

# L'éducation supérieure à l'ère de l'IA générative

Dave Anctil



L'année 2023 aura permis au milieu de l'éducation de réaliser l'état des avancées en intelligence artificielle (IA) grâce à l'application ChatGPT, devenue une sensation mondiale instantanée. Le réseau de l'enseignement supérieur comprend maintenant que des outils d'IA peuvent rédiger automatiquement des travaux et examens dans toutes les matières et disciplines. Des outils similaires, bien que moins performants, existent depuis plusieurs années. Ils étaient utilisés par un nombre croissant d'étudiants et d'étudiantes, souvent à l'insu des directions et du corps enseignant, mettant ainsi à mal l'intégrité intellectuelle et la justice scolaire. Maintenant que le sujet est médiatisé, il ne sera plus possible de pratiquer l'ignorance volontaire à propos de ces pratiques documentées dans la recherche sur l'intégrité académique.

Mais la grande accessibilité et la performance croissante des nouveaux systèmes algorithmiques occasionnent une perte de repères et des inquiétudes légitimes. Il ne fait aucun doute que l'IA transformera notre profession, et cela d'une manière difficile à imaginer. Les technologies de l'information et de la communication (TIC) ont déjà forcé des adaptations continues chez

les professionnels de l'éducation, en particulier depuis le tournant des années 1990-2000. Cependant les TIC n'ont pas changé fondamentalement notre rôle professionnel dans la transmission et l'évaluation des compétences et connaissances, ni non plus le rôle et les responsabilités des apprenants et apprenantes. Or, ce sont en partie ces rôles et conceptions traditionnels de l'enseignement que l'IA générative vient bouleverser.

Le présent article propose d'abord un rapide état des lieux vulgarisé de la technologie de l'IA générative. À la lumière de la recherche en intégrité académique et sur les pratiques numériques de la communauté étudiante, nous explicitons ensuite les principaux enjeux qui devraient retenir notre attention et orienter notre action. Nous aimerions offrir, en dernier lieu, notre perspective dans le débat plus large sur l'intégration des technologies d'IA en éducation supérieure.

## L'IA générative

La technologie de génération automatique de contenus (ou « IA générative ») existe depuis plusieurs

années. Si cette technologie est devenue un sujet d'actualité majeur en 2023, c'est à cause de l'application ChatGPT, rendue disponible gratuitement au grand public le 30 novembre 2022 par la société OpenAI<sup>1</sup>. S'en est suivi le début d'une course accélérée entre les sociétés technologiques qui rivalisent maintenant pour perfectionner ces systèmes et en faciliter l'accès, notamment à travers des applications comme ChatGPT, mais aussi les navigateurs et moteurs de recherche de Google et de Microsoft, et bientôt la suite Office 365 de Microsoft. Le nombre de nouvelles applications d'IA générative accessibles se compte déjà par milliers<sup>2</sup>.

<sup>1</sup> Le lectorat intéressé à en savoir plus sur le modèle de langage ou désireux de tester l'application ChatGPT peut consulter : [openai.com/blog/chatgpt].

<sup>2</sup> Les milliers d'applications d'IA accessibles au public, gratuitement ou par abonnement, sont maintenant répertoriées et classées par les utilisateurs et utilisatrices sur le site Futurepedia – The Largest AI Tools Directory [futurepedia.io].

Qu'est-ce que ChatGPT ? L'interface utilisateur de cette application est celle d'un agent conversationnel (ou *chatbot*). Il permet à un individu sans compétence technique particulière de générer du texte par des « invites » (ou *prompts*), c'est-à-dire des questions et suggestions rédigées en « langage naturel », comme le français et l'anglais, ou un mélange de langues, comme du code informatique ou différents langages mathématiques et scientifiques. Les applications les plus récentes permettent même d'introduire des images, des graphiques ou des fichiers de divers formats.

Les nouveaux systèmes d'IA générative ne traitent pas l'information comme un moteur de recherche conventionnel, par exemple comme un algorithme de classification basé sur l'association statistique de mots-clés pour fouiller une base de données ou l'Internet. Les entrées sont plutôt traitées par un système mathématique beaucoup plus puissant que les moteurs de recherche. Ce système fut développé pour comprendre la signification et le contexte d'une requête afin de prédire statistiquement les meilleures réponses possibles, en construisant ces dernières à partir de ses propres connaissances internes. À chaque requête, le système génère donc une sortie originale, que l'on peut à loisir préciser ou améliorer automatiquement par d'autres invites : « réduire le texte de moitié », « ajouter deux exemples », « appliquer une fonction », etc. Les applications ChatGPT et Playground (OpenAI), Bard et LaMDA (Google) ou Bing AI (Microsoft et OpenAI) ne sont que les exemples les plus performants et populaires de ces systèmes.

