

Intégrité Académique et Intelligence Artificielle. Réflexions prospectives sur la base du cas Speedwrite

Ludovic Jeanne

Mots-clés: intégrité académique, intelligence artificielle, imitation, plagiat, prospective, typologies.

Keywords: academic integrity, artificial intelligence, imitation, plagiarism, prospective, typologies.

Résumé : L'intégrité académique a partie liée avec la lutte contre la fraude académique. L'équilibre des ressources et des stratégies joue un grand rôle : la capacité des organisations et de leur personnel dédié à déceler et à établir une fraude est l'un des termes majeurs de la lutte contre celle-ci. Or, l'intelligence artificielle semble offrir un vaste champ de sophistication et de professionnalisation des pratiques de fraude académique. Cette communication se concentre sur un aspect particulier de la remise en cause de l'équilibre actuel : l'anticipation des potentialités offertes par l'intelligence artificielle en matière de production de textes sans travail rédactionnel de l'auteur putatif. À partir du cas d'une plateforme d'intelligence artificielle dédiée à la génération semi-automatique de textes « originaux », nous analysons l'efficacité et les implications d'un tel outil ainsi que de proposer des hypothèses prospectives quant à l'évolution du champ de la fraude académique.

Abstract: Academic integrity is linked to the fight, in whatever form, against academic fraud. The balance of resources and strategies plays a critical role: the ability of organizations and their dedicated staff to detect and establish fraud is one of the major terms of the fight against it. Artificial intelligence seems to offer a vast field for the sophistication and professionalization of academic fraud practices. This paper focuses on a particular aspect of the challenge to the current balance: the anticipation of the potentialities offered by artificial intelligence in terms of text production without editorial work by the alleged author. Using the case of an artificial intelligence platform dedicated to the semi-automatic generation of "original" texts, the paper will attempt to analyze the efficiency and implications of such a tool as well as to propose prospective hypotheses regarding the evolution of the field of academic fraud.

Publié dans *Actes du 2ème Colloque IRAFPA*, 2022, 211-229

<https://doi.org/10.56240/cmb9919>

Tout droit de reproduction réservé

« Il n'est pas possible que celui qui fait en secret une des choses à propos desquelles on a mutuellement passé un contrat visant à ne pas faire de tort ni en subir, soit certain que cela restera inaperçu, même si cela passe inaperçu des milliers de fois dans le présent. Jusqu'à ce qu'il soit mort, en effet, impossible de savoir si cela restera encore inaperçu. »

Épiqueure, Maxime Capitale XXXV (trad. Pierre-Marie Morel)

Introduction¹

L'intégrité académique (IA)² a nécessairement partie liée avec la lutte, quelle qu'en soit la forme, contre la fraude académique ou scientifique. Dans cette lutte, l'équilibre des ressources et des stratégies (celles du fraudeur potentiel et celles du référent intégrité académique ou scientifique, qui sont pour partie celles de son institution) joue un grand rôle : la capacité des organisations et de leur personnel dédié à déceler et à établir une fraude est l'un des termes majeurs de lutte contre celle-ci. Ne serait-ce que parce que, dans une perspective de sociologie de la déviance (Becker, 1985), l'anticipation que fait le fraudeur potentiel de la facilité ou de la difficulté avec laquelle il peut être démasqué et socialement sanctionné est l'une des données fondamentales de définition de sa conduite. Or cet équilibre n'est pas une donnée stable dans le temps, ni même dans l'espace. Et l'émergence de nouvelles technologies nous le rappelle avec une acuité toujours renouvelée. Cette communication se concentre sur un aspect particulier de la remise en cause de l'équilibre actuel : l'anticipation des opportunités offertes par l'intelligence artificielle en matière de production d'un texte « original », mais sans travail rédactionnel de l'auteur putatif et donc sans la construction afférente de sa pensée.

