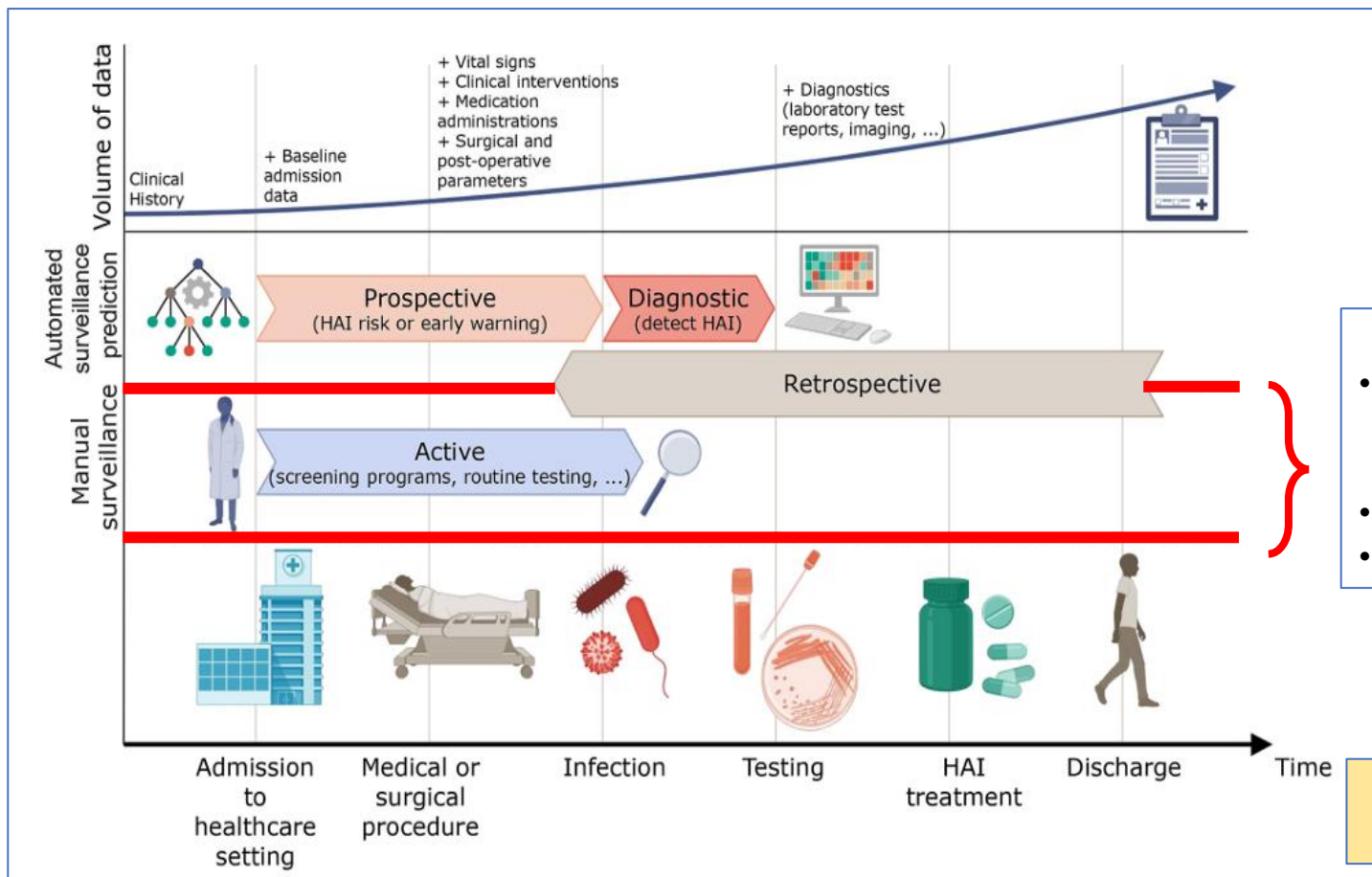
Flux de travail classique:



Cozzolino C *et al.*. Are AI-based surveillance systems for healthcare-associated infections ready for clinical practice? A systematic review and meta-analysis. *Artif Intell Med.* 2025 Jul;165:103137. doi: 10.1016/j.artmed.2025.103137. Epub 2025 Apr 22. PMID: 40286586.

Les IAS sont fréquentes, coûteuses
Surveillance manuelle = biais systémique de sous-détection

La promesse numérique (exemple des IAS)



Est-ce qu'il y a eu une infection ?

- Surveillance manuelle :
 - rétrospective
 - focalisée sur les services à haut risque
- Décalage temporel : infection détectée après les signaux
- Perte d'opportunité : diagnostic tardif

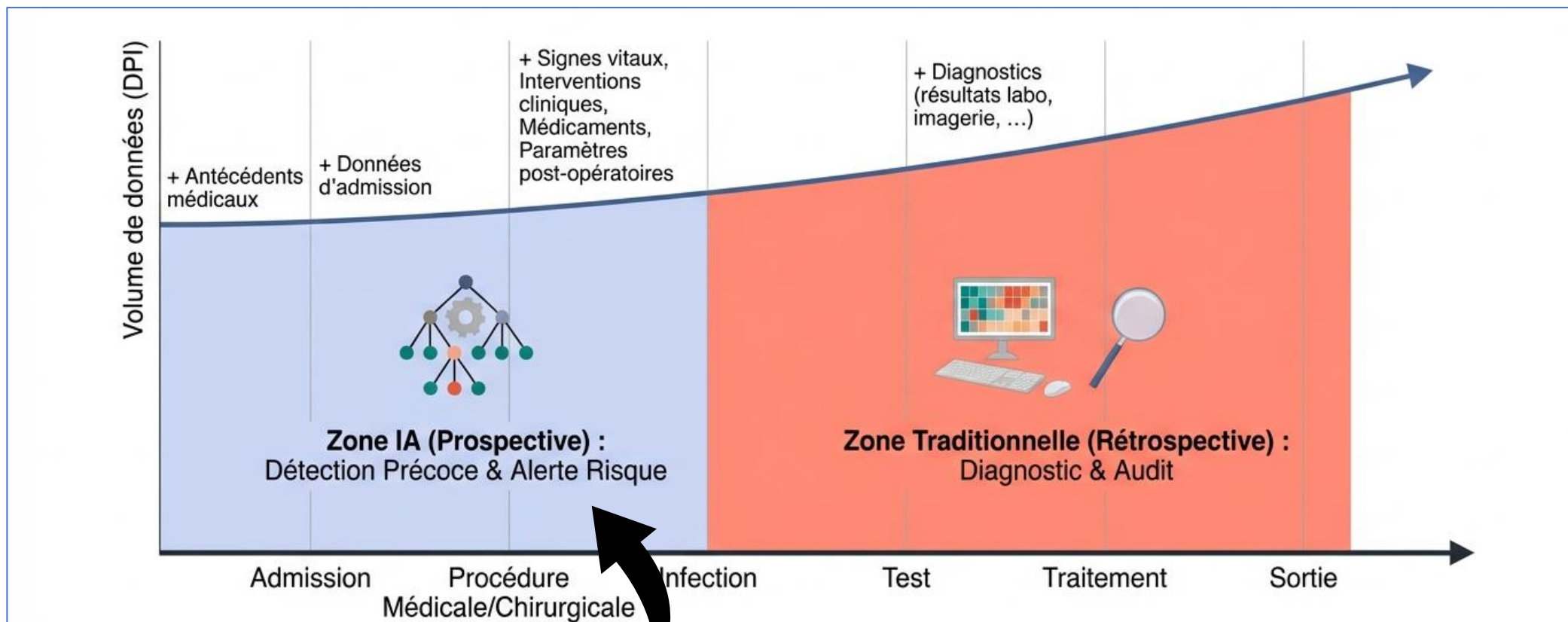


**Ne répond pas à la question :
Qui va faire une infection ?**

Cozzolino C *et al.*. Are AI-based surveillance systems for healthcare-associated infections ready for clinical practice? A systematic review and meta-analysis. *Artif Intell Med.* 2025 Jul;165:103137. doi: 10.1016/j.artmed.2025.103137. Epub 2025 Apr 22. PMID: 40286586.

