

# Intégration d'outils cliniques numériques dans le cadre d'un système de santé apprenant : quelles questions se poser ?

2024-06-03 : École d'été interdisciplinaire en numérique de la santé

#### Jean-François Ethier

#### Professeur titulaire

Département de médecine FMSS – Université de Sherbrooke Chaire en informatique de la santé UdeS

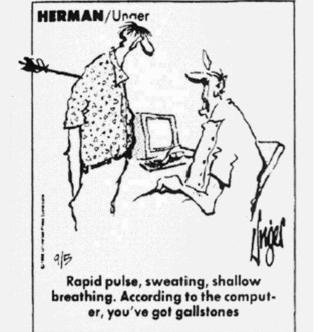


#### Codirecteur scientifique

Groupe de recherche interdisciplinaire en informatique de la santé (GRIIS.ca)

#### Directeur technologique

Réseau de recherche sur les données de santé du Canada













## Après la séance, vous ...

- Comprendrez la place de l'IA
- Saurez poser les questions importantes pour évaluer si un outil basé sur l'IA est :
  - pertinent;
  - sécuritaire;
  - offre une valeur ajoutée.









### Numérique en santé: pourquoi et comment?

#### 1. Tâches et défis pour l'IA

- IA vs statistiques
- Tâches de l'IA
- Défis de l'IA

#### 2. Les IA

- Apprentissage machine
- Apprentissage profond
- Réseau de neurones
- Modèles fondationnels

- 3. IA en cliniques : état des lieux
- 4. Outils numériques en clinique : desiderata
  - Utilité
  - Confiance et évaluation
  - Ressources











### Statistiques vs IA

Caractéristiques d'une population vs prédiction pour un individu













#### Statistiques vs IA: comment et pourquoi?

#### **Statistiques inférentielles**

- Accent sur l'analyse des liens entre prédicteurs dans une population
  - Ampleur du lien et signification statistique
- Implique la création d'un modèle de relation

#### Intelligence artificielle

- Accent sur la capacité de prédiction individuelle
- Corrélations
- Trop d'intrants
- Modèle de données non-structurées
  - Images
  - Les pompes à insuline
  - Les phrases













Et si on pouvait générer le modèle automatiquement à partir des données...











#### **Question A2**

- Quelle phrase décrit une limitation significative de l'intelligence artificielle (IA) dans le domaine médical?
  - Les systèmes d'IA manquent de capacité pour traiter efficacement de grandes quantités de données médicales.
  - L'IA est incapable de gérer les processus décisionnels complexes impliqués dans le diagnostic et le traitement médical.
  - Les algorithmes d'IA peuvent avoir du mal à interpréter des indices contextuels subtils et les émotions des patients lors des interactions.
  - L'IA dans le domaine médical est imperméable aux biais, assurant des résultats de soins de santé justes et impartiaux.













#### **Question A3**

- Quel rôle majeur l'IA joue-t-elle dans le domaine médical ?
  - a) L'IA en médecine se concentre principalement sur l'automatisation des tâches administratives, telles que la planification des rendez-vous et la facturation.
  - b) L'IA est principalement utilisée en médecine pour remplacer les médecins humains dans le diagnostic et le traitement des patients.
  - c) L'IA en médecine implique l'utilisation d'algorithmes avancés et d'apprentissage automatique pour analyser les données médicales, aider dans les diagnostics et personnaliser les plans de traitement.
  - d) Le rôle de l'IA en médecine se limite à la création de simulations de réalité virtuelle à des fins de formation médicale.













## **Question A4**

- Est-ce que des outils basés sur l'IA peuvent être « créatifs » ?
  - Oui
  - Non











# Quels types de tâches pour l'IA?













#### Exemples de tâches : trouver des patterns

- **Classer** des données
  - Exemple : identifier le sujet principal d'une image
- **Grouper** des clients avec des intérêts commerciaux similaires
- Suggérer des actions
  - Exemple: conduite autonome













#### Exemples de tâches : « générer » du nouveau contenu

- « Generative AI »
  - On y reviendra avec les « fondational models »
- « Créer »
  - Des images (p. ex. : DALL-E)
  - Du texte (p. ex. : **Chat**GPT)
  - Du code informatique (p. ex. : Skynet ??)
  - De l'audio (p. ex. : VALL-E)















### Quand on parle d'IA, on parle de quoi?

AI vs ML vs Deep learning vs LLM vs Generative AI





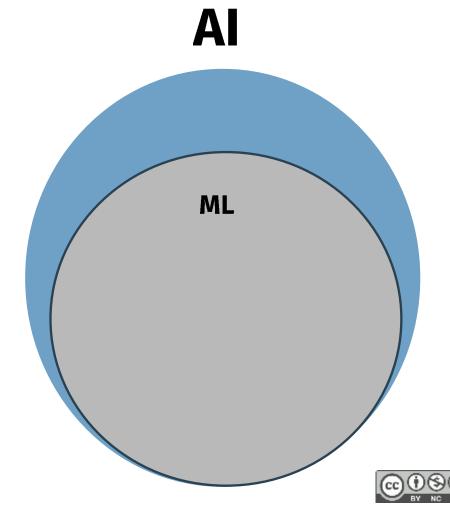






### Intelligence artificielle

- "[...] the science and engineering of making intelligent machines." \*
- Good, Old-Fashioned AI (GOFAI)
  - Engins de règles, systèmes experts, graphes de connaissances, etc.
- Apprentissage machine











# **Apprentissage machine**

Machine learning







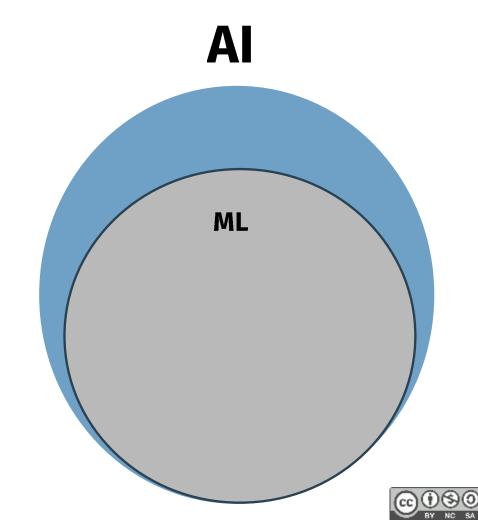






## **Apprentissage machine**

- Capacité de se modifier lorsqu'exposé à plus de données.
  - Dynamique, n'a pas nécessairement besoin d'une intervention humaine
  - Pas de programmation additionnelle