Avant d'aller plus loin dans l'explication, notons d'abord que les systèmes de traitement du langage naturel (ou NLP pour : *natural language processing*) existent depuis longtemps. Différentes méthodes permettent aux ordinateurs de traiter les langues. Nous les utilisons depuis plusieurs années pour la reconnaissance de la parole, la classification et l'analyse de textes, la traduction et la correction automatiques, l'analyse des contenus subjectifs (sentiments) et bien sûr la génération de contenus. Pour l'IA générative, il faut surtout admettre que ce n'est pas l'interface conversationnelle qui procure aux ordinateurs les capacités de comprendre le langage et de générer du contenu, mais bien ce que les chercheurs en IA appellent un « modèle génératif » d'IA. Il s'agit d'un logiciel très puissant, développé par apprentissage automatique sur des superordinateurs spécialisés, au coût de millions de dollars, pour être ensuite logé sur de très gros réseaux de serveurs donnant accès au modèle via des applications. Par exemple, le modèle GPT-3.5 (pour *generative pre-trained transformer version 3.5*) constitue un modèle particulier, mais il motorise plusieurs interfaces et applications différentes disponibles en ligne.

L'explosion récente de la performance des modèles génératifs fut rendue possible grâce à des innovations majeures dans le domaine de l'IA appelées « apprentissage profond » (*deep learning*), qui datent de 2017. Elles ont permis de développer de « grands modèles langagiers » (ou LLM pour *large language models*). Et ce sont ces derniers qui démontrent des capacités inouïes, notamment en battant des records dans les tests standardisés de performance en IA (*AI benchmarks*),

mais aussi en réussissant des examens universitaires et professionnels (nous y reviendrons).

Expliquer le fonctionnement des grands modèles langagiers sans utiliser un vocabulaire technique est difficile. Ces modèles se démarquent des autres systèmes d'IA parce qu'ils sont réellement capables de transformer les données d'entraînement en connaissances organisées, et non simplement d'assister un utilisateur par un algorithme de recherche. Cette différence est cruciale. Elle démarque l'apprentissage profond des autres algorithmes et approches de l'IA qui sont plus familiers au grand public.

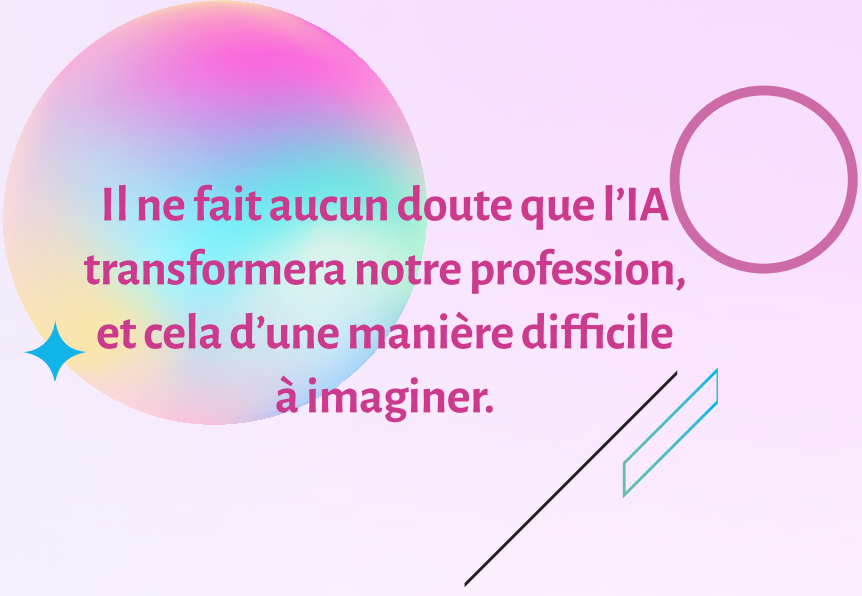
En effet, les modèles langagiers sont construits sur un réseau composé de milliards de neurones artificiels, dont l'organisation imite mathématiquement le fonctionnement biologique des neurones et synapses du cerveau. Pendant la phase d'entraînement, qui alimente ce réseau de neurones de milliards de données de divers formats (*training datasets*), une architecture programmée permet alors au réseau d'organiser lui-même des téraoctets d'informations en milliards d'objets et relations statistiques entre ces objets, sous la forme de vecteurs mathématiques. Le modèle est ensuite testé et perfectionné à l'aide d'une méthode d'apprentissage basée sur la rétroaction humaine appelée « RLHF » (pour *reinforcement learning from human feedback*). Cela permet de corriger des erreurs structurelles dans le modèle en fonction des finalités poursuivies par les ingénieurs-développeurs.

Une application générative de type LLM peut ainsi effectuer toutes les tâches susmentionnées, autrefois réalisées par des systèmes d'IA

distincts et spécialisés, à un niveau de performance en progression constante. De cette façon, les ingénieurs de l'IA produisent des modèles toujours plus performants, et ce, en augmentant la taille du réseau de neurones, la qualité des données d'entraînement ainsi que le peaufinage des méthodes de RLHF. Les applications en LLM peuvent désormais non seulement générer du texte, mais aussi coder, résoudre des problèmes mathématiques et scientifiques ainsi que concevoir des travaux dans la plupart des domaines techniques ou non techniques. Elles commencent même à générer de la musique et de la vidéo (OpenAI, 2023 ; Google Research, 2023).