La surveillance actuelle est tardive par construction = prévention impossible
Un système conçu pour détecter des événements passés ne peut pas prévenir des événements futurs

Ce que promet l'IA



Surveillance « hospital-wide »

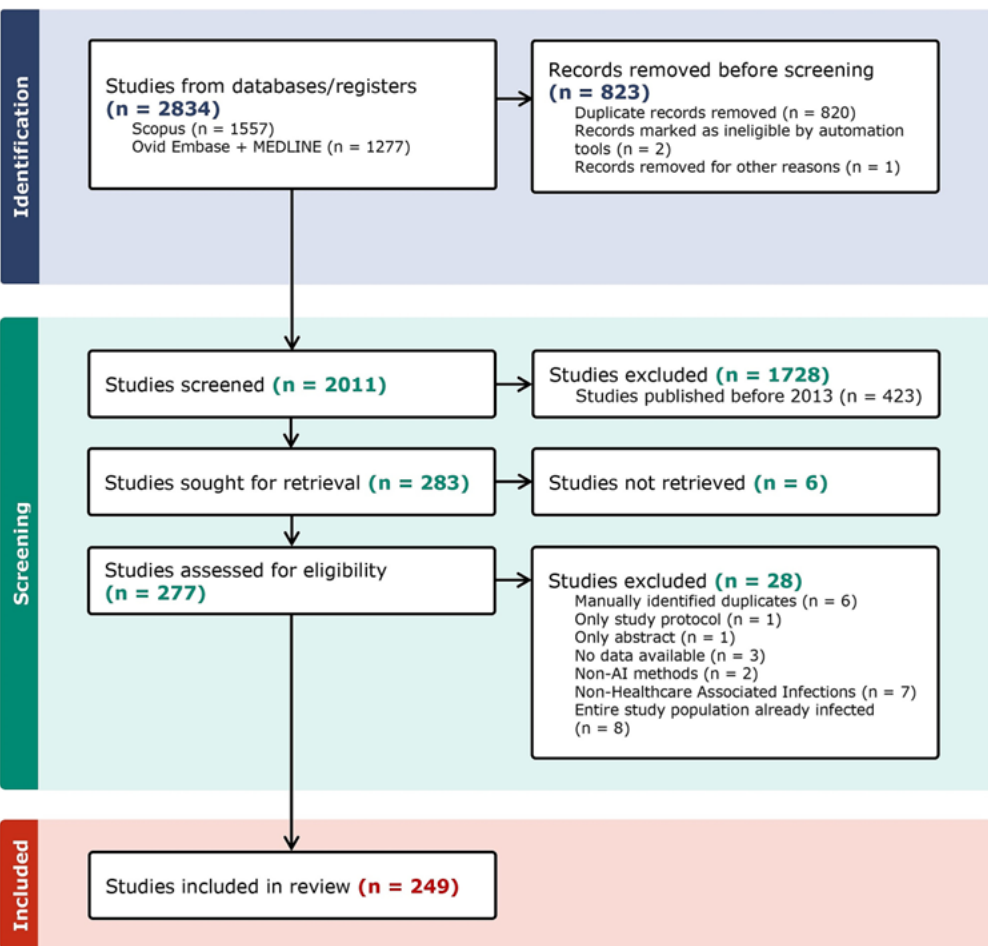
- 100 % des patients
- Automatisation continue 24h/24, sans fatigue humaine
- Standardisation et réduction de la variabilité inter-équipes
- Analyse multivariée : signes vitaux + biologie + actes + traitements
- **Objectif: détection précoce + alerte de risque prospective**

Changement de paradigme = du rétrospectif au prédictif

Les systèmes d'IA sont-ils prêts ?

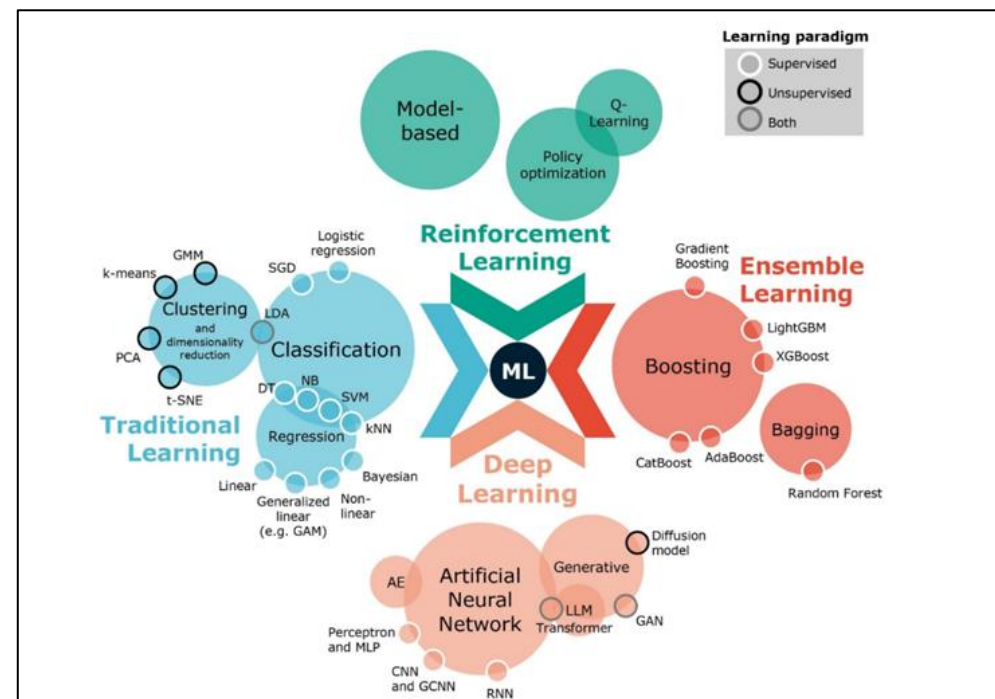
Méthode :

- Période : 2013-2023
- Bases : MEDLINE, Embase, données EHR



Résultats:

- Nombreuses méthodes de ML



- Performances toutes IAS confondues :

- **AUC ROC globale : 0,864 (IC 95 % : 0,852 – 0,877)**
- Sensibilité : 0,835 (IC 95 % : 0,803 – 0,862)
- Spécificité : 0,899 (IC 95 % : 0,869 – 0,922)
- Accuracy : 0,880 (IC 95 % : 0,858 – 0,898)

Cozzolino C *et al.*. Are AI-based surveillance systems for healthcare-associated infections ready for clinical practice? A systematic review and meta-analysis. *Artif Intell Med.* 2025 Jul;165:103137. doi: 10.1016/j.artmed.2025.103137. Epub 2025 Apr 22. PMID: 40286586.

Discrimination élevée des modèles d'IA pour les IAS, avec une AUC globale à 0,86

Les systèmes d'IA sont-ils prêts ?

Question Wooclap :

Vous êtes infectiologue dans un hôpital. On vous propose un nouvel outil d'IA pour la surveillance des IAS. C'est un excellent outil, son AUC-ROC est de 0,86.

Avec cette seule information (AUC = 0,86), que savez-vous réellement ?

- A. L'outil détectera correctement 86 % des infections
- B. L'outil classe globalement mieux les patients infectés que les non infectés
- C. L'outil permettra de réduire le nombre d'IAS
- D. L'outil est prêt à être utilisé en pratique clinique
- E. L'AUC ne mesure que la discrimination, pas la décision

Les systèmes d'IA sont-ils prêts ?

Question Wooclap :

Vous êtes infectiologue dans un hôpital. On vous propose un nouvel outil d'IA pour la surveillance des IAS. C'est un excellent outil, son AUC-ROC est de 0,86.

Avec cette seule information (AUC = 0,86), que savez-vous réellement ?

- A. L'outil détectera correctement 86 % des infections
- B. L'outil classe globalement mieux les patients infectés que les non infectés
- C. L'outil permettra de réduire le nombre d'IAS
- D. L'outil est prêt à être utilisé en pratique clinique
- E. L'AUC ne mesure que la discrimination, pas la décision

Une AUC élevée nous dit que le modèle classe bien, pas qu'il aide à décider

Comment évaluer les performances des systèmes d'IA ?