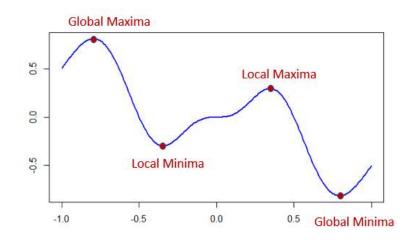


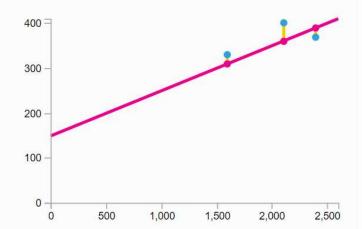
#### Apprendre quoi?

Optimiser le résultat d'une corrélation

- « Minimiser le résultat de la fonction »
  - Des dizaines ou des centaines de paramètres

 Classiquement, choix des paramètres par le scientifique







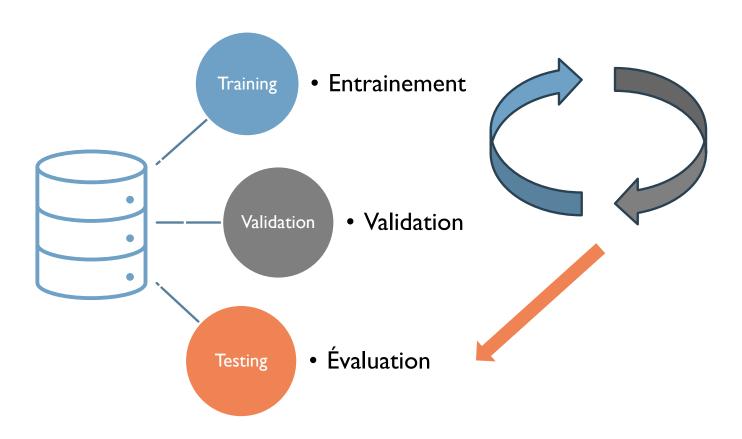








#### **Processus**















## Big data vs fat data

- **Big** data
  - 1. Volume
  - Vélocité
  - 3. Variété

- **Fat** data
  - Plusieurs types de données











# **Apprentissage profond**

Deep learning







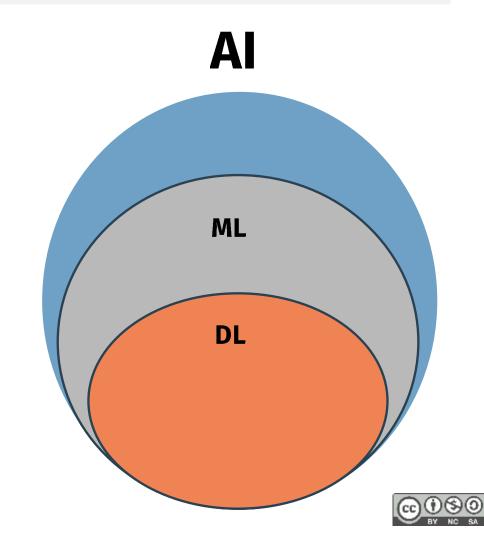






### **Apprentissage profond**

- Souvent utilisé comme synonyme de « réseau de neurones »
- Fonctionne particulièrement bien sur des données non structurées (p. ex. des pixels)
- Nécessite beaucoup plus de données



