## Revues générales

# À quoi va servir l'intelligence artificielle en ophtalmologie ?

**RÉSUMÉ : L'intelligence artificielle est un domaine de recherche en pleine évolution dont les progrès récents vont modifier notre pratique en améliorant la rapidité, la sécurité et la précision de nos soins. Malgré l'ébullition médiatique qui accompagne chaque avancée dans ce domaine, les algorithmes actuels sont incapables de construire un raisonnement et ne possèdent pas de capacité d'abstraction. Leur mécanisme est purement mathématique et loin d'être exempt de biais.**
**La compréhension de la méthodologie du *machine learning* et la démystification du fonctionnement des algorithmes sont utiles aux médecins car elles peuvent leur permettre de garder un regard critique sur les publications scientifiques, ainsi que sur les logiciels et les services qui sollicitent de plus en plus leur attention dans ce domaine.**

**G. DEBELLEMANIÈRE**
Fondation Ophtalmologique
A. de Rothschild, PARIS.

Les médecins seraient-ils voués à disparaître, remplacés par des algorithmes surentraînés, techniquement supérieurs, infaillibles et infatigables, mis quotidiennement à jour des nouvelles avancées scientifiques ? Comment trier ce qui relève de la réalité de ce qui tient du roman d'anticipation ?

Cette mise au point vise à lister les progrès récents en intelligence artificielle (IA), à en définir les limites encore insurmontées, puis à tenter d'extrapoler de ces enseignements les grandes lignes des changements que nous, ophtalmologistes, devrons intégrer dans notre pratique dans un futur proche.

### De quoi parle-t-on ?

Si chacun a une idée personnelle assez précise de ce que signifie "intelligence artificielle", il n'existe étonnement aucune définition consensuelle de ce terme [1]. Le cours *What is AI*? donné par le pionnier de cette discipline John McCarty [2] propose cette dernière : "*l'IA est le domaine scientifique et industriel visant à la création de machines intelligentes, et en particulier de programmes informatiques intelligents*", l'intelligence étant définie par le même auteur comme "*la partie calculatoire de la faculté permettant d'accomplir un objectif*".

Selon le contexte, l'IA sera vue comme mimant parfaitement l'être humain ou au contraire comme lui étant supérieure par une exhaustivité de connaissance, une plus grande rapidité et précision, et/ou une plus grande objectivité. Les facultés d'apprentissage, d'adaptation, de résolution de problème et/ou de perception de stimuli extérieurs sont en général sous-entendues par le terme intelligence artificielle.

Plus pragmatiquement, les avancées récentes dans le domaine de l'IA relèvent de l'apprentissage automatique (*machine learning*), qui n'est pas une discipline nouvelle et dont la finalité

consiste à permettre à un programme de construire des règles décisionnelles à partir de données, sans que ces règles ne soient explicitement programmées par l'humain. Derrière cette définition intimidante se cache une réalité pas toujours complexe : l'algorithme de *machine learning* le plus basique est, tout simplement, la régression linéaire, qui permet d'inférer une règle (le coefficient directeur et l'ordonnée à l'origine de la droite de régression) à partir de données.

L'apprentissage profond (*deep learning*) constitue une famille d'algorithmes particuliers d'apprentissage automatique constitués de plusieurs couches de réseaux de "neurones", ces derniers étant des nœuds de calcul prenant plusieurs entrées et leur appliquant des poids et une fonction pour générer une valeur de sortie (***fig. 1***). La fonction appliquée à la somme des entrées multipliées par les poids est nommée "fonction d'activation" et détermine l'activation ou pas du neurone et l'amplitude de cette activation.