La matrice de confusion

		Méthode de référence	
		Positif	Négatif
Prédiction	Positif	Vrai positif	Faux positif
	Négatif	Faux négatif	Vrai négatif

$$\text{Accuracy} = \frac{\text{Vrai positif} + \text{Vrai négatif}}{\text{Vrai négatif} + \text{Faux positif} + \text{Vrai positif} + \text{Faux négatif}}$$

$$\text{Sensibilité} = \frac{\text{Vrai positif}}{\text{Vrai positif} + \text{Faux négatif}}$$

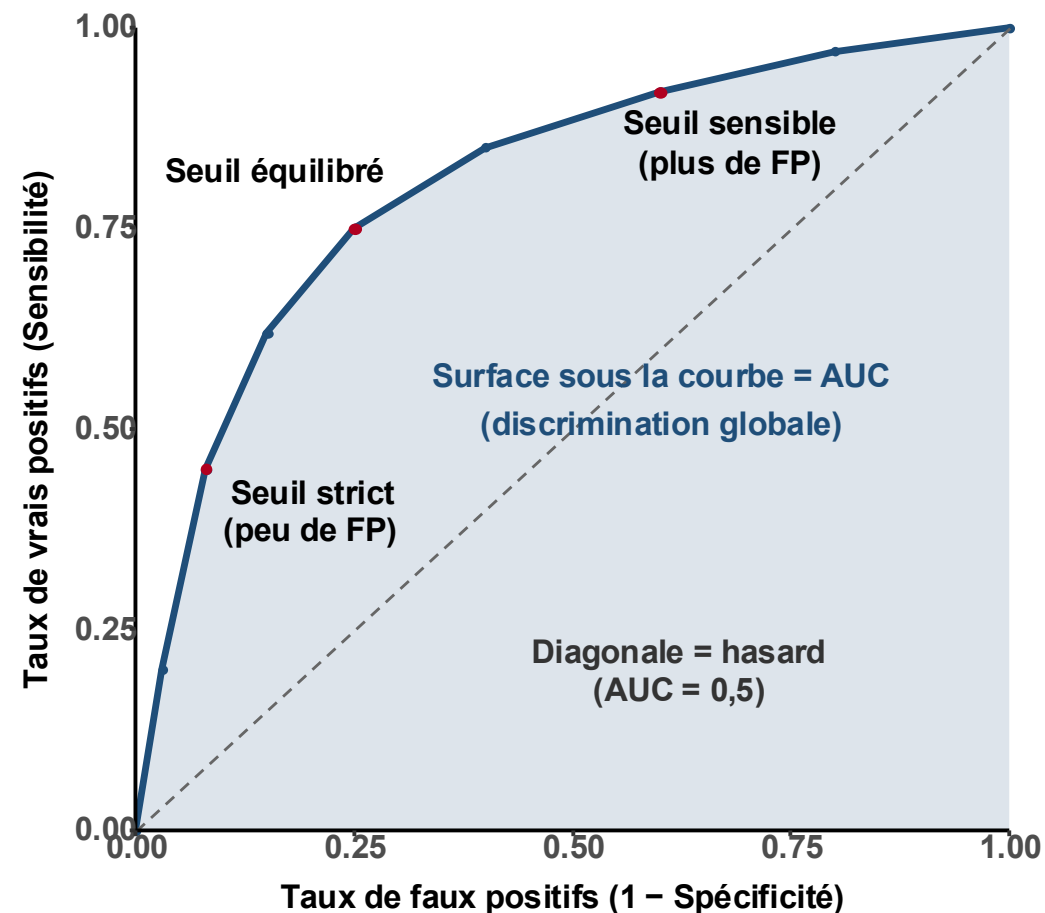
$$\text{Spécificité} = \frac{\text{Vrai négatif}}{\text{Vrai négatif} + \text{Faux positif}}$$

$$\text{VPP} = \frac{\text{Vrai positif}}{\text{Vrai positif} + \text{Faux négatif}}$$

$$\text{VPN} = \frac{\text{Vrai négatif}}{\text{Vrai négatif} + \text{Faux négatif}}$$

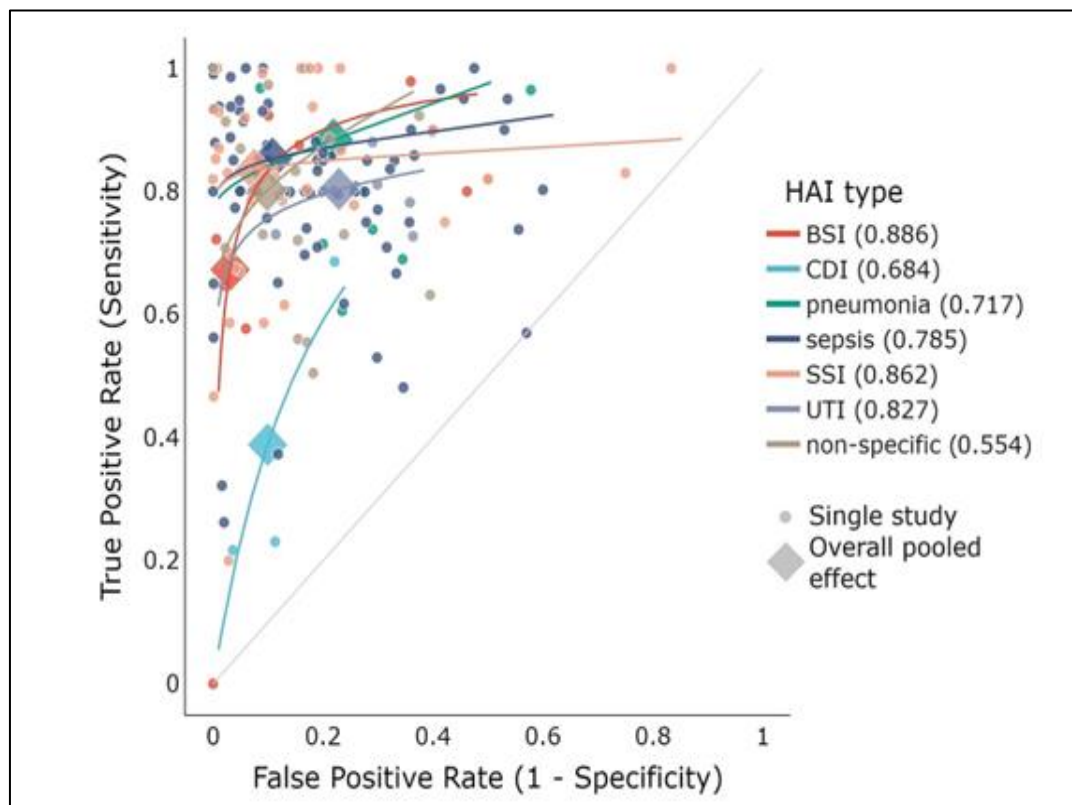
AUC-ROC : $P(\text{score}_{\text{malade}} > \text{score}_{\text{Nonmalade}})$

⇒ résume tous les compromis sensibilité / faux positifs



Différentes métriques = différentes informations

Performances détaillées



- Sepsis :
 - AUC : 0,872
 - Sensibilité : 0,853
 - Spécificité : 0,892
- Infection du site opératoire (SSI) :
 - AUC : 0,865
 - Sensibilité : 0,835
 - Spécificité : 0,925
- *Clostridioides difficile* (CDI) :
 - AUC : 0,806
 - Sensibilité : 0,389

Tendance générale à favoriser la spécificité au détriment de la sensibilité : le système rate la majorité des infections

Performances des modèles pour les IAS sont globalement élevées, mais extrêmement hétérogènes = les modèles privilégient la discrimination, pas l'aide à la décision

Question Wooclap :

Vous êtes infectiologue dans un hôpital. On vous propose un nouvel outil d'IA pour la surveillance des IAS. C'est un excellent outil, son AUC-ROC est de 0,86. Sa sensibilité est de 0,83 et sa spécificité est de 0,92.

Comment l'utiliseriez vous ?

Question Wooclap :

Vous êtes infectiologue dans un hôpital. On vous propose un nouvel outil d'IA pour la surveillance des IAS. C'est un excellent outil, son AUC-ROC est de 0,86. Sa sensibilité est de 0,83 et sa spécificité est de 0,92.

La prévalence des IAS dans votre hôpital est de 14%, calculer la VPP et la VPN sur 1000 patient hospitalisé ?