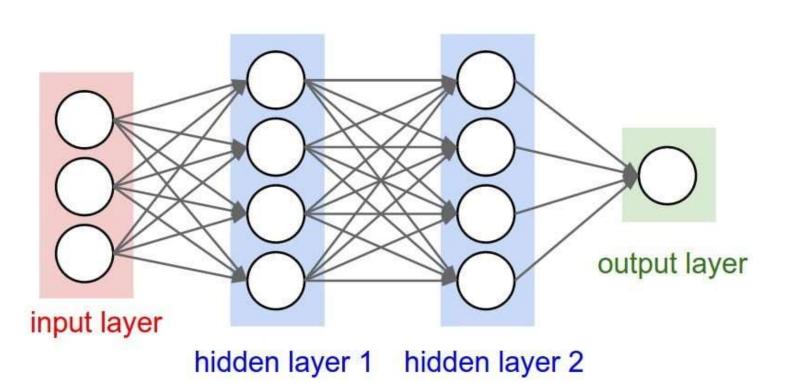


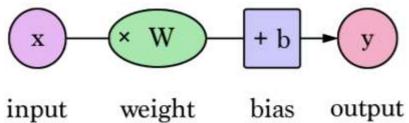


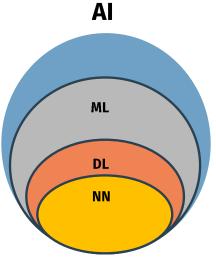




#### Réseau de neurones











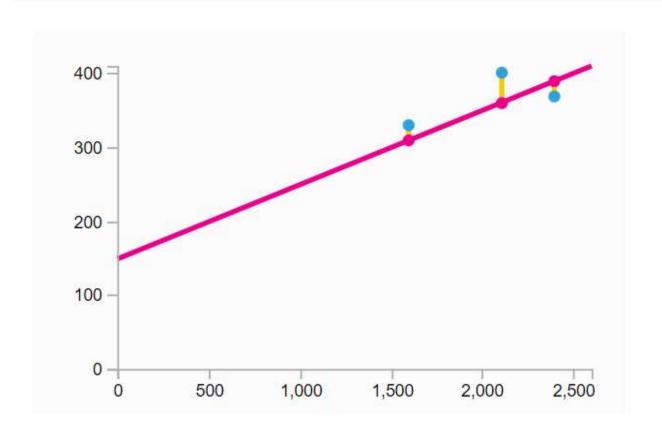


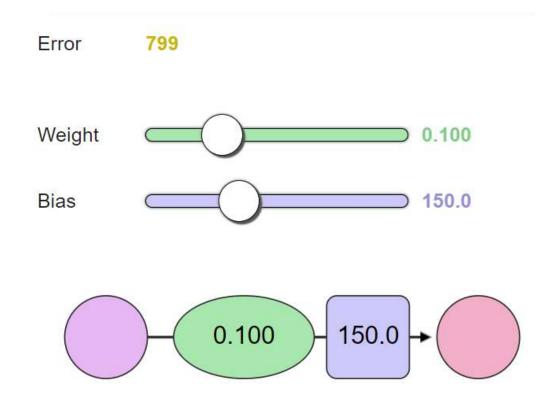






#### Réseau de neurones





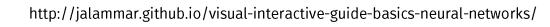










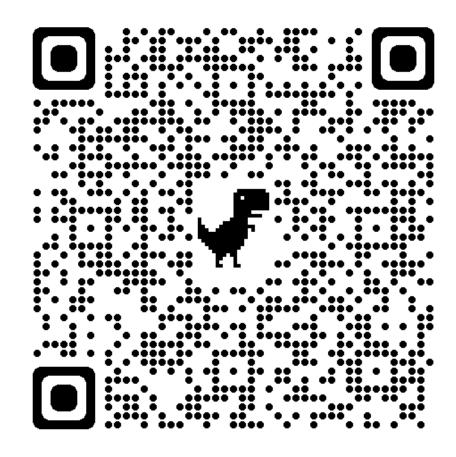




# **Exemple interactif**

https://url.griis.ca/exia

http://jalammar.github.io/visual-interactive-guidebasics-neural-networks/#train\_your\_dragon













# Supervision d'apprentissage

Apprentissage machine













#### **Question 4**

- Quelle serait une utilisation possible des grands modèles de langage (modèles fondationnels) centrés sur la médecine comme Med-PaLM Multimodal (Google) et LLaVA-Med (Microsoft) en pratique clinique ?
  - a) Prédire l'âge et le sexe d'un patient avec la photo d'une rétine.
  - b) Prédire la réponse d'une tumeur à une molécule d'immunothérapie avec la photo d'une lame de pathologie.
  - c) Prédire la mortalité d'un patient hospitalisé dans la prochaine année.
  - d) Toutes ces réponses.
  - e) Aucune de ces réponses.





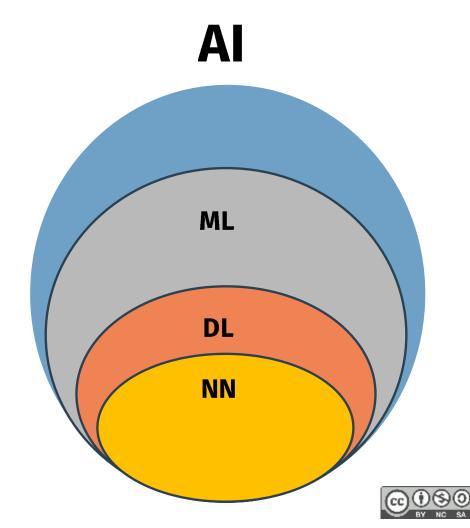






## Apprentissage profond supervisé

- Utilisé surtout pour classifier
- On donne la bonne réponse avec les données
- Peut être très « labour intensive »







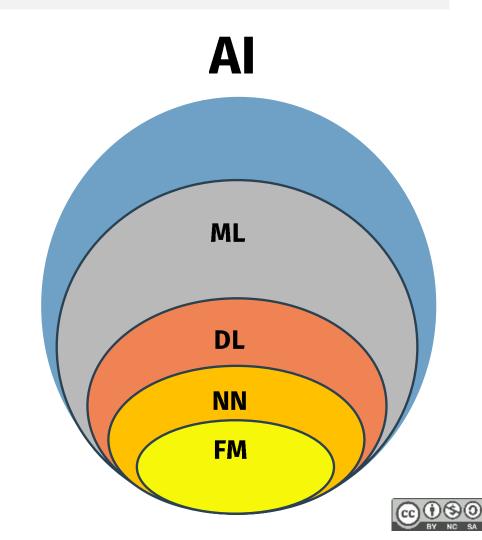




### Apprentissage profond non supervisé\*

#### Foundational models

- Large language model (GPT-4)
- Capte les associations largement présentes dans le domaine
- **Prédire** le **symbole suivant** basé sur le précédent
- Intégration images et textes
  - Med-PaLM Multimodal (Google), LLaVA-Med (Large Language and Vision Assistant for biomedicine) –
     Microsoft





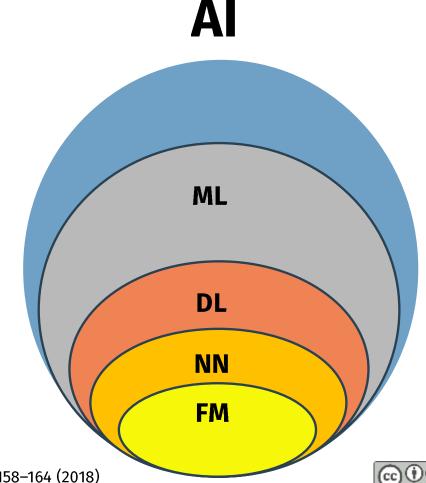






#### **Foundational models**

- Apprendre ce qu'est un œil sans avoir besoin d'un jeu d'image annoté
  - BEAUCOUP de données requises











# Combinaison de FM avec d'autres approches













#### FM, classification et generative Al

- FM + apprentissage supervisé
  - Moins d'images annotées requises
    - Ajout d'images avec étiquette de rétinopathie diabétique
    - Age et genre
  - Cliniciens ne peuvent pas le faire
  - "The model hasn't yet been tested in a clinical setting."