Les générateurs d'images, tels DALL-E (OpenAI) et Midjourney (laboratoire indépendant), ont été développés comme des modèles distincts durant les dernières années. Cependant, en 2023, les bases de données textuelles, d'images et audiovisuelles ont été graduellement fusionnées pour donner lieu à de nouveaux modèles multimodaux, c'est-à-dire qu'un même modèle pourra désormais générer tous les types de contenus. C'est notamment le cas du modèle GPT-4 (OpenAI et Microsoft), qui est également capable de créer des applications et des sites Internet.

L'explosion de la performance des applications d'IA générative signifie qu'il sera désormais possible, et ce dès 2023, de leur confier la plupart des tâches cognitives enseignées et évaluées en éducation supérieure, de la création artistique textuelle ou sur support visuel, audio ou audiovisuel jusqu'au raisonnement philosophique, mathématique et scientifique, en passant par la programmation ou la planification stratégique (OpenAI, 2023).



**Il ne fait aucun doute que l'IA transformera notre profession, et cela d'une manière difficile à imaginer.**

Nous avons pleinement conscience du fait que l'affirmation précédente sur la cognition artificielle de l'IA générative paraîtra étonnante aux non-spécialistes. Elle est pourtant corroborée par des dizaines de laboratoires de recherche qui se spécialisent dans l'évaluation empirique de la progression des capacités cognitives des systèmes d'IA (Wei et collab., 2022). À titre d'exemple : bien que le modèle GPT-4 soit sorti le 14 mars 2023, une version préliminaire fut testée par les chercheurs de Microsoft Research. Leur rapport d'évaluation est fondé sur des dizaines de tests standardisés pour mesurer et comparer la cognition humaine avec celle des modèles. Il démontre que dans toutes les tâches mesurées, les performances de GPT-4 sont proches ou supérieures aux performances humaines, et dépassent largement le modèle antérieur GPT-3.5 (Bubek, 2023). Il nous

semble important de méditer sur la signification de la conclusion provisoire de cette première recherche de vaste ampleur portant sur la cognition des grands modèles de langage comme GPT-4 :

La démonstration centrale de notre recherche est que GPT-4 atteint une forme d'intelligence générale, montrant même des indices d'intelligence générale artificielle. Cela est démontré par ses capacités mentales centrales (telles que le raisonnement, la créativité et la déduction), l'éventail des sujets sur lesquels il a acquis une expertise (tels que la littérature, la médecine et le codage) et la variété des tâches qu'il est capable d'accomplir (par exemple, jouer à des jeux, utiliser des outils, s'expliquer...). [Notre traduction<sup>3</sup>] (Bubek, 2023, p. 92)

<sup>3</sup> Le passage original : « The central claim of our work is that GPT-4 attains a form of general intelligence, indeed showing sparks of artificial general intelligence. This is demonstrated by its core mental capabilities (such as reasoning, creativity, and deduction), its range of topics on which it has gained expertise (such as literature, medicine, and coding), and the variety of tasks it is able to perform (e.g., playing games, using tools, explaining itself, ...).

D'autres laboratoires indépendants évaluent présentement les capacités étonnantes de GPT-4, et nous apprendrons bientôt son étendue.

En dépit des recherches, d'aucuns ont du mal à s'imaginer que les systèmes d'IA puissent faire plus que mémoriser et répéter les productions issues de la cognition humaine. Mais ce n'est pas le cas. Les grands modèles comme GPT-4 ne possèdent pas en mémoire les productions culturelles humaines dans une base de données. Comme nous l'avons expliqué plus haut, il s'agit plutôt d'immenses modèles de connaissances organisés mathématiquement par le biais des techniques en apprentissage profond. Bien qu'entraînés sur les productions intellectuelles et culturelles humaines, ils génèrent bel et bien leurs propres processus cognitifs dans leurs productions de contenus. Le formatage humain des requêtes rédigées (*prompting*) peut cependant amener le modèle à imiter des styles humains, par exemple en composant une illustration « dans le style de Picasso » ou un poème ou une musique « dans le style de Baudelaire » ou « dans le style de Mozart ». Cependant le modèle ne contient pas ces œuvres dans une base de données ; il a plutôt assimilé des structures d'informations (*tokens*) des œuvres et des textes qui décrivent ces œuvres ou en discutent. En ce sens, il ne « copie » pas ces œuvres, mais génère une production originale, basée sur sa propre compréhension de ces œuvres, comme le feraient des créateurs humains s'inspirant des grands maîtres. Cela explique aussi pourquoi des experts en littérature, en arts visuels ou en composition musicale obtiendront de bien meilleurs résultats que les non-experts : en précisant les détails des requêtes à

partir de leurs propres connaissances des arts et des sciences, les utilisateurs du modèle langagier augmentent (statistiquement) la qualité de la prédiction du modèle et, par conséquent, la conformité du résultat obtenu au résultat désiré.

