

# Écrire à l'université à l'heure des IA génératives : trouble dans l'auctorialité (1/2)

Par **Bilel Benbouzid**

SOCIOLOGUE

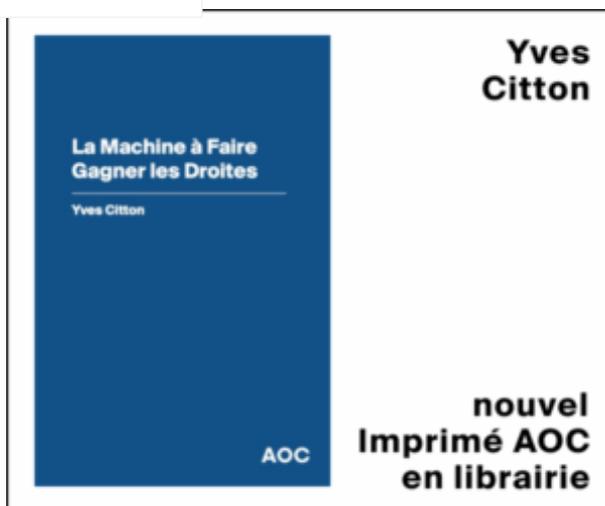
L'introduction des intelligences artificielles génératives (IAG) dans l'enseignement supérieur et la recherche pourrait constituer un point de bascule. Si l'on ne peut pas encore en anticiper pleinement les effets à long terme, des signes de transformation sont déjà repérables. Les IAG ébranlent les fondements de l'auctorialité et de la déontologie universitaire.

- favoris
- agrandir
- partager

- bluesky
- facebook
- linkedin
- copier le lien
- mail

En ma  
artifici  
du pro  
*learnii*  
passan

se Sakana annonce que son système d'intelligence  
e *AI Scientist V2*, est parvenu à automatiser l'ensemble  
article scientifique dans le domaine du *machine*  
hypothèses à la conduite des expérimentations, en  
sations et la rédaction du manuscrit.



publicité

L'article, généré sans intervention humaine directe, a été soumis à un workshop de l'ICLR, une conférence internationale en intelligence artificielle, où il a passé avec succès l'évaluation en double aveugle par les pairs, bien que les workshops soient généralement moins exigeants que les panels de la conférence principale. Conformément à un accord préalable avec les organisateurs, le papier a été retiré avant publication. L'objectif n'était pas de créer un précédent, mais de susciter un débat sur les nouvelles normes de la recherche automatisée. Cette expérience, qui n'a impliqué qu'une forme de curation humaine en amont (l'article soumis ayant été sélectionné parmi plusieurs versions générées), montre que la communauté scientifique a désormais besoin de [définir des règles](#) en matière d'attribution et de labélisation de ce type de contenus.

Bien qu'encore marginale et loin de pouvoir s'étendre à toutes les disciplines, cette expérience illustre un questionnement qui traverse l'ensemble du champ académique : que reste-t-il de l'auteur lorsque le texte n'est plus que partiellement attribuable à un sujet humain identifiable ? C'est en raison de ce trouble dans l'auctorialité que, dans l'enseignement supérieur et la recherche, les problèmes posés par les IAG se cristallisent essentiellement autour de l'écriture. Dès l'arrivée de ChatGPT à l'université, une même interrogation s'est imposée tant aux étudiants qu'aux enseignants-chercheurs : faut-il s'interdire d'y recourir pour rédiger ses travaux ?

Depuis deux ans, de nombreux acteurs du champ académique ont engagé une réflexion autour de ce dilemme moral[1]. Le débat s'articule principalement autour de trois repères normatifs associés à la pratique de l'écriture : la valeur d'intégrité académique, l'idéal éthique d'authenticité et le fondement même de la moralité, l'autonomie. Or, ces repères ne vont plus de soi. Face à ces nouvelles technologies, ils semblent même avoir perdu une part de leur force prescriptive. À tel point que, comme le titrait récemment [un article du journal \*Le Monde\*](#) au sujet des IAG dans l'enseignement supérieur : « La question n'est plus de savoir comment s'opposer, mais comment vivre avec ».

