



Ambroise **Baillifard**
UniDistance (Suisse)

L'intelligence artificielle : miroir de l'apprentissage humain ?

<http://dx.doi.org/10.18162/fp.2025.a353>

CHRONIQUE • Numérique en éducation

Peut-être serez-vous d'accord pour dire que les personnes intéressées par l'intelligence artificielle (IA) en éducation appartiennent à deux tribus bien distinctes. Les premières utilisent l'IA pour enrichir les processus d'apprentissage, sans forcément chercher à comprendre ce qu'il y a « sous le capot ». Pragmatiques, elles ont développé des compétences empiriques pour générer du contenu, personnaliser l'apprentissage ou automatiser des tâches. Elles savent qu'un modèle de langage (tel que Chat GPT) fonctionne par prédiction du mot suivant et parviennent à l'exploiter lorsque c'est pertinent, conscientes des enjeux sociétaux et environnementaux du numérique (Baillifard et al., 2024).

L'autre tribu parlait de GPT déjà en 2018. Ses membres, familiers avec des concepts spécialisés comme *Gating Networks* ou *Fine-Tuning*, peuvent discuter des implications du test de Turing ou de la conférence pionnière de Dartmouth (1956). Maîtrisant la programmation de réseaux de neurones et l'optimisation des performances d'un modèle, ces technologues peuvent entraîner leurs propres modèles sur des corpus éducatifs.

Loin d'être antinomiques, ces deux postures se complètent et s'enrichissent (Crompton et Burke, 2023). Ce n'est pas surprenant puisque c'est la rencontre entre sciences formelles (mathématiques, informatique, logique) et humaines (philosophie, neurosciences, linguistique) qui a permis à des machines d'imiter des processus tels que l'apprentissage, l'adaptation ou l'autocorrection (Popenici et Kerr, 2017). Ainsi, l'IA surgit après des décennies d'essais-erreurs pour mieux approcher les fonctions cognitives humaines et, naturellement, quand on plonge dans son univers, on aperçoit, en filigrane, les principes fondamentaux de l'apprentissage.

Répétition, erreur, rétroaction : piliers de l'apprentissage

Qu'il s'agisse d'un cerveau ou d'une IA, l'apprentissage repose sur un processus itératif où les connexions neuronales s'ajustent progressivement pour produire des réponses pertinentes. Un neurone, qu'il soit biologique ou artificiel, s'active ou produit une réponse lorsqu'un signal dépasse un seuil d'activation.

Dès 1957, l'informaticien et psychologue Rosenblatt conçoit le perceptron, un modèle de neurone artificiel capable de séparer des informations en deux catégories. Pour y parvenir, le perceptron ajuste l'importance de chaque connexion et le seuil à partir duquel il s'active. Le perceptron pourrait être utilisé pour filtrer des courriels en spam ou non-spam selon la présence de certains mots-clés. Cependant, cette approche présente la limite fondamentale de ne résoudre que des problèmes linéaires, c'est-à-dire ceux où une séparation entre deux catégories est possible. Il faut attendre 1986 pour qu'Hinton et Rumelhart popularisent l'algorithme de rétropropagation du gradient qui permet l'entraînement des réseaux de neurones multicouches. Grâce à cette technique, le réseau ajuste automatiquement ses paramètres après chaque itération d'apprentissage, en cherchant à améliorer progressivement ses prédictions.

Ce principe (*deep learning*) marque l'émergence de machines capables d'apprendre par elles-mêmes. Il reflète des aspects universels de l'apprentissage, tels que le besoin de répétition ainsi que les rôles de l'erreur et de la rétroaction. Tout comme un processus itératif permet à un réseau de neurones artificiels de s'ajuster, un élève apprend en expérimentant, en se trompant et en corrigeant ses réponses grâce à des retours fréquents et pertinents (Baillifard, 2024).

Diversité et transfert essentiels pour émanciper

Si l'apprentissage profite du couple essais-erreurs, il ne saurait s'y réduire. L'adaptabilité et le transfert des connaissances sont tout aussi essentiels.

Prenons l'exemple d'AlphaZero, qui a atteint un niveau surhumain aux échecs en jouant sans cesse contre elle-même. Cette recette faite de répétitions, d'erreurs, et de rétroactions ne lui permet pas de transférer ses compétences à un autre jeu, comme le poker. AlphaZero est enfermée dans son apprentissage, semblable à un élève récitant ses tables de multiplication sans avoir appris à multiplier : il serait en peine de trouver le résultat d'une multiplication qu'il n'aurait pas apprise par cœur.

En IA, quand les données d'entraînement manquent de variété, les modèles peinent à généraliser. Ainsi, certains algorithmes de diagnostic détectent moins bien les maladies de peau sur des patients noirs, car les images d'entraînement comportent principalement des peaux claires. Pour la même raison, des systèmes de reconnaissances faciales ont montré un taux d'erreur de 34 % sur les visages de femmes noires, contre seulement 0,8 % pour les hommes blancs (Buolamwini et Gebru, 2018).

Une manière de remédier à ces limitations consiste à exposer les modèles à un large éventail de situations, car la diversité permet le transfert des compétences d'un contexte à un autre. Quand il s'agit d'apprentissage humain, la variété prend de multiples formes : pluralité des médias, des types de tâches, des modalités d'apprentissage, des sources d'information, des modes de raisonnement, des contextes d'application, des niveaux d'analyse, etc.

Encadrer pour dépasser l'illusion du savoir

Toutefois, que ce soit pour l'humain ou la machine, la capacité de transfert ne garantit ni la compréhension ni la pensée critique.