Question Wooclap :

Vous êtes infectiologue dans un hôpital. On vous propose un nouvel outil d'IA pour la surveillance des IAS. C'est un excellent outil, son AUC-ROC est de 0,86. Sa sensibilité est de 0,83 et sa spécificité est de 0,92.

La prévalence des IAS dans votre hôpital est de 14%, calculer la VPP et la VPN sur 1000 patient hospitalisé ?

	IAS réelle	Pas d'IAS	Total
Alerte IA +	117 (TP)	87 (FP)	204
Alerte IA -	23 (FN)	773 (TN)	796
Total	140	860	1000

$$VPP = \frac{\text{Vrai positif}}{\text{Vrai positif} + \text{Faux négatif}} = \frac{117}{204}$$

57%

$$VPN = \frac{\text{Vrai négatif}}{\text{Vrai négatif} + \text{Faux positif}} = \frac{773}{796}$$

97%

**AUC excellente = le système génère 1 alerte fautive sur 2 ⇒ réduction de la charge de travail ?
 ≠ d'un système d'alerte, excellente VPN ⇒ Capacité à rassurer ? Quid des 3% restant ?**

Revue de la littérature : étude des performances et limites majeures

1

Biais de sélection des populations :

- monocentrique
- services à haut risque
- patients déjà très sévères

Généralisation faible :

performances sur jeux de données externe ?

3

Biais de métriques :

- AUC rapportée dans ~80 %
- PPV rapportée dans seulement 36,5 %

Discrimination ≠ utilité

2

Biais de temporalité :

- 89,3 % rétrospectives
- 9,0 % prospectives
- 1,7 % mixtes

Surperformance artificielle

4

Biais de comparateur faible :

- souvent comparé à : règles simples, scores obsolètes
- absence de comparaison humaine
- rarement comparé à : clinicien expert ou workflow réel

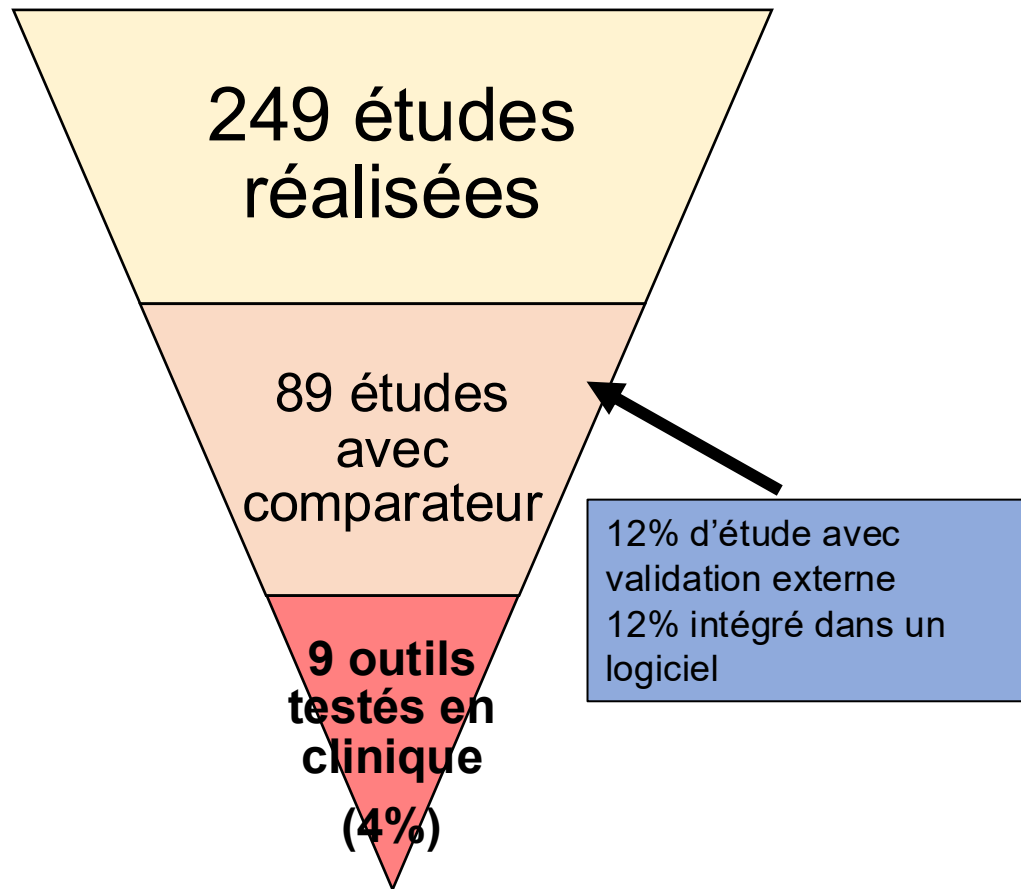
L'IA "gagne", mais contre un adversaire faible

Cap de l'implémentation ?
réduction des IAS ?
réduction de mortalité ?
réduction de charge de travail ?

Barbati C et al. Artificial intelligence use and performance in detecting and predicting healthcare-associated infections: A systematic review. *Artif Intell Med.* 2026 Feb;172:103321. doi: 10.1016/j.artmed.2025.103321. Epub 2025 Nov 27. PMID: 41338040.

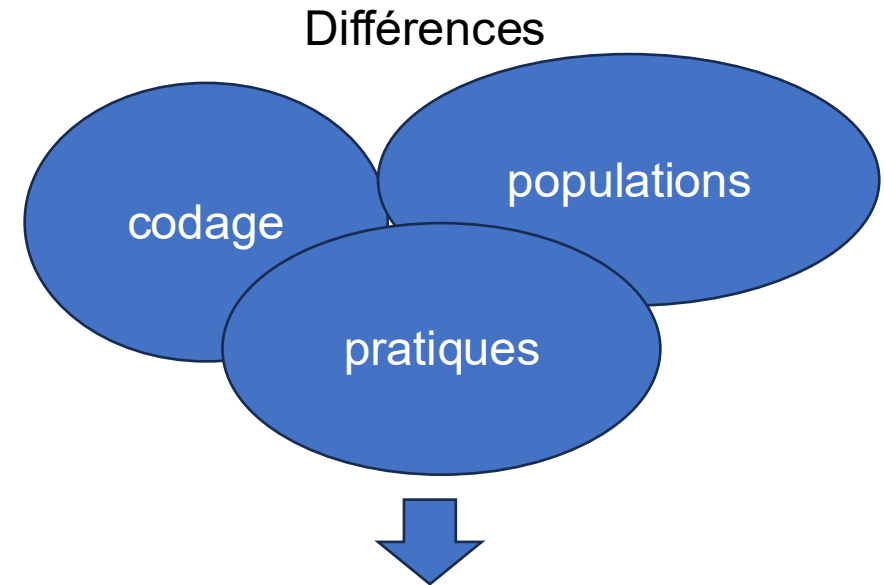
En résumé = optimisation des courbes, pas des soins : « À partir de quel risque dois-je agir ? »

« Reality Check » : le gouffre de l'implémentation

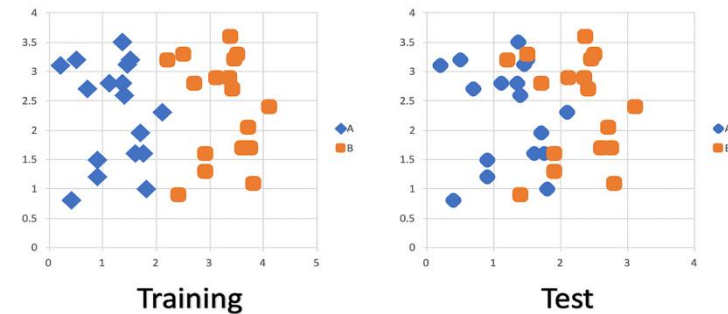


Barbati C et al. Artificial intelligence use and performance in detecting and predicting healthcare-associated infections: A systematic review. *Artif Intell Med.* 2026 Feb;172:103321. doi: 10.1016/j.artmed.2025.103321. Epub 2025 Nov 27. PMID: 41338040.

Cozzolino C *et al.*. Are AI-based surveillance systems for healthcare-associated infections ready for clinical practice? A systematic review and meta-analysis. *Artif Intell Med.* 2025 Jul;165:103137. doi: 10.1016/j.artmed.2025.103137. Epub 2025 Apr 22. PMID: 40286586.