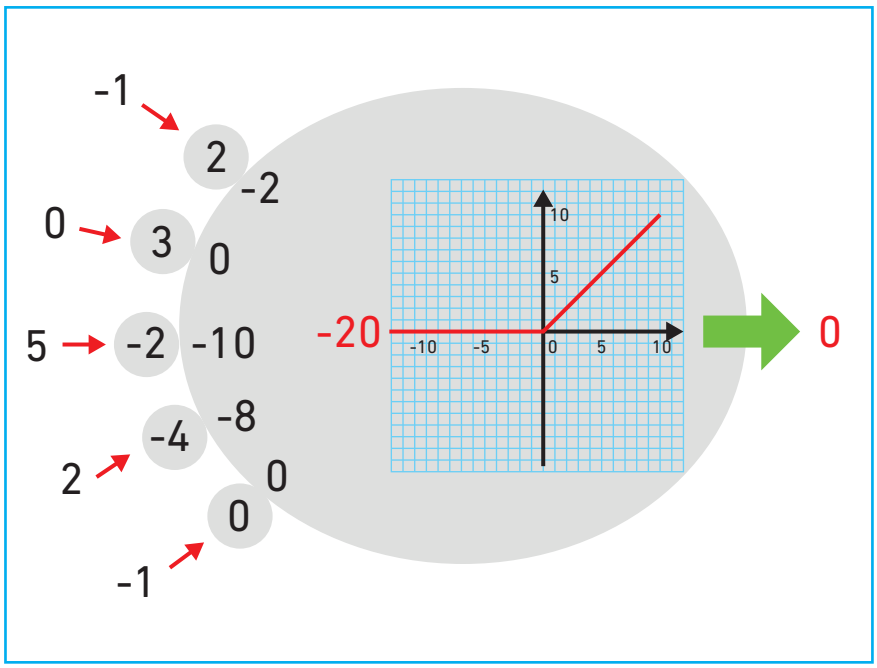


**Fig. 1 :** Cet exemple illustre le fonctionnement d'un neurone possédant 5 entrées et une fonction d'activation ReLU (*Rectified Linear Unit*). Les entrées sont multipliées par les poids correspondants, puis la somme du résultat est entrée dans la fonction d'activation. La fonction ReLU retourne un résultat nul quand son entrée est ≤ 0 et retourne la valeur d'entrée quand elle est > 0.

## Quelles avancées ont eu lieu ces dernières années ?

Les avancées en vision par ordinateur et en traitement du langage permises par l'avènement des réseaux de neurones convolutifs (*Convolutional Neural Network* [CNN]), qui sont une forme d'architecture de réseaux neuronaux utilisant le *deep learning*, ont été spectaculaires depuis 2012 et sont responsables du boom actuel de l'IA (***fig. 2***) [3, 4]. À tel point que les termes "intelligence artificielle" et "*deep learning*" sont souvent confondus par le grand public.

La puissance de cette technologie réside principalement dans sa capacité à faire évoluer ses filtres décisionnels de manière autonome, rendant inutile la phase de sélection de paramètres (*feature selection*) et de création de paramètres (*feature engineering*), indispensables dans les algorithmes plus classiques de