Il n'existe, à l'heure actuelle, aucun consensus sur le niveau d'intelligence des modèles langagiers puisqu'il n'existe aucun consensus scientifique sur l'intelligence comparée des espèces dans les sciences qui étudient ces questions. À titre d'exemple, les animaux non humains manifestent des formes d'intelligence très différentes dans leurs structures et fonctionnements par rapport à celle de l'humain. Pourtant, ces animaux sont capables de réaliser certaines tâches cognitives à un niveau comparable ou supérieur à celui de l'humain. La comparaison avec les systèmes d'IA est à la fois plus simple, puisqu'ils peuvent comprendre le langage, et donc passer des tests humains, mais aussi plus complexe, puisque leurs processus cognitifs ne sont pas biopsychologiques, et que l'interprétation du fonctionnement des plus gros réseaux de neurones artificiels est excessivement difficile.

Ce que les études ci-haut mentionnées mesurent, c'est donc la performance des modèles comme GPT-4 dans des tâches effectuées normalement par des humains, et non les processus cognitifs sous-jacents dans la résolution de ces tâches. La cognition artificielle (mathématique, informatique) est profondément différente de la cognition naturelle (biopsychologique), et les sciences cognitives appliquées aux deux genres de cognition commencent à peine à étudier les propriétés et fonctionnements des grands modèles développés en apprentissage profond.

Si nous revenons à notre sujet principal, l'éducation supérieure, nous devons reconnaître que les capacités des derniers modèles de l'IA générative sont maintenant capables de productions intellectuelles d'un niveau humain dans une foule de tâches, et que la progression de leurs capacités cognitives croît exponentiellement depuis une dizaine d'années.



## Enjeux pour l'éducation supérieure<sup>4</sup>

Les politiques d'intégrité intellectuelle des cégeps et universités sont fondées sur le principe que les étudiants et étudiantes doivent effectuer tout le travail intellectuel dans leurs évaluations. Cependant, la recherche en intégrité dans le milieu universitaire a montré que ce principe était dès lors dépassé par la réalité du numérique. En voici une preuve : une étude de 2009 concluait déjà que les outils logiciels de détection automatique de fraude scolaire seraient toujours en retard face à l'innovation et aux pratiques numériques permettant de les contourner (Bretag et Mahmud, 2009). Cette prédiction fut abondamment confirmée.

Donnons quelques exemples concrets, bien connus des chercheurs. Des outils numériques pour paraphraser automatiquement des extraits plagés, comme Rephraser, permettent depuis au moins cinq ans d'échapper

au traçage des détecteurs. D'autres outils d'IA augmentent énormément les capacités rédactionnelles, comme Grammarly et InferKit. Ils permettent à des rédacteurs médiocres d'atteindre un résultat élevé par la révision, la correction, la reformulation et la complétion assistées de leurs travaux, et cela sans plagier aucune source.

En ce qui concerne le plagiat de sources, il est depuis quelques années très aisé d'utiliser la traduction automatique pour contourner le dépistage du personnel enseignant et des logiciels de détection. La version neuronale de Google Translate (2016-) et de DeepL Translate (2017-), applications développées en apprentissage profond, permettent aux plagiaires de copier-traduire-coller des morceaux de textes pour les rendre indétectables avec un peu de camouflage artisanal ou par l'utilisation de Rephraser ou de Grammarly. L'étudiant tricheur peut même vérifier la traçabilité de son texte frauduleux avec Turnitin et Google Search, souvent directement à partir de ces applications...

La recherche récente sur l'intégrité à l'ère numérique devrait nous faire reconnaître la réalité suivante : un nombre important d'étudiants et d'étudiantes utilisent déjà une grande variété d'outils d'IA pour générer leurs travaux ou pour améliorer leurs performances scolaires, en toute impunité, et cela depuis plusieurs années (Ahsan, Akbar et Kam, 2022 ; Surahman et Wang, 2022 ; Eaton et Hughes, 2022).

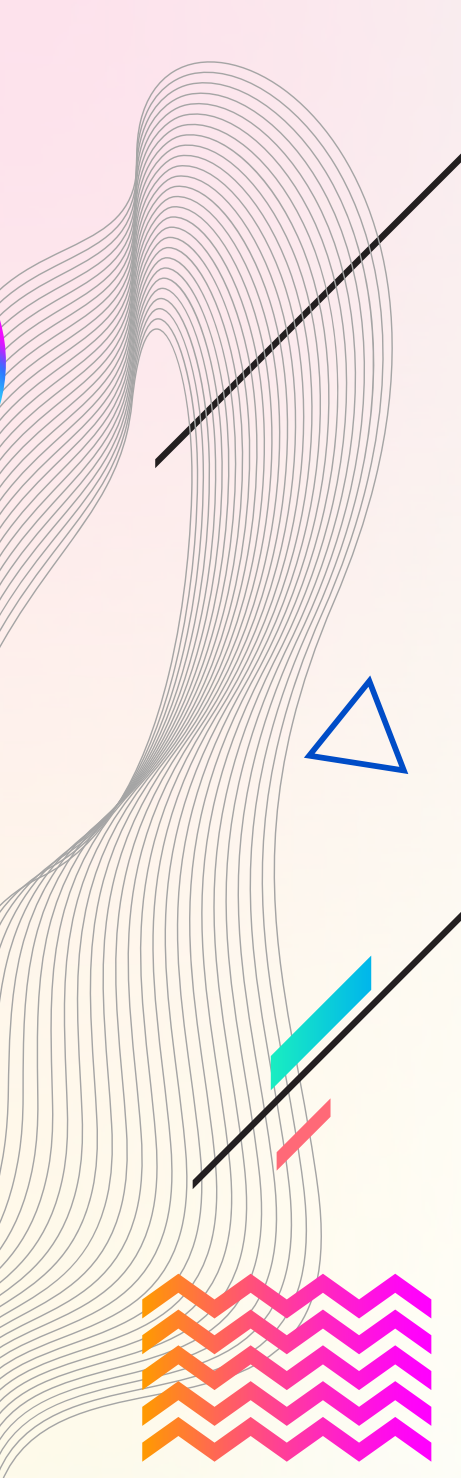
Reconnaissons aussi qu'il n'est pas du tout clair, en vertu des énoncés des règlements scolaires, que toutes ces pratiques sont frauduleuses ou malhonnêtes, puisque les contenus générés