Problématique :
« domain shift »



Majorité des modèles = prototypes « *in silico* »
Environ 96% des modèles ne sont jamais testés en clinique

Conclusion : revue de la littérature

L'IA ne remplace pas le clinicien mais agit comme un outil puissant pour améliorer la précision diagnostique et thérapeutique, à condition de surmonter les défis d'intégration et de biais

Exemple pratique d'un succès = PERISCOPE

Development and validation of artificial intelligence models for early detection of postoperative infections (PERISCOPE): a multicentre study using electronic health record data



Siri L. van der Meijden,^{a,b,*} Anna M. van Boekel,^a Laurens J. Schinkelshoek,^b Harry van Goor,^c Ewout W. Steyerberg,^d Rob G. H. H. Nelissen,^e Dieter Mesotten,^{f,g} Bart F. Geerts,^b Mark G. J. de Boer,^h and M. Sesmu Arbous,^a on behalf of the PERISCOPE Groupⁱ



^aIntensive Care Unit, Leiden University Medical Centre, Leiden, the Netherlands

^bHealthplus.ai B.V., Amsterdam, the Netherlands

^cGeneral Surgery Department, Radboud University Medical Centre, Nijmegen, the Netherlands

^dDepartment of Biomedical Data Sciences, Leiden University Medical Centre, Leiden, the Netherlands

^eDepartment of Orthopaedics, Leiden University Medical Centre, Leiden, the Netherlands

^fDepartment of Anaesthesiology, Intensive Care Medicine, Ziekenhuis Oost-Limburg, Genk, Belgium

^gFaculty of Medicine and Life Sciences, Limburg Clinical Research Centre, UHasselt, Diepenbeek, Belgium

^hDepartment of Infectious Diseases, Leiden University Medical Centre, Leiden, the Netherlands

Contexte et problématique

- 1/5 patient = infection après une chirurgie (plaie, pneumonie...)
- ↗ temps de séjours + taux de morbidité
- le diagnostic est souvent tardif (5^{ème}-9^{ème} jour) = complications
- personnel surchargé par le volume de données à analyser

Comment guider des décisions cliniques concrètes (surveillance, examens, sortie), et non simplement de prédire une infection a posteriori ?

Exemple pratique d'un succès = PERISCOPE

Le design idéal

Objectifs de PERISCOPE :

- estimations précises du risque d'infection postopératoire avant, pendant et/ou après l'intervention chirurgicale
- permettre une meilleure prise de décision
- prévention et surveillance personnalisées
- ↘ l'impact des infections postopératoires

60+ Paramètres Analysés



Signes vitaux
(FC, Température, TA)



Labo
(CRP, Leucocytes, Hémoglobine)

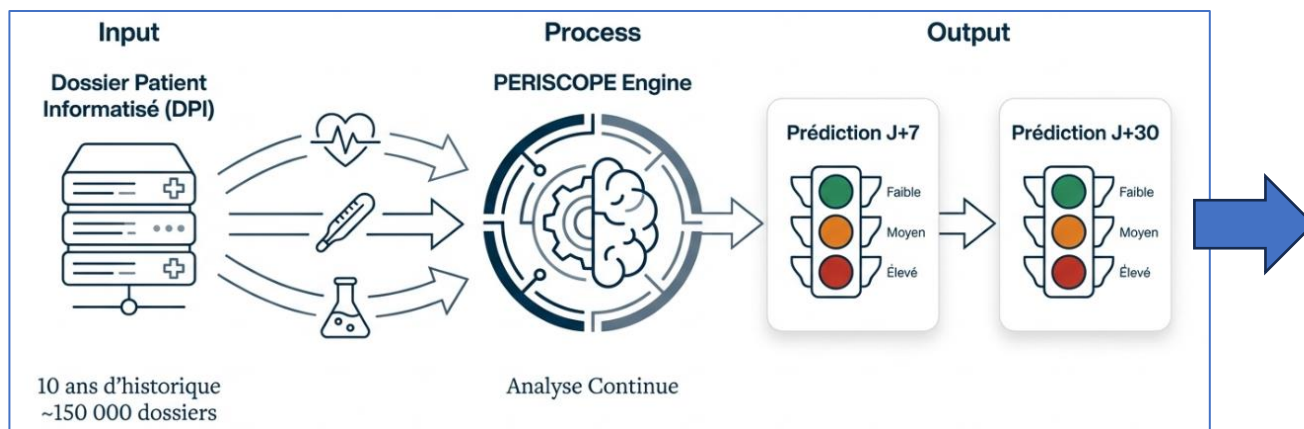


Détails opératoires
(Durée, Perte sanguine, Urgence)



Comorbidités
(Diabète, Score ASA, IMC)

Méthode



Score de risque continu,
⇒ mesure/ estimation du risque

La vraie valeur ajoutée : à partir de ce risque, agir améliore les soins

Exemple pratique d'un succès = PERISCOPE

Comment répondre aux biais cités précédemment ? (1)

1

Etude multicentrique internationale :

- 2 hôpitaux universitaires (Pays-Bas)
- 1 hôpital non universitaire (Belgique)
- données EHR réelles, intégrées au flux de soins
- 253 010 interventions chirurgicales
- ~24 000 infections postopératoires ($\approx 9,4\%$)
- Patients « non sélectionnés »

Diversité clinique non filtrée
Pas de biais de sélection

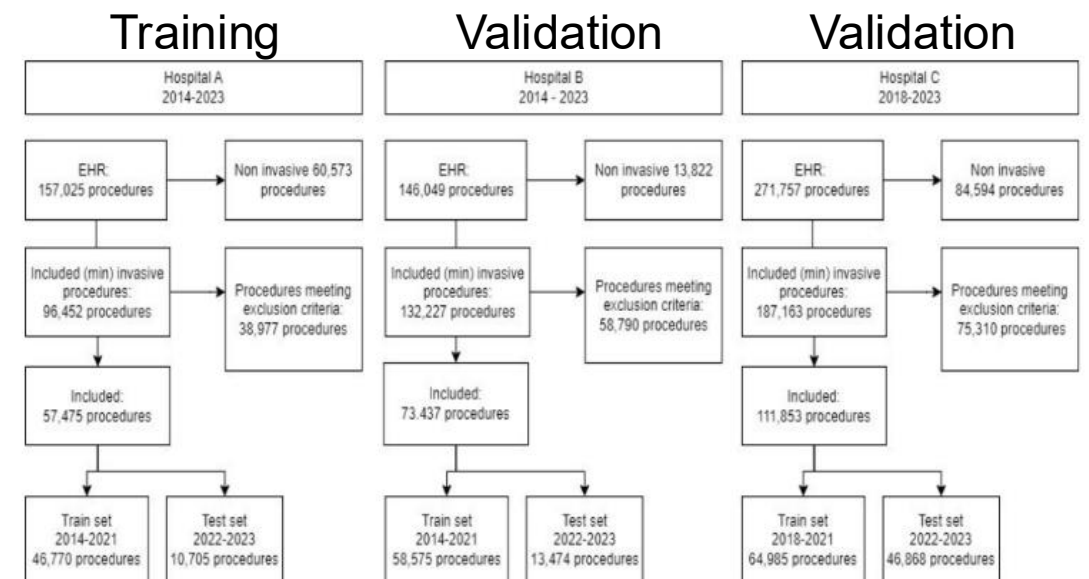


2

Modèle de développement temporel

- Développement rétrospectif mais validation prospective
- Données utilisées strictement disponibles à l'instant t

Prédiction avant le diagnostic d'IAS
Pas de biais de temporalité



Exemple pratique d'un succès = PERISCOPE

Comment répondre aux biais cités précédemment ? (2)

3

Intégration au système clinique = favoriser l'adhésion

- Intégration dans le workflow clinique
- Stratégies claires :
- Risque élevé → surveillance plaie J+1, labo J+3
- Risque bas → sortie J+2
- Évaluation du « Net Benefit » évalué avec différents seuil *in situ*

