

ChatGPT : la riposte doit être pédagogique

Jean-Philippe Boucher



Au moment d'écrire ces lignes, GPT-4 est disponible depuis moins de 48 heures et déjà nous savons qu'il peut réussir la majorité des tests standardisés et plusieurs examens d'ordres professionnels (OpenAI, 2023). En comparaison avec son prédécesseur GPT-3.5, lancé quatre mois plus tôt, soit en novembre 2022, la nouvelle mouture fait mieux pour la grande majorité des tests et obtient des résultats jusqu'à deux à trois fois plus élevés.

Nous ne pouvons que spéculer sur les capacités que les intelligences artificielles génératives (IAG) auront à la rentrée d'automne 2023. Voulons-nous vraiment embarquer dans une course à l'armement technologique contre Silicon Valley ? Et si notre saufconduit était pédagogique ? Dans cet article, je veux aborder le phénomène des IAG sous l'angle pédagogique, notamment pour l'épineuse question de la validité des évaluations.

La recherche est morte, vive la génération !

Clarifions d'abord rapidement ce qu'est GPT-4 avant d'aller plus loin. GPT-4 est le cœur de ChatGPT. C'est un modèle de traitement de langage qui a accès à une montagne de données et qui est en mesure de générer, à l'aide de cette masse d'informations, une réponse à une question précise. Ce qui rend GPT-4 impressionnant repose principalement sur deux aspects :

1. Il est en mesure d'analyser et de comprendre des questions complexes.
2. Il est en mesure de fournir une réponse dans une forme textuellement quasi parfaite, et le format de la réponse peut être ajusté selon nos besoins (nombre de mots, style d'écriture, etc.).

Concrètement, GPT-4 comprend des questions telles que « Peux-tu m'écrire un texte de 600 mots établissant un lien entre l'œuvre de Rousseau et le réchauffement climatique ? », puis il fournit des réponses respectant (approximativement) la longueur convenue et le sujet précisé. La qualité des réponses varie, mais dans beaucoup de cas, cette application est à même de produire une réponse ne nécessitant qu'un léger travail humain d'édition pour produire un travail scolaire acceptable.

GPT-4 est aussi doté d'une mémoire. Donc, si la réponse n'est pas parfaite, il est possible de raffiner la question avant d'effectuer le travail d'édition. Pour reprendre l'exemple sur Rousseau, nous pourrions demander d'ajuster la réponse afin que le texte soit rédigé à la première personne du singulier et d'ajouter trois citations ;

GPT-4 serait probablement apte à satisfaire aux deux demandes.

L'avancée majeure des IAG comme GPT-4, c'est que la réponse fournie est générée ; elle n'est pas simplement retrouvée dans les données. Il n'y a pas dans ses données la présence de textes préécrits de 600 mots portant sur chacun des sujets possibles. Le mouvement de fond qui s'opère est un remplacement probable des outils qui recherchent du contenu, comme un moteur de recherche ou Wikipédia, par des IAG qui génèrent du contenu sur demande. C'est un réel changement de paradigme.

L'insaisissable sentience

Selon ce qu'il est possible de savoir, car il est une propriété privée d'OpenAI, GPT-4 ne demeure qu'un programme

informatique qui dépend du savoir produit par l'humanité. Il n'est pas plus proche d'atteindre la sentience, soit la capacité de raisonner, que les premiers programmes d'intelligence artificielle du siècle dernier. Ainsi, il est totalement dépendant de sa montagne de données et est limité à générer du contenu à partir de celle-ci. Comme, GPT-4 n'est pas sentient, il n'a aucune métacognition et il ne peut pas se valider ou se corriger sans l'aide d'un humain. Il est aussi incapable d'avoir une pensée divergente du contenu sur lequel il s'appuie.

Autre caractéristique de GPT-4, il n'aime pas beaucoup dire qu'il ne sait pas. Alors, il lui arrive d'affirmer des choses totalement fausses en ce qui concerne soit des faits, soit des références. Afin de garder un côté mystique, les concepteurs d'IAG font un peu d'anthropomorphisme et nomment ces erreurs des « hallucinations ». Donc, GPT-4 n'est pas parfait et il se trompe régulièrement, mais comme il ne connaît pas le doute, il rédige ses propositions avec autant d'assurance que la plus évidente des vérités.

Bref, GPT-4, c'est comme si une personne analysait tout ce qui est disponible sur Internet pour un sujet donné et le synthétisait en un agrégat syntaxiquement parfait. Ce sera beau, lisse, recherché et unique dans sa formulation. Mais s'il y a de l'inédit, ce sera dû à des hallucinations.