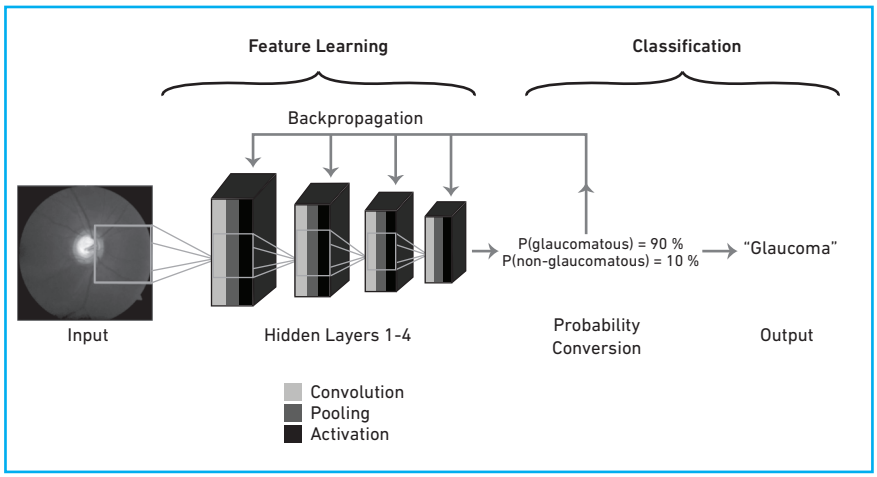


**Fig. 2 :** Architecture d'un réseau convolutif utilisé pour le diagnostic du glaucome sur rétinophotographies [5].

*machine learning*. L'architecture multicouche des réseaux neuronaux profonds permet l'organisation hiérarchique de leurs apprentissages, des concepts de complexité croissante étant représentés au fur et à mesure des couches. Il est possible de visualiser les zones d'intérêt retenues par le modèle grâce à des *heat maps*, mettant en évidence sur l'image analysée les anomalies qu'il détecte (***fig. 3***).

Ces particularités permettent une augmentation drastique de la précision des tâches de reconnaissance d'objets dans des images, de reconnaissance vocale ou d'analyse de texte, sous réserve que l'on dispose d'un set d'entraînement

# Revues générales

très important (de l'ordre de la dizaine de milliers d'exemples) pour entraîner l'algorithme. Des applications médicales des CNN ont été publiées, comme la reconnaissance de mélanomes à partir de photographies [7] ou le dépistage de la rétinopathie diabétique [8].

La seconde avancée spectaculaire récente dans le domaine de l'IA a eu lieu en 2016 avec la victoire d'AlphaGo (Google DeepMind) contre le 18 fois champion du monde de go Lee Sedol, grâce à une technologie d'apprentissage par renforcement (*reinforcement learning* [RL]). Contrairement au jeu d'échecs, le jeu de go ne peut être résolu par la puissance de calcul brute en étudiant toutes les possibilités de coup à partir d'une situation donnée et en choisissant la meilleure, en raison du nombre de combinaisons de jeu quasiment illimitées sur un goban et d'une grande complexité stratégique. L'équipe de DeepMind a entraîné dans un premier temps un réseau convolutif à jouer au go grâce aux coups décidés par des joueurs professionnels sur 30 millions de combinaisons de jeu, puis en le faisant jouer contre lui-même, ses règles de décisions étant progressivement modifiées et améliorées par la découverte de l'issue de chaque jeu [9]. Le programme a été perfectionné en 2017 [10], améliorant encore ses performances et n'ayant plus besoin d'exemples humains pour débuter son apprentissage.

L'apprentissage par renforcement a également permis à l'équipe de DeepMind de créer des programmes apprenant à jouer à certains jeux vidéo, intégrant progressivement les règles par essai/erreur et développant des stratégies pour gagner. L'apprentissage par renforcement permet donc à des algorithmes de deviner les règles de situations complexes et de développer des stratégies pour gagner… sous réserve qu'il soit possible de les entraîner des millions de fois à évoluer dans cet environnement et à leur donner un retour sur la qualité de leurs décisions.