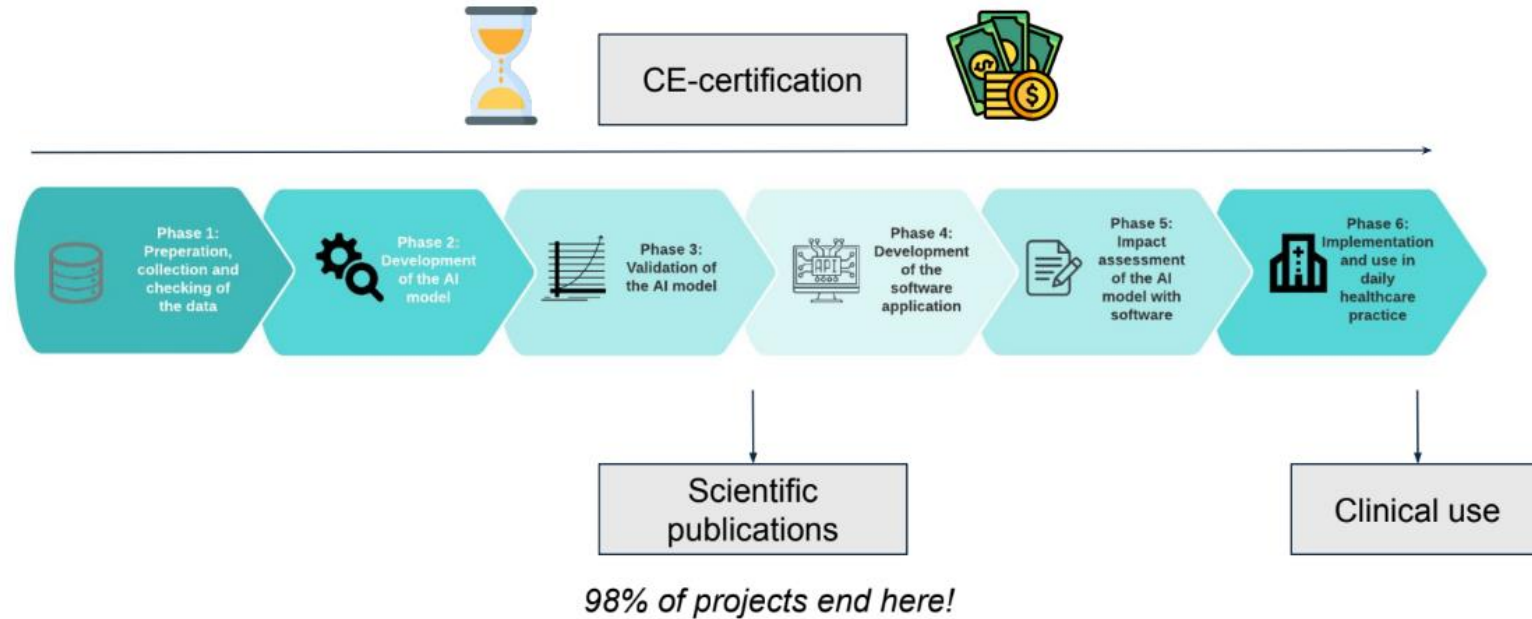


Figure issue de la présentation de Siri van der Meijden, PhD,
(Leiden University Medical Center, The Netherlands), (Healthplus.ai, October 24, 2025)

Cap de l'implémentation + certification = adhésion de l'équipe ⇒ confiance

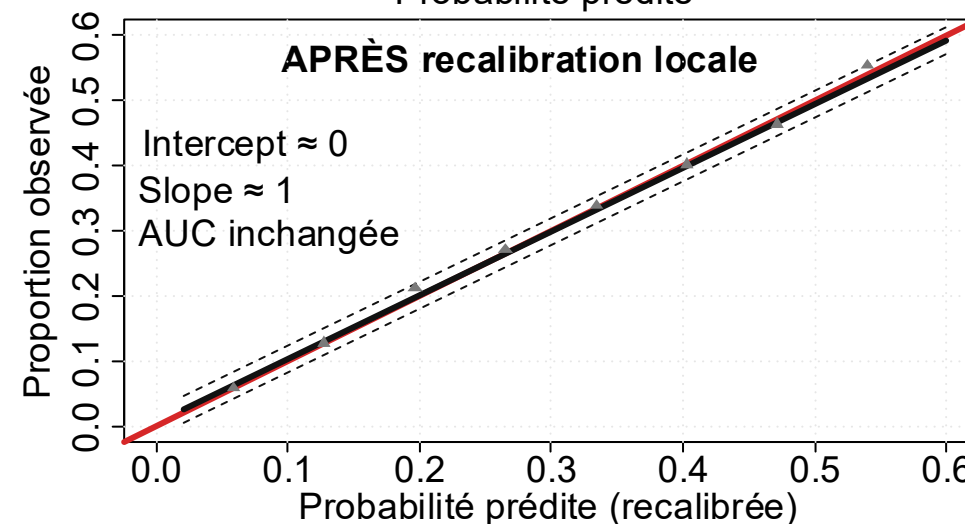
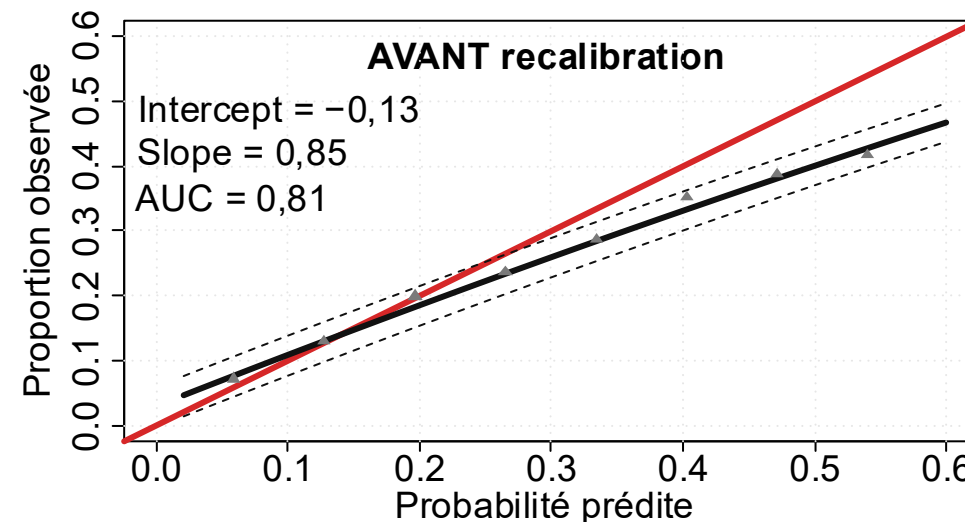
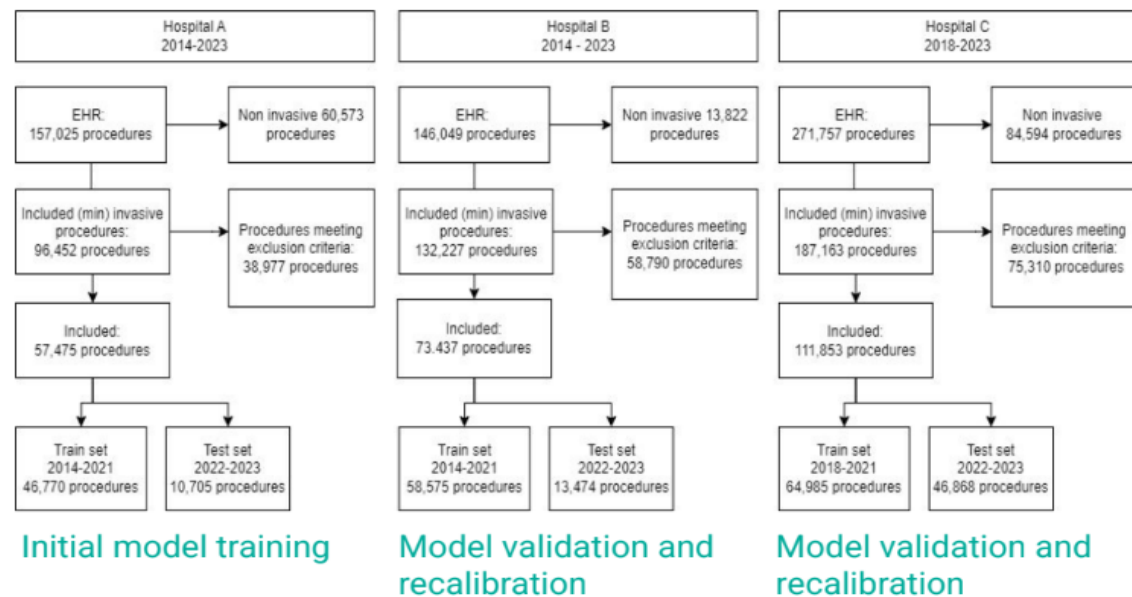
Exemple pratique d'un succès = PERISCOPE