automatiquement ne sont pas toujours des plagiat et que les outils en question sont rarement mentionnés dans la liste des pratiques interdites. Pour le dire franchement : l'enseignement supérieur n'a pas assez reconnu le défi posé par la progression extrêmement rapide des technologies numériques.

Par conséquent, face à « l'évènement ChatGPT », qui nous a fait abruptement entrer dans l'ère de l'IA générative, il est plus que temps que les cégeps et universités reconnaissent que nos politiques d'intégrité intellectuelle sont depuis longtemps en décalage avec la réalité technonumérique et les pratiques étudiantes. À notre avis, il ne suffira pas non plus d'ajouter les nouveaux outils d'IA générative à la liste des interdictions sanctionnées pour prévenir la malhonnêteté et pour développer l'intégrité intellectuelle dans la communauté étudiante.

De plus, chercher une « solution technique » à ce problème est voué à l'échec. Contrairement à une idée répandue, il est pratiquement impossible de détecter l'utilisation frauduleuse des nouvelles applications dans un travail scolaire. La raison est simple : le contenu généré par l'IA est un contenu original comparable à celui d'un rédacteur expert qui s'inspire de ses sources. Malgré les projets annoncés de logiciels détecteurs, fondés sur des méthodes statistiques, la détection ne pourra jamais offrir une prédiction hors de tout doute raisonnable ; elle demeurera toujours contestable. Et la détection des fraudes sera, comme par le passé, aisément contournable en suivant des recettes simples, disponibles dans les recoins d'Internet – ce qui est également vrai pour la pseudo solution des tatouages numériques (*digital watermarks*).

<sup>4</sup> J'enseigne uniquement dans des programmes préuniversitaires au niveau collégial, ainsi qu'au niveau universitaire. Par conséquent, je ne suis pas qualifié pour réfléchir aux enjeux des niveaux tant primaire, secondaire que collégial technique.



Il nous semble donc urgent que les directions et départements des cégeps et universités adoptent deux approches générales et complémentaires, que nous résumons ici avant de les expliciter :

- **L'approche par contournement** : protéger les principes de l'intégrité des évaluations, ainsi que l'égalité et l'équité scolaires, en rapatriant dès maintenant, dans l'environnement contrôlé de la classe, les évaluations les plus importantes pouvant mesurer les connaissances et l'atteinte des compétences.
- **L'approche par l'intégration** : doter rapidement le corps professoral d'un soutien pour recevoir une formation adéquate afin (i.) d'identifier les outils d'IA sécuritaires, éthiques et adaptés à leurs disciplines, puis (ii.) d'explorer les meilleures manières d'intégrer ces outils à l'enseignement, aux pratiques pédagogiques et aux évaluations.

L'approche par contournement permet, à court terme, de protéger l'intégrité, l'égalité et l'équité de l'évaluation scolaire. Mais elle a aussi un coût : elle pourrait diminuer la diversité des expériences d'apprentissage offertes à la nouvelle génération étudiante qui, rappelons-le, fut habituée à la « pédagogie par projets », en proposant des évaluations participatives, plus créatives et complexes qu'un simple examen (écrit ou oral). Bien entendu, l'enseignement supérieur moderne se traduit déjà par une grande variété d'évaluations actives, qui ne se limitent pas aux examens. On peut ainsi encourager différentes évaluations collaboratives et des laboratoires qui ne nécessitent pas l'utilisation du numérique,

mais qui sollicitent la cognition individuelle et collective de la communauté apprenante sans recourir à l'IA.

La deuxième approche permet de contrebalancer les désavantages de la première, en favorisant l'intégration des outils d'IA générative dans notre philosophie et nos pratiques éducationnelles. Cette intégration est naturellement plus en vogue chez les personnes conseillères pédagogiques TIC (CP-TIC) et les membres technophiles du corps professoral<sup>5</sup>. Elle consiste à créer des contenus et formats d'évaluation sollicitant un usage responsable et bénéfique, sur le plan des compétences et connaissances, des outils d'IA. Elle a également le mérite de familiariser pédagogues et personnes apprenantes à faire un usage intégré de ces outils, mais aussi de favoriser une utilisation responsable et de développer l'esprit critique, en vue d'une coopération harmonieuse des intelligences humaine et artificielle.