- FM + chatbot = generative AI application
- GPT-4 plus fonctions d'interactions = ChatGPT













#### Les défis de l'IA

Particulièrement l'apprentissage profond













# Défis d'interprétabilité











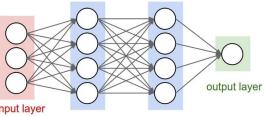


# Question de l'interprétabilité

Poids de la balle	Hauteur de chute	Temps de chute
500 g	10 m	1.5 s
400 g	7 m	1 s
•••	•••	•••

Régression (statistique) :  $y = y_0 + v_{yo} t - \frac{1}{2} g t^2$ 

Réseau de neurones : ???









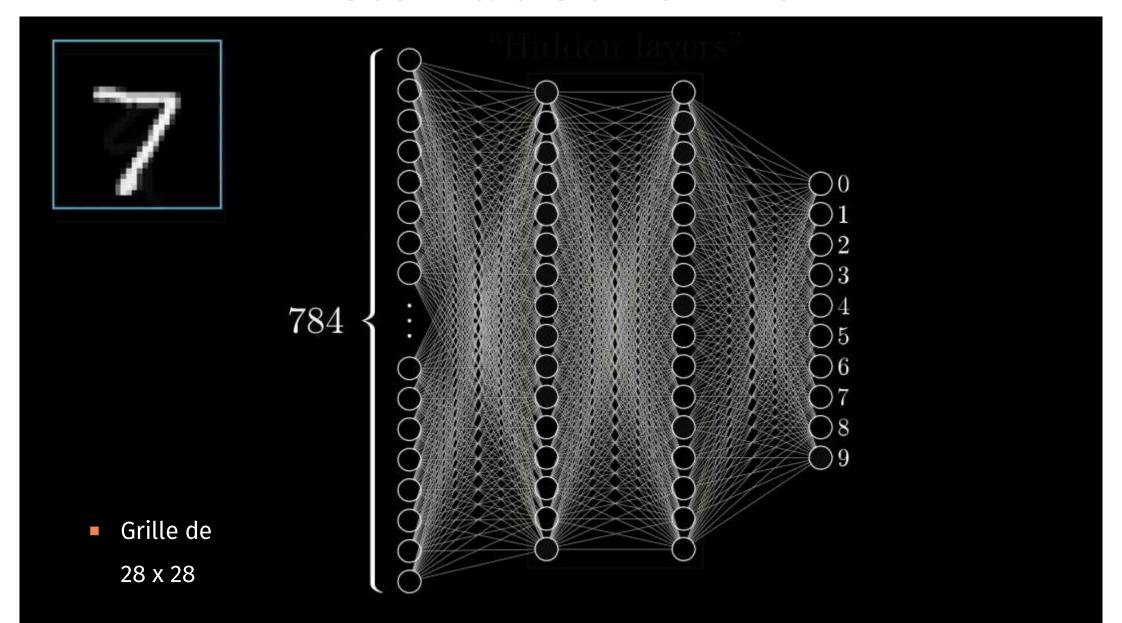








#### **Reconnaitre un chiffre**

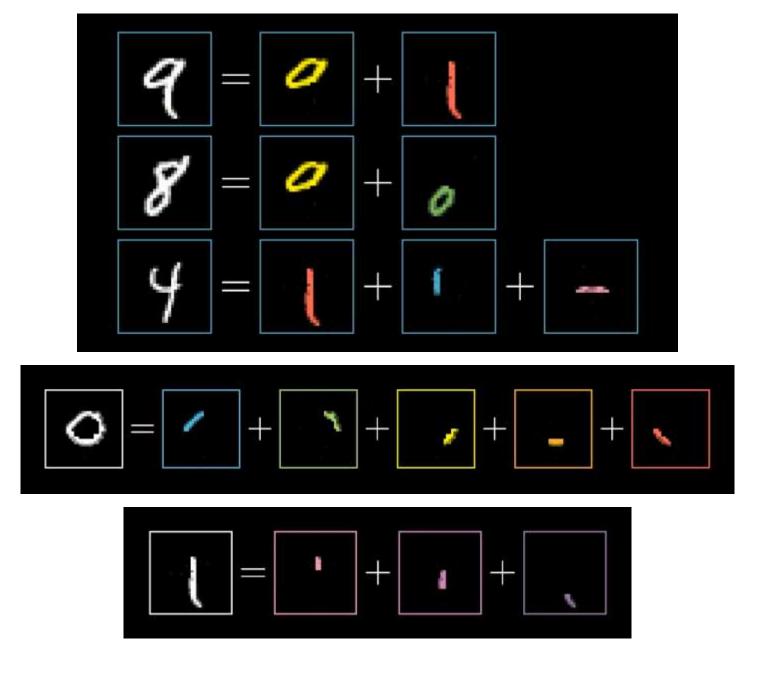


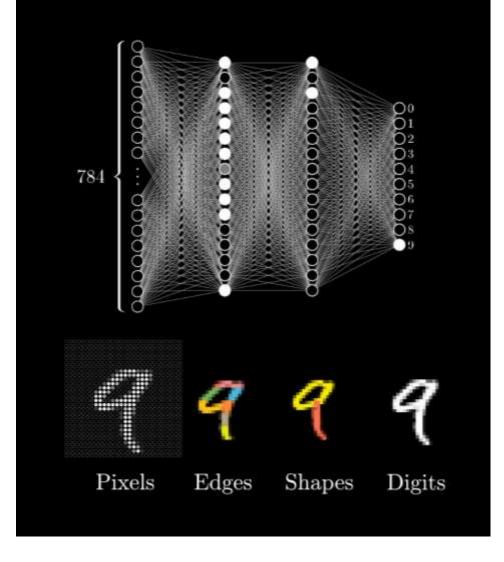














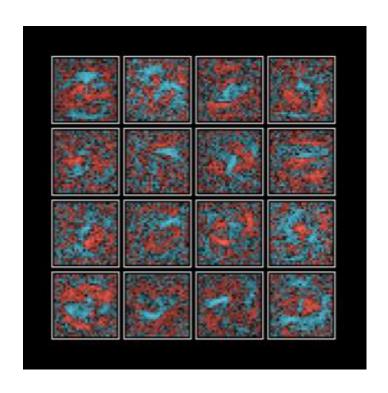


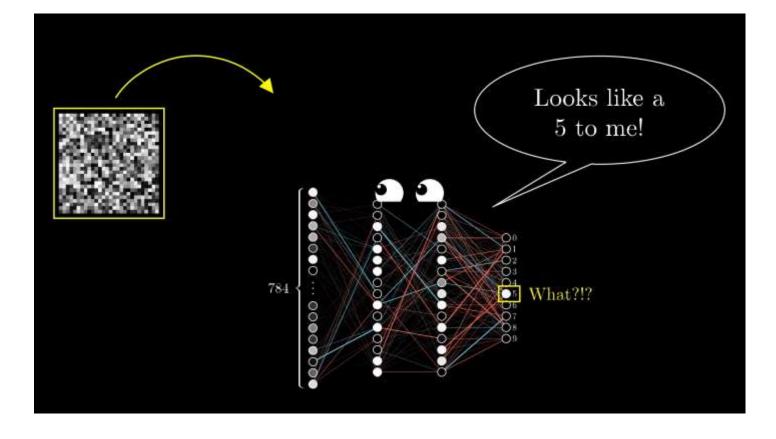






### Deuxième niveau de neurones dans les faits :













### Défis de stabilité

Apprentissage machine



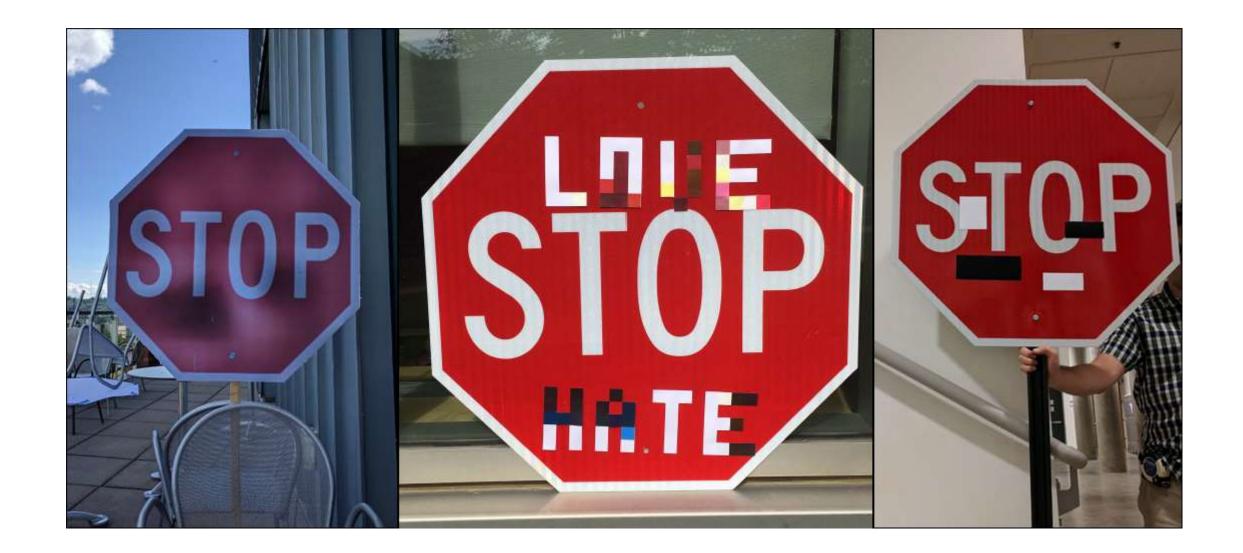












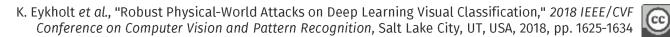














### Changement au contexte externe

- Café et cancer du poumon
  - Statut buveur de café et statut tabagique
  - Présence ou pas de cancer du poumon
  - Et si on restreint grandement les possibilités de tabagisme ?

- Exemple maison et prix
  - Superficie et nombre de salles de bain (proxy nombre d'éviers)
  - Nouvelle règle pour doubler les éviers par salle de bain





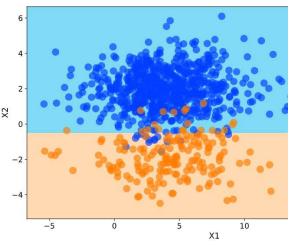






## Santé: changement de contexte en clinique

- Visite à l'urgence précovid vs durant le covid
  - Si c'est un proxy de comorbidité... ça change !
  - Mais comment le savoir qu'il faut changer si on ne peut pas interpréter