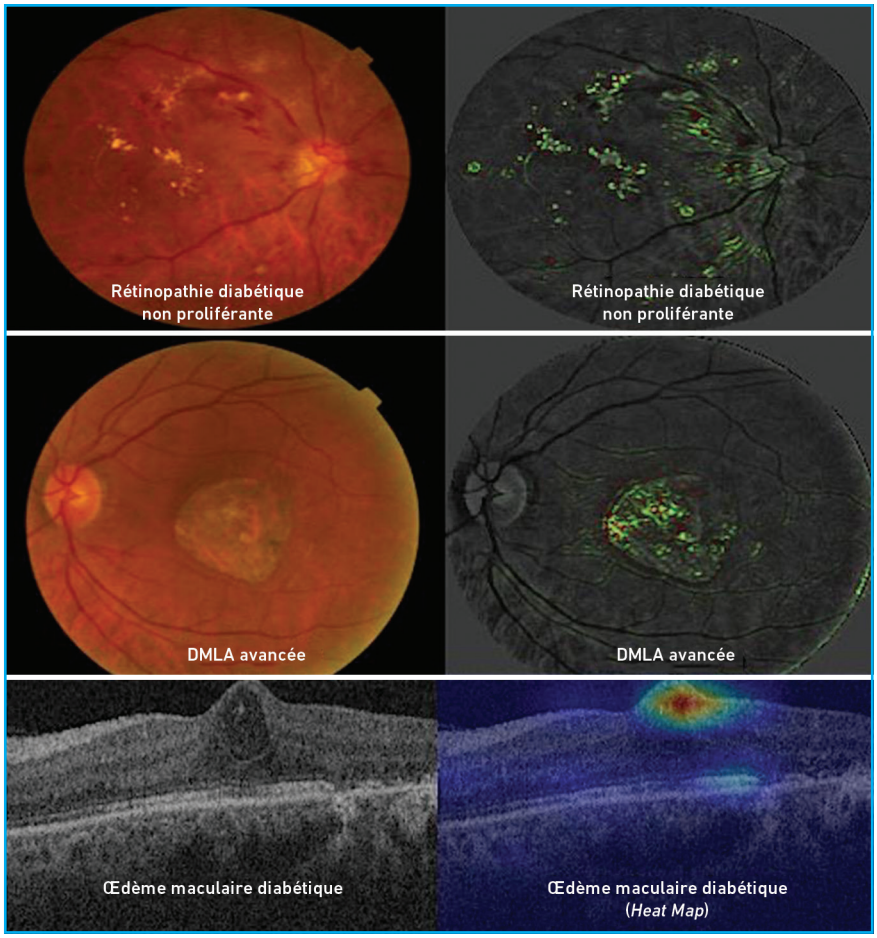


**Fig. 3 :** Les *heat maps* permettent de visualiser les zones d'intérêt retenues par le modèle de *deep learning* dans son processus de classification. Ce procédé permet de relativiser la notion de "boîte noire" souvent associée à cette famille d'algorithmes [6].

## Quelles sont les limites actuelles de l'intelligence artificielle ?

Dans le domaine intéressant particulièrement les ophtalmologistes de l'analyse d'images, le premier point limitant concerne le volume très important de données nécessaires à l'entraînement des modèles. Pour être utilisables, ces données doivent être étiquetées, c'est-à-dire associées au diagnostic que l'on souhaite faire apprendre au modèle. Pour atteindre des performances humaines de classification de la rétinopathie diabétique, l'équipe de Google a dû rassembler un *dataset* de 118 419 rétinographies pathologiques et 33 246 rétinographies saines, gradées par 54 ophtalmologistes [7]. On peut concevoir que la conception d'un modèle global de diagnostic du fond d'œil, comprenant les diagnostics rares, soit un chantier titanesque, même s'il est théoriquement possible.

Bien que potentiellement très performants dans la tâche pour laquelle ils ont été conçus, aucun algorithme n'est capable d'apprendre de manière autonome, ni d'évoluer seul vers d'autres apprentissages. Loin d'être indépendants de l'humain, la conception, les choix architecturaux, le contrôle de la validité des résultats et de la bonne généralisation du modèle nécessitent une intervention humaine.

## POINTS **FORTS**

- L'avènement de l'apprentissage profond (*deep learning*) a permis à des modèles d'intelligence artificielle d'atteindre des performances jusqu'ici inégalées en reconnaissance d'image ou dans certains jeux comme le go.
- Il convient de rester critique vis-à-vis de ces avancées : les modèles d'IA sont très dépendants de l'humain dans leur conception et leurs évolutions et ne sont pas exempts de biais.
- Loin de le remplacer, les progrès de l'IA vont apporter au médecin un gain d'efficience, de précision et de sécurité.
- La compréhension de la méthodologie du *machine learning* et du fonctionnement des algorithmes est importante pour anticiper les changements à venir et préparer notre profession aux avancées apportées par l'intelligence artificielle.