À partir du cas d'une plateforme d'intelligence artificielle (IA) dédiée à la génération automatique de textes « originaux », la communication s'efforcera d'analyser les implications de ces nouveaux outils tant du point de vue du fraudeur potentiel que des organisations de l'enseignement supérieur et de la recherche qui s'efforcent de former à l'intégrité et de sanctionner la fraude. Deux questions principales structurent la réflexion prospective : quelles sont les conséquences potentielles en matière d'intégrité académique d'une telle émergence technologique ? L'intelligence artificielle appliquée à la production de textes annonce-t-elle la disparition du plagiat au profit de la dominance d'un autre type de fraude ? La communication s'efforce d'ailleurs de clarifier dans quelle mesure et dans quelles

¹ Cette recherche a bénéficié des éclairages et des compétences scientifiques de M. Alexis Lechervy (Maître de Conférence et chercheur au GREYC - UMR CNRS 6072, Université de Caen) et de M. Denis Escudier (Directeur associé de MomentTech et Vice-Président du Datalab Normandie).

² Par commodité de rédaction, l'expression « intégrité académique » visera dans ce texte aussi bien les activités pédagogiques que de recherche.

circonstances le recours à de l'intelligence artificielle, sous la forme d'un générateur semi-automatique de texte, constituerait bien une fraude académique.

Cette première exploration vise à tenter d'identifier les principales mutations que l'émergence de l'intelligence artificielle est susceptible de produire en matière de fraude académique. C'est indispensable pour ensuite penser les implications et les nouveaux horizons à explorer en matière d'intégrité académique. Toute cette exploration peut nous mener à examiner cette question abyssale, sur un plan épistémologique, scientifique et éthique : quel sera le statut demain d'un texte adressé à une revue scientifique pour lequel les « auteurs » auront « entré » dans un outil IA des sources bibliographiques et des résultats (peut-être) mais n'auront rien écrit (ou presque) tout en produisant un texte « original » ?

Produire des textes « originaux » mais « sans auteur » : une exploration des capacités de la plateforme Speedwrite

Le meilleur moyen d'examiner les potentialités actuelles et à venir d'une nouvelle technologie en matière d'intégrité académique, c'est de l'expérimenter. C'est donc à une expérimentation sur la plateforme Speedwrite que nous nous sommes livrés. Cette expérimentation a consisté principalement à sélectionner un texte et à le faire traiter par la plateforme pour produire un nouveau texte, « original ». De nombreux choix ont dû être faits pour mener cette expérimentation. Le premier choix a été de retenir un texte en anglais. En effet, la plateforme Speedwrite a d'abord été conçue, aux dires même de ses créateurs, pour opérer en anglais ou dans des langages proches. Sur la plateforme on peut en effet trouver la mention suivante³ :

“Speedwrite works best on English, and is optimized for the generation of English texts. You might also find that it works fairly well on languages close to English, for example Spanish, and French. The more different a language is from English, the worse it will work. For example Arabic or Mandarin will not work well, at all.” (speedwrite.com)

Il a donc paru opportun d'en rester à l'anglais pour une première exploration des capacités de l'outil dans des conditions optimales.

Le texte choisi porte sur un domaine de spécialité de l'auteur et est en rapport avec l'éthique, dans le but de faciliter les adaptations et les améliorations post-production. Le but était de disposer d'un texte final « original » susceptible de passer le test d'une plateforme de détection de similarités tout en étant en mesure

³ Le texte a été souligné par l'auteur.

d'intervenir le moins possible sur le texte produit. La plateforme de détection de similarités utilisée est Turnitin.

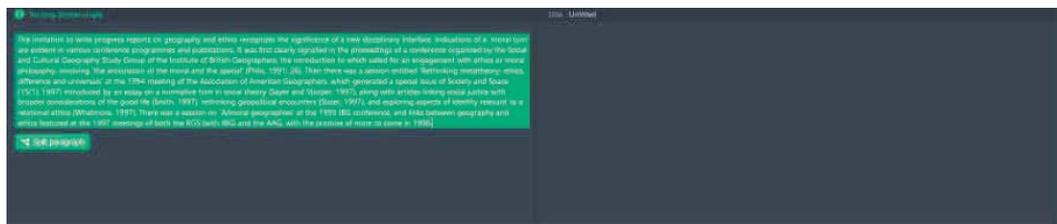
Le choix de ne retenir qu'un seul texte pour le transformer à travers la plateforme Speedwrite vise à mettre les capacités de celle-ci à l'épreuve de façon optimale. En effet l'une des « promesses » faites par la plateforme est que le texte produit ne peut pas être caractérisé comme plagiat en s'appuyant sur les outils existants de détection de similarités. Pour des raisons statistiques, la multiplication de textes en entrée ne peut que représenter un facteur facilitant pour le processus de production d'un texte original en sortie. Le défi paraît bien plus grand s'il y a un seul texte en entrée. La plateforme elle-même encourage explicitement l'utilisateur à multiplier les textes en entrée.