Le fait qu'un outil puisse ainsi ébranler les fondements mêmes de la déontologie universitaire indique que les enjeux politiques qu'il soulève excèdent de loin la seule opposition entre bon et mauvais usage. Cette opposition, qui considère que la technologie n'est ni bonne, ni mauvaise, mais dépend de ses usages, se fonde souvent sur une vision instrumentale de l'éthique des technologies qui cadrent les politiques actuelles de régulation de l'IA. Mais on peut aussi estimer, comme l'a formulé efficacement l'historien des techniques Melvin Kranzberg, qu'« une technologie n'est ni bonne, ni mauvaise, ni neutre » – elle transporte avec elle une certaine vision du monde, une manière d'en redistribuer les normes. Saisir la portée politique des transformations induites par l'introduction des IAG dans le monde académique consiste alors à examiner

successivement comment ces technologies troublent les trois repères normatifs mentionnés plus haut et reconfigurent les cadres de l'agir académique.

## **L'intégrité académique à l'épreuve**

Dès le départ, l'usage des IAG a été perçu comme une transgression morale dans la production des écrits. Ce cadrage initial du problème a placé la question de l'intégrité académique au premier plan des préoccupations. L'intégrité est souvent présentée comme une valeur fondamentale qui engage enseignants, chercheurs et étudiants à adopter des pratiques honnêtes, rigoureuses et responsables, aussi bien dans la production que dans la transmission des savoirs. Elle implique la reconnaissance du travail d'autrui, la transparence des évaluations et la loyauté intellectuelle. Chacune de ces dimensions de l'intégrité se trouve aujourd'hui mise à l'épreuve par l'usage des outils génératifs.

Un exemple éclairant nous vient de l'édition scientifique. Les revues et éditeurs peinent à encadrer de manière cohérente l'usage des IAG par les auteurs. Si la plupart s'accordent à interdire que ces outils soient crédités comme co-auteurs, [les lignes directrices restent profondément hétérogènes](#) sur les modalités de divulgation de l'usage d'une IAG dans le processus d'écriture. Ce flou normatif, qui s'étend aussi aux travaux académiques des étudiants, contribue à l'ambiguïté entourant les notions de plagiat ou de triche, souvent mobilisées pour [qualifier les problèmes](#) posés par les IAG. L'usage des IAG constitue-t-il une forme de plagiat, dans la mesure où les textes produits ne sont pas cités comme provenant d'une IAG ? Ou bien s'agit-il d'une forme de triche, dans la mesure où ces outils remplacent une partie du travail intellectuel attendu ?

L'usage d'une IAG ne constitue pas exactement un cas de plagiat (l'utilisation d'idées ou de textes d'autrui sans attribution correcte) car il ne s'agit pas d'un auteur humain dont on aurait volé les idées. Le problème de la citation avec les IAG tient à un paradoxe : d'une part, leurs énoncés ne disposent pas d'une valeur référentielle claire — on ne peut pas les citer pour appuyer un argument ou les commenter ; et, d'autre part, il est attendu que les lecteurs soient informés lorsqu'un texte a été généré par une IA. En ce qui concerne la triche, les universités la définissent généralement comme tout acte procurant un avantage injuste à un étudiant. Si l'on applique cette définition à la lettre, alors l'usage d'une IAG dans un contexte où elle est interdite — comme lors d'un examen sur table — constitue sans ambiguïté une tricherie. Mais en dehors de ces situations balisées, la ligne devient plus floue. Est-ce encore tricher que de s'appuyer sur ChatGPT pour un

devoir à la maison, comme on le ferait en demandant de l'aide à un parent ou à un ami ? Le caractère « injuste » de l'avantage devient alors difficile à établir. Cette ambiguïté touche également les chercheurs[2] : au-delà des cas flagrants de fraude (génération fictive de données/résultats), se pose la question de l'équité dans la compétition scientifique et de ce qui constitue une concurrence loyale entre chercheurs, notamment dans l'évaluation de leur qualité – des modalités d'évaluations scientifiques tendant elles-mêmes à être automatisées par les IAG.