<sup>5</sup> Je recommande au personnel enseignant de s'autonomiser en consultant les nombreuses ressources en ligne, par exemple le site Eduactive, qui propose de tels contenus pédagogiques. Yann Houry, directeur de l'innovation pédagogique et technologique au Lycée français international de Hong Kong, propose des expérimentations et des capsules technopédagogiques novatrices et avancées, notamment à partir de sa page LinkedIn, que je recommande chaleureusement.

Donnons ici un exemple, à titre illustratif, que nous pouvons qualifier « d'approche par rétroaction conversationnelle » :

Apprendre à raisonner logiquement, notamment à bien expliquer et argumenter dans la rédaction de textes, est l'une des compétences les plus importantes et transversales de l'éducation supérieure. Mais cette compétence large est aussi difficile à acquérir, puisqu'elle demande beaucoup de pratique et que la personne enseignante ne dispose pas d'un temps suffisant pour superviser individuellement la construction grammaticale, syntaxique et logique des phrases, paragraphes et textes de chaque individu composant ses groupes. Les applications des modèles GPT-4 (OpenAI) ou Claude+ (Anthropic) peuvent produire une bonne rétroaction selon différents exercices rédactionnels. Il s'agit de donner des exercices avec requêtes préédigées afin d'encourager les étudiants et étudiantes à s'exercer avec l'IA, de manière autonome et encadrée, plutôt que de demander au système d'effectuer le travail à leur place. La personne apprenante rédige d'abord des paragraphes d'explication et d'argumentation, basés sur les contenus du cours.

Ensuite, elle entre chaque texte composé dans la boîte de dialogue. Enfin, elle demande à l'IA des requêtes du genre :

Voici un raisonnement [explicatif ou argumentatif dans telle matière] que j'ai composé. Je veux connaître les points forts et faibles de mon texte en vertu des [critères suivants...]. Je veux que tu m'expliques de manière détaillée mes erreurs, et que tu me donnes des suggestions pour améliorer mon raisonnement...

La rétroaction conversationnelle devient une forme d'assistance bien connue en tutorat, qui consiste à faire pratiquer les bases d'un apprentissage. Elle amène rapidement la personne apprenante à constater ses forces et faiblesses sans avoir peur du jugement ou de déranger. Il s'agit pour la personne enseignante de la guider en vue d'une utilisation métacognitive de l'IA. La personne apprenante obtient alors de l'assistance dans les défis individuels de son apprentissage des tâches et s'exerce à mieux comprendre ses erreurs pour s'améliorer. Les professeurs et professeures peuvent demander ensuite de recevoir le résultat de cette collaboration, sous la forme d'un document où la personne apprenante fait état de sa propre progression, pour évaluer les efforts et l'amélioration avant les examens passés sans assistance.

Nous n'avons pas l'espace pour développer d'autres exemples. Mais comprenons que la recherche pédagogique sur les bons usages de l'IA générative en est encore à ses balbutiements. Il faudra effectuer des études scientifiques pour connaître les pratiques les plus susceptibles d'améliorer l'apprentissage et la cognition, tout en protégeant l'intégrité des évaluations. En revanche, l'intégration de l'IA en éducation ne doit pas se faire sans se questionner philosophiquement et socialement sur l'éducation supérieure à l'ère de l'IA cognitive.

### Le débat sur l'intégration de l'IA au collégial

La perspective esquissée jusqu'ici s'oppose à deux conceptions dominantes dans le vaste débat sur la place de la technologie en éducation supérieure. Elle s'oppose à une conception néolibérale, tournée vers le clientélisme et l'adaptation béate aux exigences du marché (définies par les besoins à court terme des entreprises). Mais elle s'oppose aussi à une conception de l'éducation que nous pourrions qualifier de « romantique », laquelle

consiste à essentialiser les pratiques éducationnelles classiques en évoquant la nécessité d'une mobilisation du corps professoral contre le spectre de la « déshumanisation technologique ».

Pour mieux cerner notre propre perspective dans le débat, nous précisons ces critiques sans toutefois les approfondir en tenant compte de toutes les nuances.

La conception néolibérale n'est pas recommandable, non seulement parce qu'elle trahit le principe fondamental



de l'éducation supérieure en tant que bien public<sup>6</sup>, mais aussi parce qu'elle est contradictoire avec la réalité de l'évolution technosociale. Former des étudiants et étudiantes pour la « réalité » du marché du travail nous semble réductionniste et contreproductif, puisque la vitesse de transformation de cette dernière ne cesse de s'accélérer. Il est donc beaucoup plus avisé, de l'aveu même des meilleurs économistes spécialisés sur la question, d'offrir une formation intellectuelle robuste et polyvalente, favorisant le développement de meilleures capacités cognitives, d'adaptation et d'autonomisation des étudiants et étudiantes face aux défis et transformations socioprofessionnels accélérés par l'IA (Brynjolfsson, 2023).