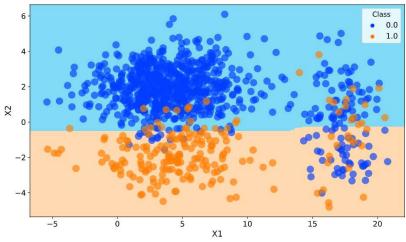




## Santé : changement de contexte en clinique

- Visite à l'urgence précovid vs durant le covid
  - Si c'est un proxy de comorbidité... ça change!
  - Mais comment le savoir qu'il faut changer si on ne peut pas interpréter

- Pompe à insuline
  - Ajout de GLP1... impact?

















# Section grande respiration et étirement!











# OK... Défis pour l'IA en clinique











### **Question B4**

 Est-ce que vous croyez que les outils basés sur l'IA seront plus utiles comme ...

- a) Outils prédictifs (p. ex. : prédire le risque de faire un AVC sur 5 ans)
- o) Outils de recommandation (p. ex. : anticoaguler ou pas)











#### En théorie

- Les humains ont une variabilité dans les performances
  - fatigues physiques, perte de focus, difficulté avec les tâches répétitives
- "capable of handling complex interactions in large datasets to predict outcome with greater accuracy, but the models need a greater number of input-output pairs to learn from"











## Situation présente

Plus de 500 outils basés sur l'IA autorisés par la FDA

- Entre 10 % et 30 % des médecins ont utilisé un outil basé sur l'IA aux États-Unis
  - réaction allant de l'optimisme prudent au manque de confiance total

- Santé Canada
  - modèles d'IA qui pourront se modifier pendant leur utilisation en clinique











### Beaucoup d'espoirs...

It was one of those amazing "we're living in the future" moments. In October 2013 [...] "MD Anderson is using the IBM Watson cognitive computing system for its mission to eradicate cancer."