Finalement, ni tout à fait triche, ni véritable plagiat, l'usage des IAG soulève une question plus fondamentale : quelle est la nature de l'effort intellectuel légitime à fournir dans un contexte académique ? Dans une situation d'apprentissage ou dans celle de la production de connaissance scientifique, définir un seuil d'effort légitime reviendrait à réguler sur le seul mode répressif. Cette approche strictement répressive supposerait de surveiller chaque usage et d'être en mesure de **détecter automatiquement** la présence de textes générés par IA. Or, il est difficile de distinguer un usage « passif » de l'IAG d'un usage « actif ». Par exemple, dans une expérimentation récente, Ludovic Jeanne a simulé **la production d'un article scientifique** à partir d'un texte existant, qu'il a soumis à un processus de reformulation automatique, sans ajout substantiel de contenu intellectuel. Le texte généré, après des retouches minimales et une relecture légère, échappe aux détecteurs de plagiat car il respecte formellement les critères d'originalité. L'auteur de l'expérimentation conclut que « le texte peut être doté de sens, original (dans le sens juridique du terme : sans antécédence), tout en étant frauduleux car non-authentiquement de l'auteur. » Ainsi, l'application du principe d'intégrité académique au cadrage des usages des IAG se heurte à une impossibilité pratique de contrôle, tant ces outils redéfinissent la frontière entre l'originalité formelle d'un contenu et l'expression authentique de son auteur.

## **Être un cyborg fidèle à soi-même**

Dès lors, le problème se déplace. Il ne tient plus seulement à la capacité de détecter l'usage de l'IA ou à mesurer un effort d'écriture, mais à une interrogation sur ce qui fait qu'un texte nous engage réellement. Derrière cette question se joue une réflexion éthique sur le rapport à soi dans l'acte d'écrire, qui renvoie à une notion centrale de la philosophie morale contemporaine : l'authenticité, entendue comme cet idéal éthique qui renvoie à la capacité d'un individu à vivre en accord avec lui-même, en restant fidèle à ce qu'il reconnaît comme constitutif de son identité.

Cette quête de l'authenticité de l'individu moderne est bien plus qu'un programme de développement personnel. Charles Taylor l'envisage comme une quête du sens de la vie collective, des valeurs et des convictions qui structurent notre existence sociale. *Pour Taylor, l'authenticité ne consiste pas à se replier sur soi, mais à se constituer en tant que sujet dans un rapport dialogique avec autrui[3]. L'identité se façonne au fil des relations, des appartenances et des normes partagées. Elle n'est jamais donnée d'avance, mais se construit comme un récit que l'individu découvre en l'écrivant avec les langages, les formes d'expressions, les références et les horizons de sens de sa communauté morale.* Quelle place accorder à l'authenticité dans un monde où nos modes d'expression sont désormais médiatisés par des technologies qui bouleversent notre rapport à l'écriture ?

Cette interrogation trouve un écho particulier dans le contexte universitaire, où les étudiants se trouvent dans une phase décisive de construction de soi et de leurs manières de penser à travers l'acquisition d'une multitude de compétences[4]. De sorte que l'enseignement supérieur est un espace de formation « politique », dans le sens où l'on apprend, au contact des enseignants, des chercheurs, des autres étudiants et de tout un ensemble de ressources matérielles et immatérielles, à exprimer une voix singulière dans un langage commun, en assumant le poids moral de ses influences. Dans la diversité des médiations qui participent à la construction de l'authenticité à l'université, quel statut accorder aux IAG ?

Le problème ne réside plus, comme pour l'intégrité, dans le fait que l'étudiant pourrait copier ou paraphraser des contenus sans appropriation critique. Il tient plutôt à la manière dont l'outil, même mobilisé comme vecteur de stimulation intellectuelle – pour clarifier une idée, explorer un argument ou fluidifier un texte –, exerce une influence diffuse, souvent imperceptible. *Les modèles génératifs fonctionnent comme des co-auteurs implicites : ils orientent les formes de pensée tout en laissant à l'utilisateur l'impression d'une autonomie pleine et entière. Même lorsque l'usage paraît réflexif, le risque d'une dépossession subtile demeure. Face aux influences tacites des IAG, il est difficile de discerner les lignes de partage entre l'expression authentique de soi et les effets normatifs induits par la machine.*