*A contrario*, la conception militante-romantique suppose de considérer ce progrès technologique comme une « force déshumanisante » qu'il conviendrait de « combattre », en définissant l'éducation comme une contre-force « opposée à la destruction de la culture », « de ce qui fait l'humain, le commun, la nature, la vie en général » (Martin et Mussi, 2023, p. 9-10). Bien que nous souscrivions à la conception humaniste québécoise de l'éducation – former des citoyens et citoyennes, ainsi que des personnes autonomes,

capables d'exercer pleinement leurs libertés de pensée et d'expression –, le discours romantique évoquant la « déshumanisation technique » pour rejeter le numérique et l'IA nous semble servir des intérêts politiques militants, plutôt qu'éducationnels.

Les technologies algorithmiques redéfinissent la condition humaine comme l'ont fait d'autres grandes inventions de l'histoire, telles que l'écriture et l'imprimerie, la machine à vapeur ou l'électricité. Essentialiser l'éducation en érigeant ses traditions comme des remparts face au changement technosocial revient à imposer une seule conception doctrinaire de l'humain, au bénéfice d'un militantisme contradictoire avec la vocation première de l'éducation supérieure : la recherche libre de la vérité et la transmission ouverte des connaissances. Si la critique des technologies doit, bien entendu, avoir sa place dans cette recherche et cette transmission, nous n'avons pas le droit d'imposer une idéologie particulière, avec ses biais et ses limites intrinsèques, au prétexte de « protéger » la communauté étudiante contre la civilisation technologique à laquelle elle appartient de fait et de droit.

Comme pédagogues et spécialistes de nos disciplines, notre responsabilité première est de donner la meilleure éducation possible, en effectuant des choix fondés sur nos expertises, notre expérience et des études fiables. Bien sûr, nous ne pouvons pas ignorer non plus le fait que les multinationales derrière l'IA générative, comme Microsoft et Google, cherchent à réaliser des profits, et non à nous aider dans la mission qui nous habite comme professionnels et professionnelles de l'éducation.

Nous espérons ainsi contribuer à ouvrir un espace de débat en vue d'une « troisième voie » entre les approches néolibérale et romantique. La conception pragmatiste et démocratique que nous proposons vise, de fait, à protéger notre profession de l'influence excessive de tous les intérêts privés, idéologiques et militants, ainsi que d'une polarisation au sujet de l'IA qui risque de nuire aux grandes missions de l'enseignement supérieur comme bien public.

Cette conception reconnaît la place de l'autonomie individuelle et les bénéfices collectifs du numérique dans toutes les sphères de la vie sociale. Et il en va de même de l'IA, qui occupe et occupera des fonctions essentielles dans le développement humain avec non seulement des bénéfices importants, mais aussi des pertes et risques mal connus, que nous avons le devoir d'identifier le plus rapidement possible pour les prévenir et en minimiser les effets nuisibles. Très prochainement, nous allons contribuer à cette recherche et à cette veille avec les équipes spécialisées de l'Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA). Pour le moment, nous proposons, en guise de conclusion, trois grandes orientations en vue de l'intégration plus harmonieuse de l'IA en enseignement supérieur.

**Une formation continue sur le numérique et l'IA.** En grande majorité, le personnel enseignant du réseau collégial ne possède pas les connaissances et compétences essentielles sur le numérique et l'IA pour les comprendre et se les approprier. Le ministère de l'Enseignement supérieur devrait investir pour développer un programme de formation conti-

<sup>6</sup> Comme définie dans la loi québécoise ou la Convention mondiale sur la reconnaissance des qualifications relatives à l'enseignement supérieur (UNESCO, 2019).

nue, conçu par des experts en IA et en technopédagogie, qui soit bien adapté aux besoins du réseau collégial. Un programme budgété de dégrèvement permettrait à nos professionnels et professionnelles de disposer de temps et de soutien pour acquérir des bases permettant ensuite de s'autonomiser dans leurs disciplines – qui, rappelons-le, sont toutes concernées.

### **Des assistants virtuels homologués.**

Comme nous l'avons illustré avec l'approche collaborative, les modèles de l'IA générative sont accessibles à toutes les organisations et offrent des opportunités inédites pour soutenir le travail du corps professoral. Bien que ces outils puissent fournir une certaine assistance, il ne s'agit pas d'agents logiciels experts conçus pour les tâches professorales ; ils sont souvent mal utilisés et parfois mal adaptés à l'éducation supérieure ; ils soulèvent également divers enjeux de sécurité que nous ne pouvons pas détailler ici. Ces outils servent également des intérêts économiques critiquables, notamment la création de nouveaux marchés pour les multinationales via des clientèles captives qui sont alors incitées ou même forcées à adopter leurs produits selon des dynamiques compétitives néfastes.

Il faudrait donc développer des agents d'IA spécialisés en éducation, capables de nous assister dans la correction, la supervision, la rétroaction et l'aide pédagogique. Cela libérerait du temps précieux pour les tâches valorisantes et stimulantes, comme l'innovation pédagogique, l'enseignement individualisé, la mise à jour des contenus et connaissances de nos cours ainsi que la recherche et la création. Qu'ils soient développés par l'État ou par des entreprises réglementées, les

assistants virtuels bénéfiques en éducation supérieure devraient toutefois être soumis à un processus d'homologation, permettant de garantir leur efficacité, leur sécurité et les bénéfices pédagogiques démontrés. Entretemps, les départements et les organisations des différents collèges et universités devraient rester vigilants et critiques face aux discours néolibéraux qui nous demandent de nous adapter béatement, au mépris de notre autonomie professionnelle et des finalités éducatives de nos institutions au service du bien commun de la société.