 Well, now that future is past. The partnership between IBM and one of the world's top cancer research institutions is falling apart. (2017)











## Défis liés à la validation et à l'implémentation

ARTICLES | VOLUME 1, ISSUE 6, E271-E297, OCTOBER 01, 2019

A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis

Xiaoxuan Liu, MBChB † Livia Faes, MD † Aditya U Kale, MBChB Siegfried K Wagner, BMBCh Dun Jack Fu, PhD Alice Bruynseels, MBChB et al. Show all authors Show footnotes

Open Access - Published: September 25, 2019 - DOI: https://doi.org/10.1016/S2589-7500(19)30123-2

Our review found the diagnostic **performance** of deep learning models to be **equivalent** to that of **health-care professionals**. However, a **major finding** of the review is that **few** studies presented **externally validated results** or compared the performance of deep learning models and health-care professionals using the **same sample**. Additionally, **poor reporting** is prevalent in deep learning studies, which **limits** reliable **interpretation** of the reported **diagnostic accuracy**.

- Performance souvent équivalente aux professionnels de santé
- Peu souvent validé à l'externe
- Lorsque validé, échantillon souvent différent
- Publication des résultats sous-optimale





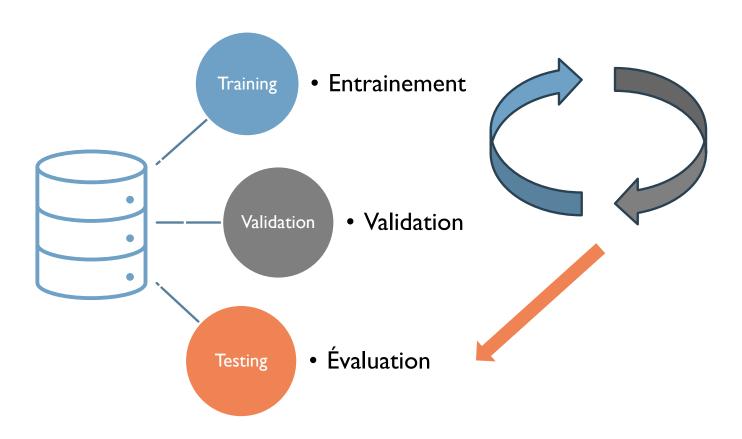








#### **Processus**

















# Desiderata pour l'utilisation de l'IA en clinique

Questions que vous pourriez vouloir poser avant de choisir un outil











## **Question 2**

- Étant donné la grande quantité de données utilisées pour entrainer les outils d'IA, quel avantage peuvent-ils procurer par rapport aux outils prédictifs usuels (p. ex. : le score de CHADS2) pour votre pratique?
  - Meilleure précision des prédictions a)
  - Moins de biais
  - Meilleure compréhension des mécanismes sous-jacents
  - A, B, C
  - Aucune de ces réponses











### **Question 3**

Comment identifier les outils basés sur l'IA qui seront utiles pour ma pratique ?

- a) Bas score de Brier pour le modèle
- b) Échantillon d'entrainement du modèle semblable à mes patients
- c) Score d'aire sous la courbe élevé des patients difficiles à diagnostiquer
- d) Échantillon de validation comportant au moins le double de la population couverte par mon hôpital









## Utile?

Pour qui et pourquoi?













### Dépasser l'aire sous la courbe

Est-ce que de prédire correctement qu'un patient en soins palliatifs va mourir devrait faire partie de l'évaluation ?

Points
10
8
9
28
14
0
7

(Assessed

External validation of the Hospital-patient One-year Mortality Risk (HOMR) model for predicting death within 1 year after hospital admission

Carl van Wakaven, Finlay A. McAlister, Jeffrey A. Bakal, Steven Hawken and Jacques Donzé CMAJ July 14, 2015 187 (10) 725-733, DOI: https://doi.org/10.1503/cmaj.159209

Research | Open Access | Published: 02 December 2017

#### The Brier score does not evaluate the clinical utility of diagnostic tests or prediction models

Melissa Assel, Daniel D. Sjoberg & Andrew J. Vickers

Diagnostic and Prognostic Research 1, Article number: 19 (2017) | Cite this article
11k Accesses | 17 Citations | 3 Altmetric | Metrics

	RF-AdminDemoDx	RF-AdminDemo	RF-Minimal	mHOMR
Internal validation <sup>a</sup>				
C-statistic (range)	0.90 (0.90-0.91)	0.86 (0.85-0.87)	0.85 (0.84-0.86)	0.86 (0.85-0.86)
Brier score (range)	0.068 (0.065-0.073)	0.079 (0.077-0.083)	0.082 (0.078-0.084)	0.081 (0.078-0.085
External validation b,c				
C-statistic (95% CI)	0.89 (0.88-0.89)	0.85 (0.84-0.86)	0.84 (0.83-0.84)	0.84 (0.83-0.85)
Brier score (95% CI)	0.074 (0.072-0.076)	0.084 (0.081-0.086)	0.086 (0.084-0.089)	0.086 (0.083-0.088
CDSS-eligible validation <sup>b</sup>	nd			
C-statistic (95% CI)	0.86 (0.85-0.87)	0.81 (0.80-0.82)	0.79 (0.78-0.80)	0.80 (0.79-0.81)
Brier score (95% CI)	0.088 (0.085-0.091)	0.10 (0.097-0.10)	0.10 (0.10-0.11)	0.10 (0.099-0.11)

Expected clinical utility of automatable prediction models for improving palliative and end-of-life care outcomes: Toward routine decision analysis before implementation 3

Ryeyan Taseen 🛎, Jean-François Ethier

Journal of the American Medical Informatics Association, Volume 28, Issue 11, November 2021, Pages 2366–2378, https://doi.org/10.1093/jamia/ocab140