La question de l'authenticité rejoint ici le problème de la persuasion, un problème au cœur de nombreuses recherches récentes sur les effets des IAG. Toute une série d'expérimentations a montré que les IAG peuvent influencer les opinions parfois de manière plus efficace que des interlocuteurs humains[5]. Elles génèrent des messages personnalisés, en adaptant leurs réponses à la logique argumentative des utilisateurs et en mobilisant des formes subtiles de rhétorique adaptative. Ce pouvoir persuasif n'est pas nécessairement problématique : il peut être mis au service d'objectifs vertueux dans des

exercices de déconstruction ou de remise en question des croyances et des opinions, en s'adaptant finement aux raisonnements des individus et en fournissant des contre-arguments ajustés, comme, par exemple, face aux croyances complotistes. Mais il peut aussi contribuer à l'homogénéisation des formes de pensée, en imposant, de façon implicite, une vision dominante du monde, au détriment de la diversité épistémique et culturelle. Ainsi, la quête de soi, qui passe par le dialogue avec les « autres », inclut désormais les IAG parmi ces interlocuteurs. Mais ces dernières occupent une place singulière dans cette formation dialogique, au point qu'elles invitent à redéfinir ce que signifie être authentique.

Si humains et IAG co-écrivent sans jamais pouvoir totalement démêler leurs voix, l'authenticité reposerait sur la cohérence entre le processus d'écriture des usagers et leur capacité à comprendre et façonner les mécanismes d'écriture. Grammarly, outil bien connu d'assistance à l'écriture, propose désormais une fonctionnalité intitulée **Authorship**, qui retrace l'origine des fragments textuels produits par l'humain ou la machine, sans toutefois lever entièrement les ambiguïtés autour de l'écriture authentique. Certaines initiatives comme **Claude for Education** proposée par Anthropic esquissent une voie : en concevant des assistants pédagogiques qui privilégient l'explicitation des raisonnements plutôt que la simple fourniture de réponses, elles engagent une écriture plus réflexive. Mais cette orientation ne garantit pas encore un accès transparent aux mécanismes internes des modèles (pré-prompts, biais, filtres, explicabilité), ce qui entrave la poursuite de l'idéal d'authenticité. Ces initiatives favorisent néanmoins une prise de conscience des médiations en jeu. Une nouvelle forme d'écriture collaborative est en train de s'imposer. Elle pourrait faire émerger une subjectivité hybride comparable à la figure du cyborg décrite par Donna Haraway : un cyborg qui trouverait son authenticité dans l'exploration critique et ironique de ses propres frontières entre humain et machine d'écriture[6].

## **L'auteur est mort, vive l'auteur !**

Bien que les enseignants-chercheurs et les étudiants soient encore loin de s'être pleinement approprié ces machines d'écriture, au point de devenir des cyborgs authentiques, cette nouvelle hybridation, en brouillant les frontières entre humain et machine, entre intention et automatisation, réactualise le débat poststructuraliste sur la figure de l'auteur.

Dans *De la grammatologie*, Jacques Derrida avait déjà critiqué le logocentrisme platonicien qui associait la vérité d'un énoncé à la présence d'un auteur. En déconstruisant la hiérarchie entre parole et écriture (l'écriture étant, chez Platon, suspectée de donner l'illusion du savoir, comme le suggère le mythe de Theuth qu'il rapporte dans le *Phèdre*, souvent réactivé aujourd'hui dans [les critiques des LLM](#)), Jacques Derrida s'en prend à l'idée d'un sens stable garanti par la présence d'un sujet conscient. À travers la notion de *différance*, il montre que le sens ne se donne jamais immédiatement, mais se construit dans un jeu de traces et de décalages, rendant illusoire toute maîtrise souveraine du langage. Sa formule célèbre « il n'y a pas de hors-texte[7] » implique que toute signification est différée, sans ancrage dans une conscience fondatrice.