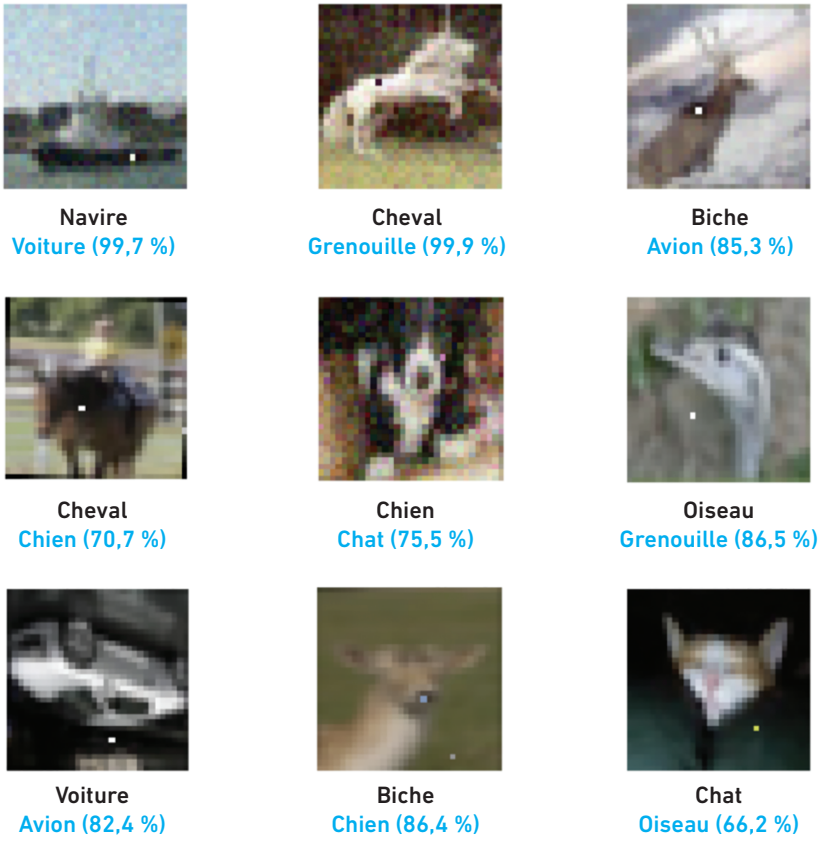


**Fig. 4 :** La suppression d'un seul pixel dans une image peut entraîner un réseau convolutif à commettre des erreurs de classification, avec un indice de confiance dans le résultat très élevé [13].

Par ailleurs, les modèles actuels d'intelligence artificielle ne peuvent exécuter que la tâche pour laquelle ils ont été entraînés : un outil de dépistage par IA de la rétinopathie diabétique sur rétinophotographies n'est ni performant dans d'autres pathologies, ni forcément transposable à d'autres type de machines et encore moins à d'autres types d'examens.

Il est important de noter l'absence de sens commun des algorithmes de *machine learning*, y compris des réseaux neuronaux. L'intelligence artificielle actuelle ne possède aucune représentation concrète du monde extérieur ou de la tâche à accomplir [11] et les réseaux convolutifs, basant leur décision sur des filtres mathématiques, peuvent changer drastiquement "d'avis" suite à la modification d'un seul pixel de cette dernière ou *via* l'insertion de gradients colorimétriques imperceptibles pour l'humain [12] (***fig. 4***). Cela entraîne des réponses parfois aberrantes des modèles d'IA et limite fortement la possibilité, en l'état actuel des connaissances, d'un système de décision totalement autonome dans le domaine médical. Il serait en effet inacceptable de constater, *a posteriori*, qu'une erreur faite par un modèle d'IA n'aurait jamais été commise par aucun médecin tellement elle apparaît énorme pour n'importe quel observateur.

L'absence de compréhension de la réalité par les algorithmes, qui ne font qu'appliquer des règles mathématiques (fussent-elles complexes) à des données numériques n'ayant aucun sens en elles-mêmes, explique les écarts de performance parfois spectaculaires d'un modèle d'IA entre le jeu de données sur lequel il a été entraîné et un nouveau jeu de données indépendant. En l'absence de validation externe d'un modèle, il n'est pas possible de conclure quant à sa performance.