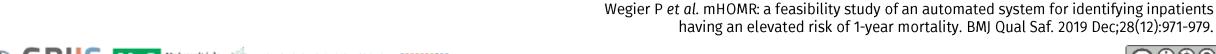




#### Utilité

- Qui étaient les patients inclus ?
- Le modèle fonctionnait pour quels patients ?
  - Pas juste une métrique pour le modèle global
- Est-ce que les patients chez qui le modèle performait moins bien sont exclus de l'analyse?
  - Intention to treat

- Est-ce que, pour les patients chez qui ça fonctionne, le modèle m'aurait aidé?
- Est-ce que le modèle identifie des patients que j'aurais manqués
  - Exemple : des patients qui n'ont pas eu de discussions de soins de vie











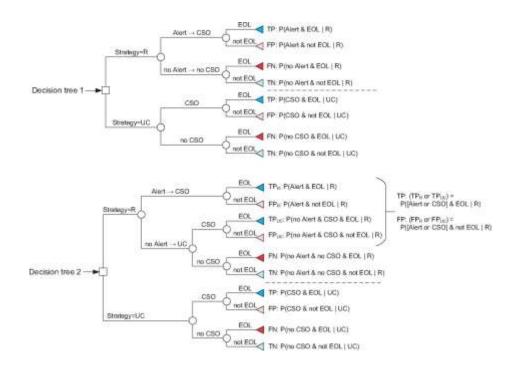
## Anticiper l'utilité clinique

- Le comparateur devrait être les soins usuels.
  - Arbre 2 : même en cas de non-alerte, en soins usuels, des actions peuvent être posées par les cliniciens.

Expected clinical utility of automatable prediction models for improving palliative and end-of-life care outcomes: Toward routine decision analysis before implementation 3

Ryeyan Taseen . Jean-François Ethier

Journal of the American Medical Informatics Association, Volume 28, Issue 11, November 2021, Pages 2366–2378, https://doi.org/10.1093/jamia/ocab140











# Ajustable pour une situation clinique donnée?













### Est-ce qu'on peut ajuster le résultat selon la question?

- Tendance à égaliser la sensibilité et la spécificité en entraînement
  - Peu utile en clinique
  - Possible d'ajuster le modèle selon la question clinique ?
    - Plus sensible ? Plus spécifique ? Meilleure VPP ?
- Est-ce seulement oui ou non ?
  - Possible d'avoir une probabilité ?
- Est-ce que l'incertitude sur la prédiction est disponible ?











## **Traitement des données manquantes**

- Est-ce que le système donne toujours une réponse?
- Est-ce qu'une valeur est ajoutée artificiellement?
  - Valeur semblable aux patients semblables qui en avaient?
  - Valeur normale?
    - Exemple: albumine et bilirubine aux soins intensifs













Utile?

Ajustable pour une situation donnée ?

# Ajustable pour un patient donné?











## Informations pour chaque prédiction?

- Pour un patient donné, sommes-nous confiants du résultat ?
- Pouvons-nous avoir une idée des intrants ayant eu un impact ?





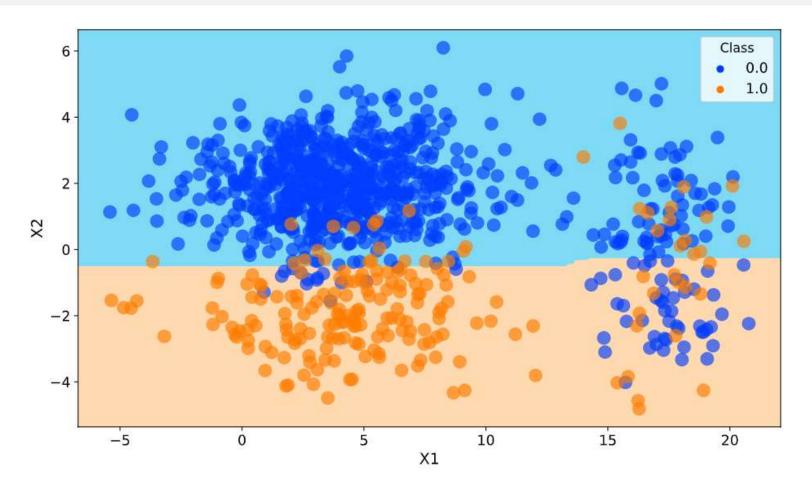








# Très bon modèle global... utile pour tous?





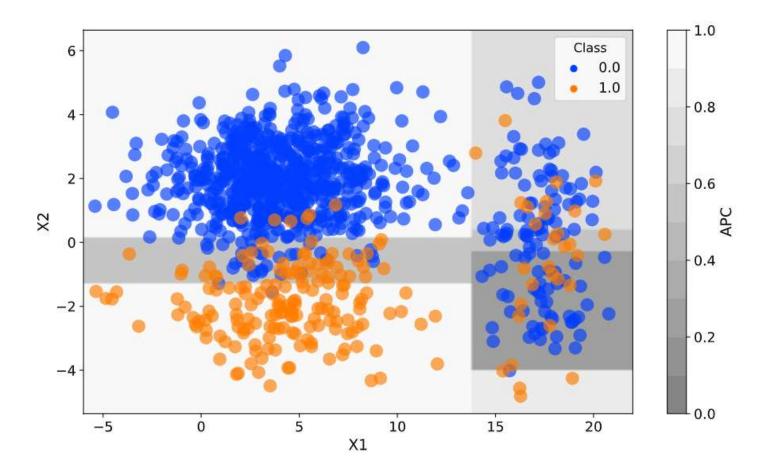








### **Confiance différentielle**













Utile?

Ajustable pour une situation donnée ?

Ajustable pour un patient donné?

# Permet un déploiement sécuritaire?









# Capacité d'évaluation en continu

Une fois déployé, le défi est que l'outil demeure sécuritaire et pertinent













### Évaluation en continu

- Au **départ**
- Durant l'utilisation
  - Impact de la contamination liée à l'utilisation du modèle
  - Extrapolation
  - Changement du contexte de soins













# Au départ

- Est-ce que notre population est similaire ?
  - Est-ce que nous avons une sous-population qui n'était pas présente dans l'entrainement ?