Dans leur article « ChatGPT: Deconstructing the Debate and Moving it Forward », Mark Coeckelbergh et David Gunkel prolongent le geste déconstructiviste initié par Derrida[8]. L'arrivée des IAG réactive une thèse déjà défendue par ce dernier : il n'y a jamais eu d'origine pure du texte, pas plus que d'auteur souverain maîtrisant entièrement le sens de ses écrits. L'écriture a toujours constitué un processus distribué, dans lequel se mêlent étroitement humains, langue et techniques. Ce trouble dans la figure de l'auteur réactive, dans le débat contemporain, les analyses de Barthes et Foucault qui, dès les années 1960, avaient remis en question l'autorité de l'auteur. Barthes annonçait la « mort de l'auteur » pour mieux restituer un pouvoir d'interprétation au lecteur[9] ; Foucault, dans *Qu'est-ce qu'un auteur ?*, soutient que la « fonction-auteur » n'a pas toujours existé telle qu'on la conçoit aujourd'hui, mais constitue un dispositif historiquement situé de régulation des discours[10].

C'est en suivant Derrida, Barthes et Foucault que Mark Coeckelbergh et David Gunkel proposent de reformuler le problème éthique des IAG. Plutôt que de demander : « Ce texte est-il l'expression authentique d'un sujet intègre ? » (ce qui suppose précisément le modèle d'un auteur individuel) il devient pertinent d'adopter un autre angle, en posant les questions : « Que fait ce texte ? Qui en est affecté ? Quel type de relation de pouvoir ou de dépendance produit-il ? » C'est là un déplacement majeur, d'une morale centrée sur l'identité du locuteur (qui parle ?) vers une morale attentive aux effets concrets des productions discursives (que fait-on avec ce texte ?). Ce trouble dans l'auctorialité induite par les IAG ne serait donc pas une perte, un abandon de son humanité, mais un geste émancipateur. La production intellectuelle devient le lieu de médiations multiples qu'il convient désormais d'apprendre à situer.

Cette relance de la déconstruction pose néanmoins de nombreux problèmes pratiques. Une éthique déconstructiviste éclaire les conditions de production des textes mais ne permet toujours pas de cadrer, en pratique, les usages des IAG. Antoine Compagnon

rappelle que, même affaiblie, cette figure reste une nécessité politique, éthique et herméneutique : elle organise nos pratiques de lecture, d'enseignement, d'édition, et rend possible l'attribution des responsabilités. Ce que les IAG troublent, ce n'est pas seulement la paternité des textes, mais tout un cadre de régulation de l'économie des discours académique[11]. Même Barthes, dans *Le Plaisir du texte*, reconnaît avoir besoin d'une figure d'auteur : non comme origine, mais comme partenaire imaginaire de la lecture. Autrement dit, on ne se débarrasse pas si facilement de l'auteur nous dit Antoine Compagnon. Sans cette figure, la lecture devient indéchiffrable, et la production intellectuelle, politiquement ingérable.

Ainsi Compagnon insiste sur la nécessité pratique de maintenir la figure de l'auteur, malgré sa déconstruction théorique. Comme l'a montré David Pontille, l'auctorialité est un concept en mutation permanente, confronté aux défis de la collaboration scientifique et aux exigences de transparence et d'équité dans la reconnaissance des contributions. Si les enjeux autour de l'auctorialité varient selon les disciplines, la persistance de la figure de l'auteur renvoie toujours à la même dimension fondamentale de la production intellectuelle : celle de la signature comme acte d'engagement. Et c'est par l'attribution nominale que les énoncés produits acquièrent un statut de validité. Pour que ce processus opère pleinement, les textes doivent être impérativement liés au nom de leurs auteurs : « Tant l'authenticité que l'autorité qui leur sont conférées en résultent [12] ».

## **L'éthique du non-usage des IAG : un devoir**

### **envers soi-même**

Cette dimension de la signature comme engagement éclaire la dimension éthique et politique de l'auctorialité. Être auteur d'un texte, c'est accepter d'en répondre, de le défendre, de l'assumer comme l'expression d'une pensée. Il s'agit d'une responsabilité au fondement même de notre constitution comme sujets autonomes. C'est pourquoi cette fonction d'auteur soutient l'un des piliers éthiques les plus valorisés dans l'enseignement supérieur et la recherche : l'autonomie, définie comme la capacité à se donner à soi-même ses propres fins, par un usage réfléchi de la raison. Si la figure de l'auteur disparaît, que devient alors la capacité de chacun à se penser comme sujet moral, capable de juger et d'agir selon ses propres raisons ? Autrement dit, que reste-t-il

de cette autonomie dans un monde où les productions intellectuelles sont de plus en plus coproduites par des dispositifs algorithmiques ?