En 2016, Microsoft a lancé Tay, un chatbot conçu pour apprendre en interagissant avec les utilisateurs d'un réseau social. En 24 heures, il est devenu haineux, raciste et conspirationniste. Pourquoi ? Parce qu'il absorbait sans filtre tout ce qu'on lui disait, sans discernement du vrai, du faux, du pertinent ou de l'inacceptable. Privé d'une base structurée de connaissances, son apprentissage était mimétique, donc dépourvu d'analyse ou de discernement.

Ce problème ne s'est pas dissous avec l'évolution des modèles. Par exemple, GPT-3 pouvait générer des absurdités, comme l'affirmation selon laquelle «les vaches pondent des œufs». Son mode de fonctionnement repose sur la prédiction probabiliste des prochains mots, sans validation intrinsèque de la cohérence ou de la véracité. Que ce soit pour une IA ou un humain, l'accumulation d'informations mène davantage au mimétisme qu'à la pensée critique.

Devant ces défis, les développeurs ont progressivement amélioré l'encadrement de l'apprentissage de leurs modèles, à l'instar d'enseignants ajustant leurs méthodes éducatives. Premièrement, ils ont affiné les bases de données utilisées pour le préentraînement (excluant les sources peu fiables). Ensuite, ils ont intégré des retours humains de plus en plus fins pour optimiser l'apprentissage des modèles. À l'apprentissage supervisé classique – où le modèle est entraîné sur des données annotées (comme des radiographies médicales) – s'ajoute une méthode (*RLHF*) où plusieurs réponses générées par le modèle sont évaluées par des humains, permettant au modèle de s'ajuster aux attentes humaines. Enfin, ils ont intégré des bases de connaissances externes auxquelles le modèle peut se référer avant de répondre à un utilisateur (Lewis et al., 2020). Par exemple, Perplexity AI est couplée à des sources vérifiées (articles scientifiques) afin de vérifier ses réponses avant de les produire.

Qu'il soit artificiel ou humain, l'apprentissage, plus que l'acquisition du savoir, concerne la capacité à identifier des sources fiables et à structurer l'information. Il s'agit moins du savoir que de la volonté de savoir.

L'apprentissage artificiel est humain

L'analogie humain-machine permet, nous l'avons vu, de mettre en lumière des piliers essentiels à l'apprentissage.

- L'effort, la répétition, l'erreur et la rétroaction nourrissent un apprentissage s'affinant par ajustements successifs. Or, ces ingrédients, à eux seuls, peuvent conduire au dressage, à l'endoctrinement ou aux croyances erronées (on parlera d'hallucinations pour l'IA).
- La qualité et la diversité des contenus, la confrontation des sources et la capacité à en évaluer la fiabilité permettent un apprentissage véritable. Sans cela, on ne peut attendre de l'élève (ni de la machine) qu'il dépasse ce qui lui est donné. Cependant, si on entraîne la machine pour qu'elle fasse la machine, «on n'éduque pas l'enfant pour qu'il le reste» (Reboul, 2010, p.26). Le sens de l'éducation c'est justement que l'élève dépasse ses maîtres, c'est-à-dire qu'il se libère de ce qui l'empêche d'être soi (Baillifard et Bonvin, 2023).

Terminons en remarquant que ce n'est pas l'IA qui apprend, mais des humains qui façonnent sa capacité à apprendre. Ce sont eux qui ajustent des architectures logicielles, qui affinent des modèles, qui définissent ce qui est appris. Grâce à leurs réseaux multicouches, Hinton et Rumelhart (1986) ont prolongé le travail de Rosenblatt, décuplant les capacités de la machine. Si l'IA reflète le progrès des connaissances de l'humanité, elle ne l'incarne pas. L'apprentissage véritable se situe là où il y a possibilité de dépassement ou d'émancipation. Profondément humain, il est câblé à notre nature émotionnelle : à la passion pour ce qui nous fascine, à la curiosité pour ce que nous ignorons et à la simple joie d'apprendre.

Références

- Baillifard, A. (2024). Apprendre : une aventure. *Cahiers pédagogiques*, 594(5), 47-49. <https://doi.org/10.3917/cape.594.0047>
- Baillifard, A. et Bonvin, M. (2023). Ni plagosus Orbilius, ni pur enseignement positif. Qu'est-ce qu'enseigner ? *Médiations et médiatisations*, 13. <https://doi.org/10.52358/mm.vi13.361>
- Baillifard, A., Ertz, O., Lecorney, S. et Martarelli, C. (2024). Mythes écologiques du numérique. *Formation et profession*, 32(1), 13. <https://doi.org/10.18162/fp.2024.a322>
- Buolamwini, J. et Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 77-91. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Crompton, H. et Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20(1), 22. <https://doi.org/10.1186/s41239-023-00392-8>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S. et Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP Tasks. *Arxiv*. <https://doi.org/10.48550/ARXIV.2005.11401>
- Popenici, S. A. D. et Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(1), 22. <https://doi.org/10.1186/s41039-017-0062-8>
- Reboul, O. (2010). *La philosophie de l'éducation*. PUF. <https://doi.org/10.3917/puf.rebou.2010.01>
- Rumelhart, D. E., Hinton, G. E. et Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533-536. <https://doi.org/10.1038/323533a0>

Pour citer cet article

- Baillifard, A. (2025). L'intelligence artificielle : miroir de l'apprentissage humain ? [Chronique]. *Formation et profession*, 33(2), 1-4. <https://dx.doi.org/10.18162/fp.2025.a353>