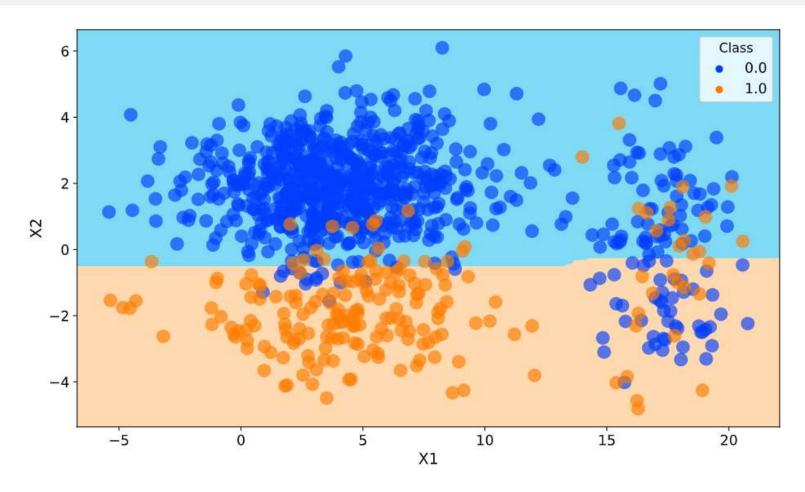








# Au départ













### **Durant l'utilisation: contamination positive**

- Une fois que les cliniciens sont sensibilisés
  - Exemple : utilité sur les discussions de fin de vie... si tout le monde le fait

• Quand faut-il penser à retirer un modèle car le retour sur investissement est trop faible ?











## **Extrapolation**

- Suivre l'utilisation clinique réelle
- Équivalent d'un médicament utilisé hors indication
- Évaluation à faire
  - Performance, utilité, risques, bénéfices











# Communication des changements du modèle













### Changements au modèle

- Est-ce que vous êtes notifié si le modèle est réentraîné ?
  - Semblable à une nouvelle machine de scintigraphie avec sensibilité augmentée
- Est-ce que le modèle peut changer en continu ?
  - Donc deux résultats différents d'une journée à l'autre pour le même patient
- Qu'est-ce qu'on va vous communiquer?











Utile?

Ajustable pour une situation donnée ?

Ajustable pour un patient donné?

Sécuritaire ?

# Permet un déploiement éthique, légal et équitable ?











## Contrôle de ce qui sort de votre organisation?

- Est-ce que vos données sont utilisées pour entrainer le modèle qui sera revendu ailleurs ?
- Est-ce que les données de vos patients sont dans le modèle ?
  - Les extrêmes y seront (p. ex. : très riche ou très pauvre) probablement

- Comment gérer le consentement ?
  - Modèle peu adapté pour un apprentissage « partiel »









### Performance différentielle

- Disons que les cliniciens experts ont une sensibilité et une spécificité de 88 % pour une question clinique donnée...
  - Et que, lorsqu'ils se trompent, c'est au hasard, ça peut tomber sur n'importe qui.

- Est-ce qu'un modèle avec une spécificité de 93 % et une sensibilité de 94 % pour la même question, mais qui se trompe presque toujours pour un même groupe (p. ex. : jeunes, hommes, roux), serait:
  - Mieux ? Pire ? Acceptable ? Désirable ?









### Groupes de patients désavantagés ?

- Est-ce que toutes les variables sont nécessaires?
- Comment réagit le système en cas de données manquantes ?
  - Association entre données manquantes et groupes de patients

- Apprendre de l'existant
  - Risque de recréer le passer
    - HTA chez les femmes











Utile?

Ajustable pour une situation donnée ?

Ajustable pour un patient donné?

Sécuritaire ?

Éthique et légal ?

# Gestion équitable des impacts sur l'organisation des soins ?

Qui gagne et qui perd?











### Coût raisonnable?

- **Combien** de **variables** sont nécessaires ?
- Est-ce des variables déjà captées électroniquement naturellement ?
- Est-ce que les données sont disponibles au moment de l'utilisation de l'outil?
  - Toujours facile de « prévoir » ce qui se passe durant une hospitalisation... une fois que l'hospitalisation est terminée
- Comment est connecté l'outil aux systèmes informatiques de l'hôpital?













### Organisation des soins

Impact coûts et ressources

- Exemple : Analyse des Rx poumons et recommandation de scan
  - Sensibilité vs spécificité ?
  - Avalanche de scans ?
  - Retards pour d'autres indication ?

#### Documentation

- Qu'est-ce qui est stocké : résultats ? Intrants ? Version ?
  - Et si le modèle change toujours, comment faire un audit?











### **Gagnants et perdants**

- Qui va gérer les conséquences ?
  - Exemple : Fin de vie tout le monde va se mettre à consulter les soins palliatifs ?

- Est-ce que les gains cliniques sont pour ceux impactés au niveau des ressources ?
  - Exemple : héparine faible poids moléculaire









## Numérique en santé: pourquoi et comment?

#### 1. Tâches et défis pour l'IA

- IA vs statistiques
- Tâches de l'IA
- Défis de l'IA

#### 2. Les IA

- Apprentissage machine
- Apprentissage profond
- Réseau de neurones
- Modèles fondationnels

- 3. IA en cliniques : état des lieux
- 4. Outils numériques en clinique : desiderata
  - Utilité
  - Confiance et évaluation
  - Ressources











### Références additionnelles

- Smith, Barry (2023). ChatGPT: Not Intelligent. Ai: From Robotics to Philosophy the Intelligent Robots of the Future – or Human Evolutionary Development Based on Ai Foundations.
  - Systèmes complexes: narrow AI vs general AI
  - https://philpapers.org/archive/SMICNI.pdf
- Gradient descent, how neural networks learn
  - 3blue1brown: <a href="https://www.3blue1brown.com/lessons/gradient-descent">https://www.3blue1brown.com/lessons/gradient-descent</a>











# École d'été interdisciplinaire en numérique de la santé https://eins.griis.ca/



jf.ethier@usherbrooke.ca