Évaluation, fraude et GPT-4

L'implication de GPT-4 est énorme pour un bon nombre de pratiques évaluatives dans le réseau collégial. Si les étudiants et les étudiantes ont dans leur poche un outil qui peut pondre un texte argumentatif de 600 mots en quelques secondes, ça change considérablement les règles du jeu. Nous entendons beaucoup parler de solutions, notamment de l'achat d'outils de détection, de formation sur l'utilisation des IAG en salle de classe ou de l'ajout de ressources pour accompagner les professeurs. Je dois avouer que la majorité de ces solutions ne me convainquent pas. Je n'aime pas que la plupart des solutions nous ramènent vers le technologique au lieu de se concentrer sur notre force : le pédagogisme.

Certes, l'arrivée de GPT-4 bouscule l'ordre établi, mais je crois que c'est possible d'utiliser positivement ce bouleversement un peu comme les professeurs d'une autre époque l'ont fait avec d'autres avancées technologiques. En fait, je pense que GPT-4 pourrait aider une (véritable) transition vers une évaluation par compétences. Au lieu de jouer au chat et à la souris avec GPT-4 pour tenter de détecter les fraudes, je crois que notre salut se trouve dans des évaluations qui ne peuvent pas être réalisées par une IAG telle que GPT-4. Toutefois, je suis conscient que cette idée est d'une difficulté variable selon la discipline enseignée.

Je vois deux axes majeurs qui permettent d'atteindre des évaluations à l'épreuve des IAG. D'abord, en amont, à l'étape de la conception de nos évaluations, puis, en continu, par une notation orientée sur le processus.

La conception des évaluations

Le Rapport sur l'état et les besoins de l'éducation : Évaluer pour que ça compte vraiment (Conseil supérieur de l'éducation (CSE), 2018) affirme que :

[d]ans la logique de l'approche par compétences, c'est l'évaluation faite en fin de parcours ou d'étape qui devrait permettre de certifier jusqu'à quel point les apprentissages ont été effectués. Il est cependant prudent de fonder le jugement évaluatif sur plusieurs traces d'apprentissage... (p. 10)

Le rapport du CSE met en garde contre le cumul de plusieurs évaluations tout au long de la session, car « en cumulant les étapes au bulletin [...], on garde aussi la trace des premières erreurs dans le résultat final (l'addition de ces étapes), ce qui fait en sorte que cette note ne témoigne pas toujours adéquatement des acquis réels de l'élève ou de l'étudiant » (2018, p. 10). De fait, « ceux qui connaissent des débuts difficiles se retrouvent particulièrement pénalisés par le cumul de point » (2018, p. 10).

L'approche par compétences prône une évaluation plus globale – dite *certificative* –, en fin de parcours, pour attester de l'atteinte de la compétence par les étudiants et étudiantes, et n'est pas en phase avec une évaluation trop morcelée. L'idée derrière cette vision de l'évaluation, c'est que nous ne pouvons pas vraiment évaluer une compétence en la décomposant en de multiples évaluations de sous-éléments. Ce n'est pas parce que je sais courir, faire une passe entre deux cônes et botter un ballon dans un but que je sais jouer au soccer. Le soccer implique des coéquipiers, des adversaires et, surtout, une tonne de microdécisions qui doivent

être prises rapidement. Alors, si je veux évaluer la compétence de jouer au soccer, je ne peux pas simplement faire des évaluations hors contexte du sport lui-même en ciblant les sous-éléments du soccer.

étudiants de prendre des raccourcis dans leurs engagements scolaires » (2018, p. 11). Un passage vers une évaluation plus certificative et moins morcelée n'est donc pas seulement un moyen de défense contre les IAG, mais en plus une voie vers un engagement et un apprentissage accrus de la communauté étudiante.

J'ajoute aussi en vrac quelques autres moyens de défense qui touchent le contenu des évaluations. D'abord, ça risque de changer éventuellement, mais pour l'instant, l'amoncellement de données de GPT-4 se limite aux années précédant l'an 2021. Par conséquent, sur n'importe quel sujet actuel, il est incapable de produire une réponse adéquate. Le choix des thèmes abordés dans nos évaluations peut alors nous prémunir des risques de fraude.