Comment répondre aux biais cités précédemment ? (3)

Métriques transparentes

- AUC-ROC / VPN / VPP
- Validation + recalibration
- Adapter les seuils à chaque hopitaux

Pas de biais de métriques



«Recalibration = faire en sorte que le pourcentage de risque affiché par le modèle corresponde à la réalité clinique locale »

Exemple pratique d'un succès = PERISCOPE

Comment répondre aux biais cités précédemment ? (4)

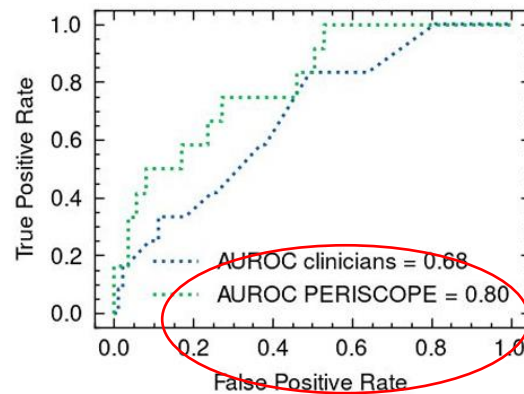
4

Follow-up study: doctor versus algorithm

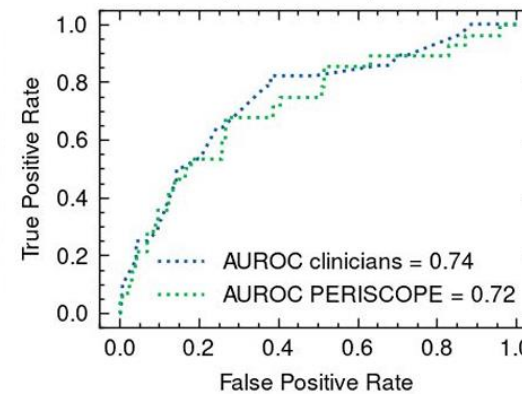
51 surgeons were asked to estimate the risk of infection directly after surgery which were compared to PERISCOPE for 501 procedures

Surgeons needed to indicate how sure they were about there predictions.

Unsure predictions (n = 101)



Sure predictions (n = 382)



Very Sure predictions (n = 60)

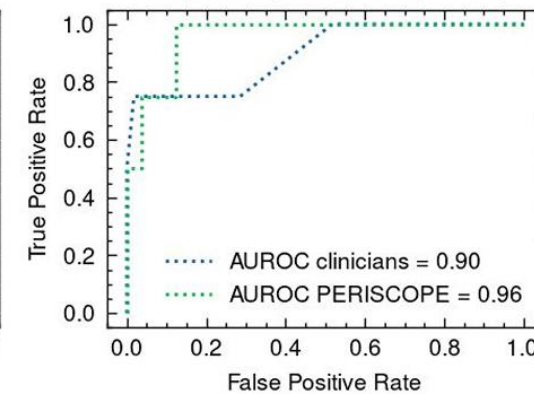


Figure issue de la présentation de Siri van der Meijden, PhD, (Leiden University Medical Center, The Netherlands), (Healthplus.ai, October 24, 2025)

Pas de biais de comparateur faible

PERISCOPE ne remplace pas l'expertise
 ⇒ un deuxième regard statistique, constant et non biaisé

Conclusion générale

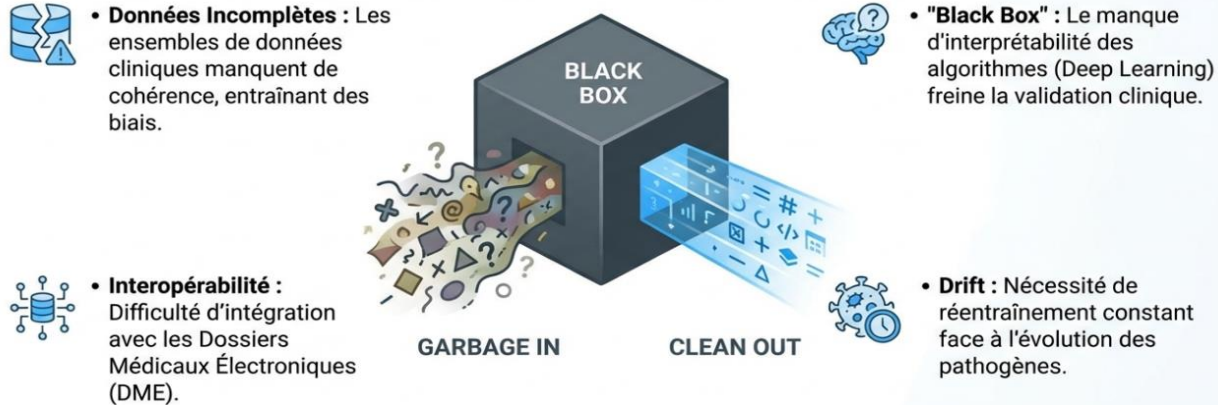
La technologie est prête
même si design lourd

Le défi
= implémentation
+ certification + adhésion

Recalibrage
indispensable

L'outil ne remplace pas
l'expert =
aide à la priorisation

Cadre éthique et légal ?



Loi bioéthique intelligence artificielle et **relation de soin**



Délégation de la
décision médicale



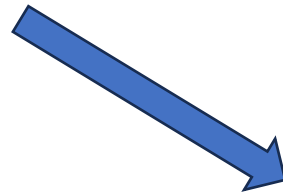
Délégation du
consentement du patient



Loi bioéthique intelligence artificielle et **relation de soin**



Délégation de la
décision médicale



- Comment s'assurer que **l'expertise** appartient au professionnel de Santé ?
- Comment veiller au risque d'**aliénation** du professionnel de santé à la machine ?
- Comment s'assurer que les professionnels sont en mesure de porter **un regard critique** ?

Loi bioéthique intelligence artificielle et relation de soin : délégation de la **décision médicale**

- Comment s'assurer que l'**expertise** appartient au professionnel de Santé ?
- Comment veiller au risque d'**aliénation** du professionnel de santé à la machine ?
- Comment s'assurer que les professionnels sont en mesure de porter **un regard critique** ?

Principe de
garantie humaine

Affirmation du rôle
de santé comme
décideur

Contrôle du
traitement
algorithmique des
données

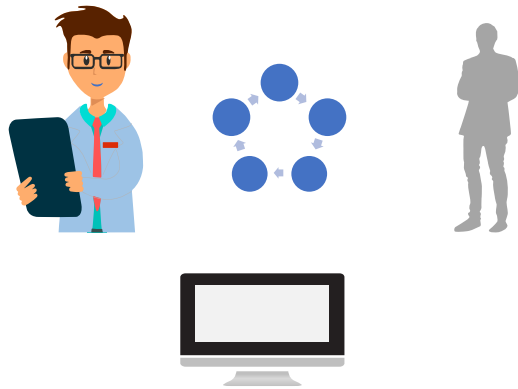
« La garantie d'une **supervision humaine** de toute utilisation du numérique en santé,
+
et l'obligation d'instaurer pour toute personne le souhaitant et à tout moment **la possibilité d'un contact humain** en mesure de lui transmettre l'ensemble des informations la concernant dans le cadre de son parcours de soins. »

Le professionnel de santé n'est pas un simple auxiliaire du traitement algorithmique
Nécessité de développer **son sens critique**

« La saisie d'informations relatives au patient dans le traitement algorithmique **se fait sous le contrôle du professionnel** de santé qui a recours audit traitement. »

Loi bioéthique intelligence artificielle et relation de soin : délégation du **consentement du patient**

- Comment éviter que les propositions de l'IA ne se substituent à **la décision partagée en santé ?**