Concernant le domaine de l'apprentissage par renforcement, l'entraînement des algorithmes nécessite la répétition, des millions de fois, de la tâche à accomplir, l'issue positive ou négative de la

# Revues générales

partie permettant l'apprentissage. Ceci n'est évidemment pas réalisable dans la plupart des tâches de la vie réelle.

## Que peut-on déduire de ces éléments pour notre pratique future ?

Si l'on extrapole à notre spécialité les notions abordées précédemment, on peut, de manière très subjective, suggérer au lecteur les éléments suivants :

>>> Des outils d'aide au diagnostic, de nouveaux scores, des indices de dépistage plus performants vont voir le jour dans le domaine de l'interprétation des rétinographies et dans les autres modalités d'imagerie (OCT, champ visuel, topographie cornéenne…).

>>> De même, certaines prises de décisions thérapeutiques dans des contextes complexes pourront être aidées par des modèles d'IA entraînés à résoudre ces situations spécifiques.

>>> Ces modèles, bien que plus complexes et performants que les scores actuels, n'en différeront pas par essence et souffriront des mêmes biais (par exemple, de la population sur laquelle ils ont été entraînés).

>>> Ces outils permettront une augmentation de la qualité des soins en diminuant les erreurs humaines dans des applications très précises, mais nécessiteront une validation médicale.

>>> Le diagnostic semi-automatisé des examens et l'analyse sémantique des dossiers médicaux permettront la production de comptes rendus et de courriers pré-rédigés à valider par le médecin, réduisant nos besoins en secrétariat. D'une manière générale, les applications générant un gain de temps ou d'efficience représenteront probablement une grande part des avancées.

>>> On peut supposer que, dans l'hypothèse d'une généralisation massive des outils d'IA appliquées à la médecine, les gains d'efficacité pourront permettre de diminuer le nombre de médecins nécessaires à une population donnée. Les ophtalmologistes de demain seront responsables d'un plus grand bassin de population, plus chirurgicaux, plus spécialistes.

>>> La problématique de la qualité des données cliniques recueillies, de leur structuration adéquate et de leur sécurité va prendre une place de plus en plus importante dans notre pratique quotidienne. La conception des logiciels médicaux va en être impactée et l'utilisation de ces données (par qui, comment et pourquoi) va faire l'objet de débats auxquels la communauté médicale doit être préparée.

## BIBLIOGRAPHIE


1. Wang P. What do you mean by "AI"? *Frontiers in Artificial Intelligence and Applications*, 2008;171:362-373.
2. Mccarthy J. What is Artificial Intelligence? 2007. www-formal.stanford.edu/jmc/whatisai/
3. Torres J. Why now this Artificial Intelligence boom? *Towards Data Science*, 2018. towardsdatascience.com/why-now-this-artificial-intelligence-boom-b50a35713090
4. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*, 2012;25:1097-1105.
5. Mayro EL, Wang M, Elze T *et al.* The impact of artificial intelligence in the diagnosis and management of glaucoma. *Eye*, 2019 [Epub ahead of print].
6. Ting DSW, Pasquale LR, Peng L *et al.* Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*, 2019;103:167-175.
7. Esteva A, Kuprel B, Novoa RA *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 2017;542:115-118.
8. Gulshan V, Peng L, Coram M *et al.* Development and validation of a *deep learning* algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 2016;316:2402-2410.
9. Silver D, Huang A, Maddison CJ *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature*, 2016;529:484-489.
10. Silver D, Schrittwieser J, Simonyan K *et al.* Mastering the game of Go without human knowledge. *Nature*, 2017;550:354-359.
11. Vincent J. Facebook's head of AI wants us to stop using the Terminator to talk about AI. *The Verge*, 2018. www.theverge.com/2017/10/26/16552056/a-intelligence-terminator-facebook-yann-lecun-interview
12. Moosavi-Dezfooli S, Fawzi A, Frossard P. DeepFool: a simple and accurate method to fool deep neural networks. *IEEE Conference on CVPR*, 2016:2574-2582.
13. Su J, Vargas DV, Kouichi S. One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, 2019:1.


L'auteur a déclaré ne pas avoir de conflits d'intérêts concernant les données publiées dans cet article.