Par la suite, les résultats que GPT-4 fournit laissent croire que tout ce qui a peu de présence sur Internet est inconnu pour lui. Ainsi, si nous travaillons avec des mises en situation « locales » ou en pédagogie authentique avec de vrais clients à proximité de nos cégeps comme de petits organismes à but non lucratif (OBNL) ou des entités internes à nos cégeps, GPT-4 ne saura probablement pas quoi dire à leur sujet. En contrepartie, plus nous utilisons des études de cas globales et internationales, plus nous facilitons la triche par les IAG.

La notation orientée sur le processus

La dualité *processus* et *produit* est un sujet très étudié, notamment là où la performance est facilement mesurable et comparable, comme dans le management et le sport. Il est généralement

En fait, je pense que GPT-4 pourrait aider une (véritable) transition vers une évaluation par compétences.

Si nous avons une pratique évaluative morcelée, nous sommes habituellement forcés de faire de petites évaluations ciblées sur des sous-éléments des compétences ou bien des évaluations qui ne touchent que les savoirs en omettant les savoir-faire et savoir-être. En compartimentant ainsi, non seulement nous nous éloignons de l'évaluation prônée par l'approche par compétences, mais nous facilitons la fraude par les IAG. Ces dernières sont beaucoup plus performantes avec les problèmes précis et circonscrits qui se concentrent sur les savoirs. Plus les évaluations sont englobantes de l'entièreté de la compétence, moins nous sommes vulnérables à l'utilisation des IAG.

D'ailleurs, même avant l'avènement de ChatGPT, le rapport du CSE recommandait de « modifier les pratiques superficielles d'évaluation des apprentissages qui permettent aux

reconnu que d'orienter les objectifs sur le processus plutôt que le produit a un effet positif sur la motivation et ultimement la performance (Williamson et collab., 2022).

Ça tombe bien, car selon moi, le meilleur moyen de défense contre la fraude par IAG est une pratique évaluative orientée sur le processus plutôt que sur le produit. Ça peut sembler être en contradiction directe avec l'argument précédent de ne pas morceler l'évaluation et d'en avoir une plus certificative si nous percevons cette pratique comme le produit et surtout comme le seul moment où nous pouvons évaluer. Toutefois, une pratique évaluative ne se limite pas à des évaluations sommatives. D'ailleurs, le rapport du CSE recommande « de fonder le jugement évaluatif sur plusieurs traces d'apprentissage » (2018, p. 10).

Comme le dit si bien Scallon (2014), la bonne voie se trouve probablement quelque part entre 15 examens hebdomadaires de 6,66 % et un examen final de 100 % :

[P]our apprécier une compétence, on ne peut pas se contenter d'additionner diverses informations collectées en cours de route, ni s'appuyer sur une situation unique, si complexe soit-elle, placée au terme de la progression. La démarche de certification doit se situer entre ces deux extrêmes. (p. 205)

Alors, à quoi ressemble concrètement une pratique évaluative orientée sur le processus qui n'est pas une addition de petites évaluations ciblées ? Et que sont ces traces d'apprentissage ? Ça ressemble simplement à une suite d'évaluations formatives qui amènent éventuellement à une évaluation certificative. Les

évaluations formatives permettent non seulement d'accompagner les apprenants et les apprenantes, mais aussi d'être dans une posture d'évaluation grâce aux traces d'apprentissage que nous obtenons.

Par exemple, pour un travail, on pourrait avoir une remise en trois temps où les deux premières remises sont formatives et la troisième est certificative. Les deux premières remises seraient des versions de moins en moins approximatives du travail final. Ainsi, comme pédagogue, nous avons deux traces d'apprentissage et une remise finale du travail. Ça nous permet d'utiliser une approche de correction qualitative et d'aller dans le sens du rapport du CSE qui dit que :

... au moment d'attribuer une note finale, l'ensemble des informations que l'on a sur les capacités développées par un élève ou un étudiant (dont les évaluations formatives) devrait être pris en compte dans l'exercice du jugement professionnel, notamment pour constater les progrès réalisés et s'assurer que la performance attendue se maintient dans le temps. (2018, p. 10)

Il est primordial ici de mentionner que de faire une remise en trois étapes ne doit pas tripler le temps consacré à la correction ; il faut plutôt adopter une notation orientée sur le processus. Les deux remises formatives permettent l'accompagnement des étudiants et étudiantes, mais donnent l'occasion aussi de démarrer la correction avant l'évaluation certificative. Ainsi, tout au long du processus, il faut « constater les progrès réalisés et s'assurer que la performance attendue se maintient dans le temps » (CSE, 2018, p. 10) en

notant les traces d'apprentissage et en commençant à remplir les outils de correction (grille critériée ou autre).