Dans leur article « Should I Use ChatGPT to Write My Papers ? », Aylsworth et Castro envisagent l'autonomie non pas seulement comme un idéal éducatif, mais comme une obligation morale envers soi-même : celle de cultiver les conditions de possibilité de sa liberté, en développant les capacités de réflexion, de jugement et d'auto-

détermination[13]. L'acte d'écriture n'est pas un simple exercice technique ou une compétence instrumentale. Il devient un acte de formation éthique, un moment où l'individu exerce et éprouve sa propre capacité à réfléchir sur le sens de ses engagements, à formuler des arguments, à construire un point de vue justifiable.

Ce qui rend leur position particulièrement intéressante, c'est qu'ils identifient une tension entre deux conceptions de l'autonomie. D'un côté, l'autonomie comme fin éducative suppose une lente élaboration de soi à travers des pratiques intellectuelles rigoureuses ; elle se cultive dans l'effort de penser, de structurer ses idées, d'argumenter – autant de gestes que l'écriture rend possibles. De l'autre, l'autonomie peut être comprise comme liberté de choix, c'est-à-dire comme la capacité à décider de ses moyens pour atteindre un but donné. Dans cette perspective, le recours à une IAG peut sembler augmenter l'autonomie de l'étudiant.

Mais pour Aylsworth et Castro, cette autonomie procédurale, centrée sur l'efficacité ou la satisfaction personnelle, entre en contradiction avec l'autonomie comme finalité morale de l'éducation. Car déléguer l'écriture à une intelligence artificielle, c'est renoncer à exercer les facultés mêmes qui rendent cette autonomie possible : la délibération, la critique, la mise en forme d'un point de vue propre. C'est là que la tension devient critique : plus l'étudiant se sent libre d'utiliser ChatGPT pour écrire à sa place, moins il cultive sa capacité à se penser comme « auteur » de ses choix et de ses jugements.

Ils apportent alors une réponse catégorique au dilemme moral que nous avons posé en introduction de cet article : les étudiants ont une obligation morale de ne pas utiliser ChatGPT pour rédiger leurs devoirs, car écrire soi-même ses textes est essentiel à la culture de son autonomie, laquelle constitue une expression du respect que l'on se doit en tant qu'être rationnel capable de fixer et poursuivre ses propres fins. C'est pourquoi, même en dehors des logiques de contrôle ou de sanction, l'université a le devoir, selon eux, d'encourager et de protéger ces moments de réflexion personnelle, en créant les conditions pédagogiques où écrire par soi-même redevient une expérience significative.

Les auteurs proposent ici une morale de la responsabilité envers soi-même, dans laquelle écrire par soi-même n'est pas seulement une tâche scolaire, mais un acte de reconnaissance de sa propre dignité morale. Ils nous rappellent que : «

*Écrire, c'est penser. Penser, c'est se construire. Et se construire, c'est honorer l'humanité en soi. »*

Selon nous, cette tension entre l'autonomie comme fin éducative et l'autonomie comme liberté de choix résume bien le dilemme auquel étudiants et enseignants sont confrontés face aux intelligences artificielles génératives. *S'il existe une éthique possible de ces outils, elle passe peut-être par la capacité et la liberté de choisir de ne pas les utiliser, justement là où leur usage est techniquement possible. Une éthique du non-usage n'a de sens que dans la possibilité de l'usage. Encore faut-il interroger les conditions réelles de ce choix : car la liberté de ne pas recourir aux IAG, loin d'être partagée également, devient elle-même un privilège social.*

**Bilel Benbouzid**

SOCIOLOGUE, MAÎTRE DE CONFÉRENCES À L'UNIVERSITÉ GUSTAVE EIFFEL ET  
CHERCHEUR AU LABORATOIRE INTERDISCIPLINAIRE SCIENCES INNOVATIONS  
SOCIÉTÉS (LISIS)