« Le professionnel de santé qui décide d'utiliser, pour un acte de prévention, de diagnostic ou de soin, un dispositif médical comportant un traitement de données algorithmique dont l'apprentissage a été réalisé à partir de données massives s'assure que **la personne concernée en a été informée** et qu'elle est, le cas échéant, avertie de l'interprétation qui en résulte. »

« Les professionnels de santé concernés sont informés du recours à ce traitement de données. Les **données du patient** utilisées dans ce traitement et les résultats qui en sont issus leur **sont accessibles.** »

« La saisie d'informations relatives au patient dans le traitement algorithmique **se fait sous le contrôle du professionnel** de santé qui a recours audit traitement. »

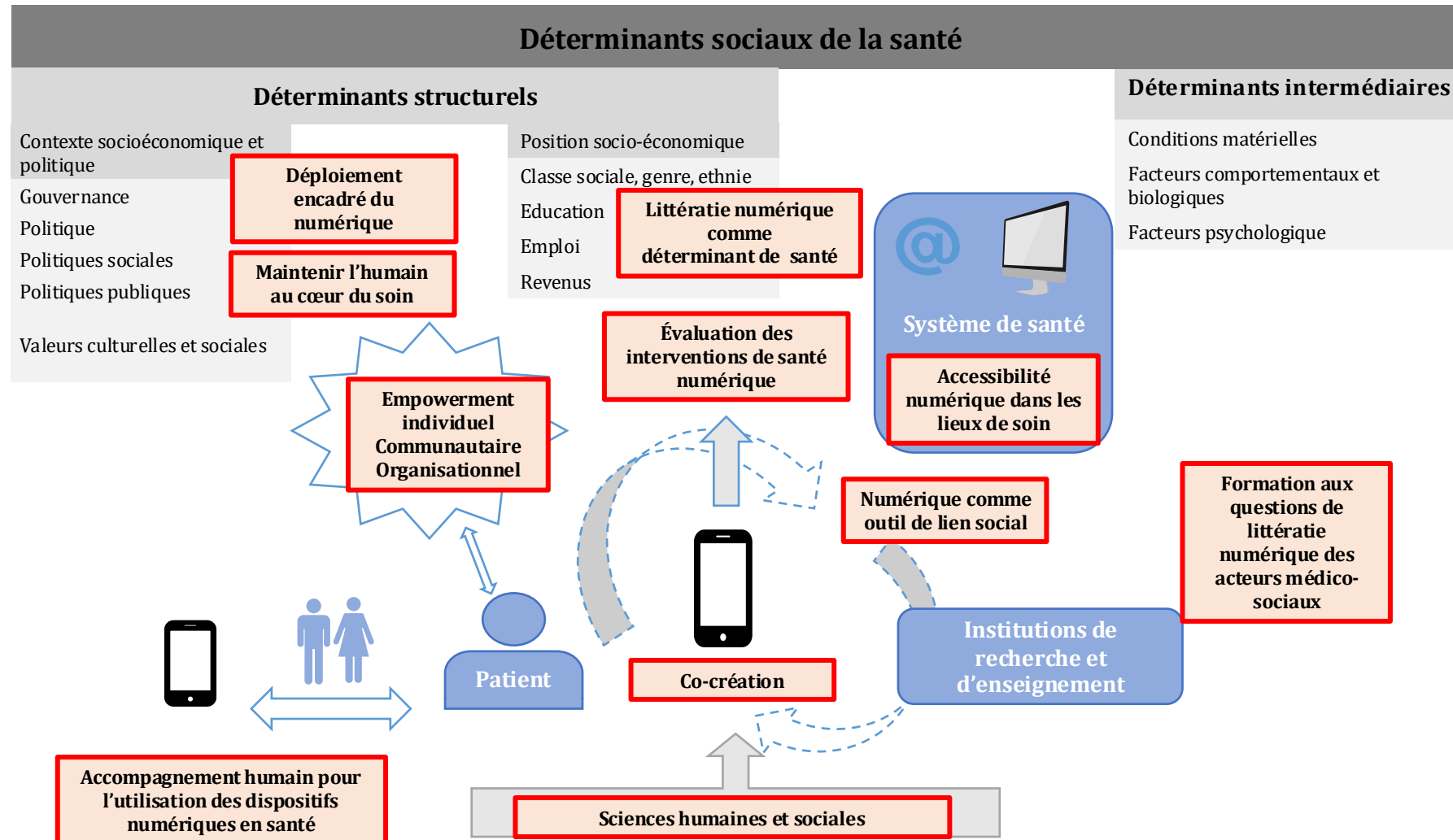
« Les concepteurs d'un traitement algorithmique mentionné au I s'assurent de **l'explicabilité de son fonctionnement pour les utilisateurs.** »

- **Question Wooclap :**
- **Quels autres enjeux éthiques vous semblent important à considérer dans l'utilisation de l'IA en santé ?**

Creuser les inégalités sociales de santé

- **15 %** de la population française en situation d'**illectronisme** en 2021
- **Illectronisme** :
 - situation d'une personne ne possédant pas les compétences numériques de base (**1.5%**) :
 - rechercher des informations en ligne
 - communiquer en ligne
 - utiliser des logiciels
 - protéger sa vie privée
 - résoudre des problèmes en ligne
 - ou ne se servant pas d'Internet (**13.9%**)
 - incapacité ou impossibilité matérielle de l'utiliser dans les **trois derniers mois**

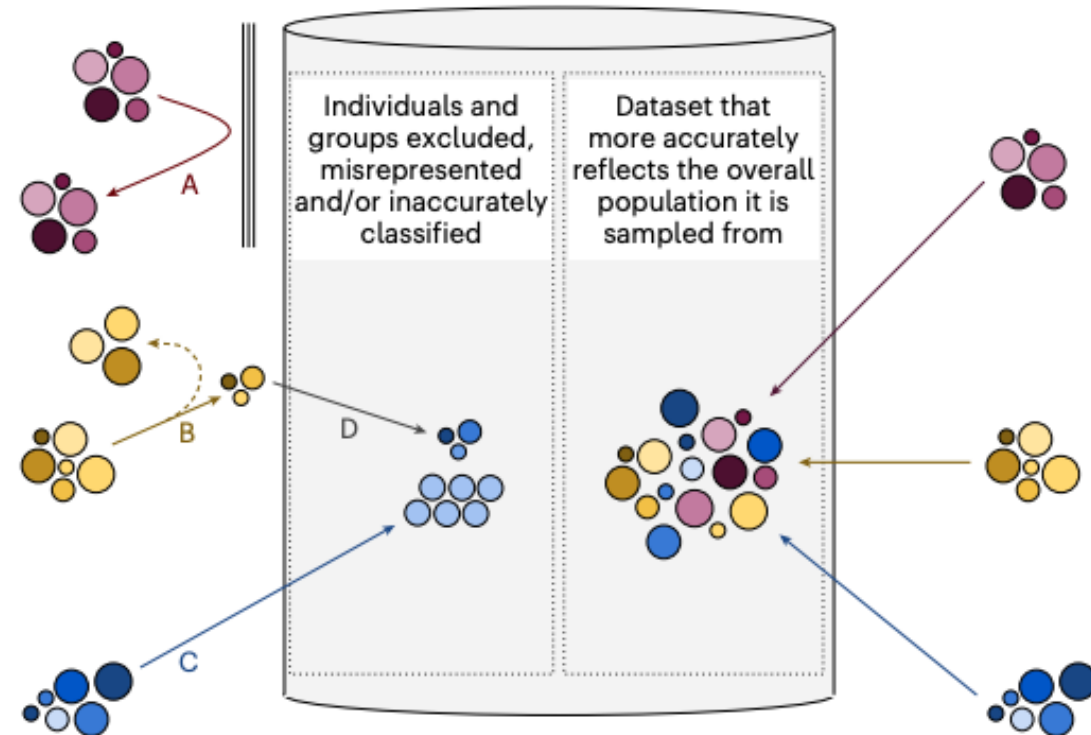
Recommandations du programme ZINGA pour l'utilisation des outils numériques à destination des patients



Inégalités sociales de santé et IA

- Barrières d'accès aux soins
- Exclusion involontaire ou délibérée de données concernant une population

- Classification (ex binaire : H/F)
- Données collectées puis supprimées, restrictions légales ou éthiques
- Données collectées mais pas suffisamment détaillées





Merci de votre attention