Grossièrement, au lieu de faire un blitz de correction tous les mois, il faut viser à en faire un peu chaque semaine. Avec une évaluation orientée sur le processus, il ne devrait pas y avoir de surprises à l'évaluation certificative étant donné que l'acte de correction est entamé avant celle-ci. Ainsi, la correction de l'évaluation certificative devient une validation de notre jugement des dernières semaines plutôt qu'une nouvelle correction complète et exhaustive. L'évaluation ne se base donc plus uniquement sur le moment précis de l'évaluation certificative – le produit –, mais elle fait aussi de la place à tout ce qui est arrivé avant, soit le processus.



Plus les évaluations sont englobantes de l'entièreté de la compétence, moins nous sommes vulnérables à l'utilisation des IAG.



Au-delà de cet exemple, nous pouvons également utiliser plusieurs autres stratégies pour obtenir les traces d'apprentissage, comme faire tenir aux étudiants et étudiantes des journaux de bord, utiliser des rapports d'avancement, tenir de courtes rencontres hebdomadaires pour les travaux d'équipe.

Une IAG peut donner de beaux produits finis, mais elle peut difficilement produire des traces d'apprentissage comme les deux remises formatives incomplètes citées en exemples. Elle ne peut pas non plus tenir un journal de bord ni prendre la parole lors d'une rencontre d'équipe.

Le pédagogique comme moyen de défense

Globalement, je dirais que pour contrer la fraude par les IAG, il faut « humaniser » nos évaluations. Ce que j'entends par *humaniser les évaluations*, c'est de les concevoir comme un acte d'apprentissage et non pas comme un acte de notation ou de classement. Le premier argument pour cette transition est implacable : GPT-4 est déjà dans le percentile 90 pour les évaluations de style classement (tests

standardisés, examens d'ordre, etc.), alors nous avons déjà perdu la bataille de ce côté.

Encore plus important, selon moi, cette transition est bénéfique pour la communauté étudiante. En utilisant des évaluations plus englobantes, des remises formatives, des traces d'apprentissage, des rétroactions sur ces traces et une orientation globale sur le processus d'apprentissage, nous nous plaçons en phase avec l'approche par compétences et nous contribuons à ce que les étudiants et étudiantes maîtrisent mieux lesdites compétences. Les phases d'apprentissage et d'évaluation deviennent plus harmonisées et moins distinctes. Pour reprendre l'analogie du soccer, il faut enseigner et évaluer le sport lui-même en le jouant le plus souvent possible ; il ne faut pas le déconstruire en trois blocs de cinq semaines où l'on enseigne et évalue d'abord à courir, puis à faire une passe et finalement à tirer au but. Il faut qu'à un moment donné, si faire se peut le plus régulièrement possible, les étudiants et étudiantes soient dans l'exécution complète de la compétence afin qu'ils puissent l'apprendre et que nous puissions les évaluer.

La prophétique révolution de l'éducation

S'il est certain que les IAG auront un impact indéniable sur notre société et si je dois admettre que leur incidence sur les pratiques éducatives est soudaine et sévère dans certains cas, notamment pour les tests standardisés, je me place toutefois en opposition avec ceux et celles qui voient dans ces IAG une révolution de l'éducation et du rôle du professeur.

Jusqu'à maintenant, aucune avancée technologique ne peut revendiquer le titre de « celle qui a révolutionné l'éducation », et ce n'est pas par manque de prophètes qui ont maintes fois depuis des siècles fait cette prédiction. La réalité, c'est que l'éducation n'a pas beaucoup changé dans les 2000 dernières années. Ça demeure un groupe d'étudiants et d'étudiantes qui apprennent avec un professeur dans un endroit physique approprié et admis, et c'est probablement parce que c'est encore la meilleure façon de la prodiguer. Les IAG rejoindront probablement le livre manuscrit, l'imprimerie, la radio, la télévision, le photocopieur, la calculatrice, l'ordinateur, l'Internet, les cours massifs en ligne et tous les autres dans la boîte des « ex-prétendants » à la révolution de l'éducation. Et ce, probablement plus tôt que tard.

Je me permets même de faire une prédiction plus risquée et d'affirmer que, contrairement à ce que certains peuvent croire, l'intégration comme telle des IAG en classe ne sera que très minimale et que ces dernières deviendront plutôt un simple outil de plus dans l'éventail de ce que les étudiants et étudiantes utilisent au côté de la calculatrice, de l'ordinateur et d'Internet. —