### **La collaboration des intelligences humaines et artificielles.**

En tant que chercheur en interaction humain-machine (IHM), il nous semble évident que l'arrivée rapide d'agents d'IA en éducation sera de plus en plus perçue comme une menace, donnant ainsi de l'eau au moulin à la perspective romantique. La recherche a bien documenté les réactions de méfiance et de défiance associées à l'anthropomorphisation d'agents artificiels dotés de capacités intellectuelles et langagières (Damiano et Dumouchel, 2018). Nos dispositions psychosociales nous poussent ainsi à considérer les agents robotiques et logiciels comme des « étrangers » et des « menaces » au cercle protégé de notre « humanité », plus particulièrement pour notre autonomie, notre dignité et notre sentiment de compétence. Cependant, la recherche montre également que la peur d'être remplacés ou supervisés par des machines diminue considérablement lorsque les individus apprennent à collaborer avec des assistants humbles et serviables, conçus pour s'intégrer harmonieusement dans les fonctionnements humains.

Pour augmenter les chances de parvenir à une collaboration intégrée des intelligences humaines et artificielles, pour créer un climat de confiance et d'ouverture face à ces technologies inédites et déstabilisantes, il faudra donc développer une stratégie tenant compte de ce grand défi et chantier. Surtout, comme le prescrit la *Déclaration de Montréal pour un développement responsable de l'intelligence artificielle*, échafaudée en 2017-2018, il est impératif que des principes comme l'autonomie, la responsabilité, le bien-être et l'inclusion soient pris au sérieux dans l'intégration de l'IA dans les milieux humains comme l'éducation. En ce qui nous concerne, professionnels et professionnelles de l'éducation, syndicats et gestionnaires, nous devons réclamer le droit de faire des choix réels au nom de notre mission fondamentale et de notre professionnalisme. La puissance de l'IA générative aura des conséquences positives et négatives sur la plupart des systèmes sociaux, et nous avons, ensemble, la charge de veiller à ce que l'éducation supérieure ne soit pas un simple terrain d'expérimentation et une source de profits pour les multinationales de la technologie. —

## Références bibliographiques

Ahsan, K., S. Akbar et B. Kam (2022). « Contract cheating in higher education: a systematic literature review and future research agenda », *Assessment & Evaluation in Higher Education*, vol. 47, n° 4, p. 523-539.

Bretag, T. et S. Mahmud (2009). « A model for determining student plagiarism: electronic detection and academic judgement », *Journal University Teaching & Learning Practice*, vol. 6, n° 1, p. 49-60.

Brynjolfsson, E. (2023). « The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence », *Dædalus*, Digital Economy Lab, Stanford University, 12 janvier, [En ligne].

Bubek, S. et collab. (2023). « Sparks of Artificial General Intelligence: Early experiments with GPT-4. », 22 mars, [En ligne].

Damiano, L. et P. Dumouchel (2018). « Anthropomorphism in Human-Robot Co-evolution », *Frontiers in Psychology*, vol. 9, p. 1-9.

Eaton, S. E. et J. Christensen Hughes (dir.) (2022). *Academic integrity in Canada: An enduring and essential challenge*, Singapore, Springer.

Google Research (2023). « MusicLM: Generating Music From Text », 15 mars, [En ligne].

Martin, E. et S. Mussi (2023). *Bienvenue dans la machine. Enseigner à l'ère numérique*, Montréal, Écosociété.

OpenAI (2023). GPT-4 Technical Report.

UNESCO (2019). *Convention mondiale sur la reconnaissance des qualifications relatives à l'enseignement supérieur*, Paris, France.

Surahman, E. et T.-H. Wang (2022). « Academic dishonesty and trustworthy assessment in online learning: A systematic literature review », *Journal of Computer Assisted Learning*, vol. 38, n° 6, p. 1535-1553.

Wei, J. et collab. (2022). « Emergent Abilities of Large Language Models », *Transactions on Machine Learning Research (TMLR)*, aout.



**Dave Anctil**, PhD. est professeur de philosophie et d'intelligence artificielle au Collège Jean-de-Brébeuf et chercheur affilié à l'Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA). Auparavant, il fut chercheur postdoctoral et chargé de cours à l'Université de Montréal et à l'Université du Québec à Montréal. Ses recherches et publications récentes portent sur l'évolution et les interactions de l'IA avec l'humain, dans une perspective d'évaluation des risques, bénéfiques et transformations psychosociaux. Il occupe également des fonctions d'éthicien, de vulgarisateur et de consultant expert auprès d'organisations et de médias d'information canadiens portant sur les enjeux éthiques et psychosociaux de l'IA et du numérique.

dave.anctil@brebeuf.qc.ca