Nous avons également fait le choix de retenir un texte relativement court pour mettre vraiment à l'épreuve les capacités de la plateforme Speedwrite. En effet, comme indiqué ci-dessus, les mécanismes d'intelligence artificielle sont fondamentalement des mécanismes statistiques : plus la taille du corpus textuel est petite, plus le risque de mettre en difficulté les capacités de reformulation de la plateforme est grand.

Le texte retenu a d'abord été « nettoyé » en appliquant les règles de suppression suivantes :

- Suppression des citations ;
- Suppression des numérotations de chapitres ;
- Suppression des traces de l'auteur original et de l'éditeur ;
- Suppression des remerciements ;
- Suppression de la bibliographie ;
- Suppression des mentions en tête et pied de page.

Puis le texte résultant, appelé « texte test nettoyé », a été entré dans la plateforme Speedwrite, paragraphe par paragraphe. La plateforme ne pouvant produire un texte nouveau et original qu'à partir de suites de mots relativement courtes, celle-ci a presque toujours demandé la division des paragraphes entrés (Document 1). Selon les cas et la taille du paragraphe initial, celui-ci pouvait être divisé par deux, trois ou quatre. Il suffisait ensuite de cliquer pour demander une « prévision » (Document 2) et la plateforme produisait un texte « original » à partir du texte initial (Document 3), comme suit (captures d'écran en cours d'opération) :



Document 1 – Étape 1 : copier-coller le texte source (sur la gauche) et le calibrer



Document 2 – Étape 2 : calibrage fait, lancer la production de « prévisions »



Document 3 – Étape 3 : texte imité (sur la droite)

Le texte résultant appelé « imitation brute » a ensuite été relu et adapté ou précisé afin de se rapprocher d'un texte finalisé (« imitation retravaillée »). L'analyse repose donc sur quatre objets textuels comme suit :

Texte	Auteur	Mots	Caractères (enc)	Variation	Remarques
Original	David M. Smith	4501	25285		Smith, D. M. (1997) 'Geography and ethics: a moral turn?', <i>Progress in Human Geography</i> , 21(4), pp. 583–590. doi: 10.1191/030913297673492951.
Test nettoyé	David M. Smith	3051	16762	-33,71%	Suppression des citations Suppression des numérotations de chapitres Suppression des traces de l'auteur original et de l'éditeur Suppression des remerciements Suppression de la bibliographie Suppression des mentions en tête et pied de page
Imitation brute (n°1)	Donald Fake	2174	11259	-32,83%	Passage dans Speedwrite
Imitation retravaillée (n°2)	Donald Fake	2363	12405	-25,99%	Relecture, reformulations (en mode Word "Révision") avec l'aide de DeepL. Réintégration de quelques références bibliographiques.

Tableau 1 – Le dispositif méthodologique

La contraction de la taille du texte d'une étape à l'autre est liée, d'une part, à des traitements volontaires de la part de l'auteur et, d'autre part, à des traitements de la

part de la plateforme Speedwrite. Ainsi la contraction entre le texte « original » et le texte « test nettoyé » est essentiellement dû aux suppressions indiquées dans le tableau 1. Par contre les contractions de la taille du texte entre le texte « test nettoyé » et les textes « imitation » 1 et 2 sont principalement dues aux traitements appliqués par la plateforme⁴. Notamment celle-ci a fait disparaître les très nombreuses références bibliographiques présentes dans le corps du texte « original » et conservées à l'étape précédente. D'autre part, le texte choisi comportait plusieurs passages constitués d'énumérations assez longues : ce sont ces passages-là qui semblent avoir été drastiquement contractés par la plateforme. Cela met donc en évidence un effet du traitement appliqué par la plateforme pour produire le texte de sortie : celui-ci aura tendance à être contracté en comparaison de la taille du matériau textuel mis en entrée et cette contraction n'affecte pas de la même manière différentes formes syntaxiques.

La production du texte d'« imitation retravaillée » a ensuite permis de clarifier et de structurer la présentation du texte comme suit :

Geoethics: geography and the turn of ethics

By Donald Fake, Department of Geography, Clark University, USA

The invitation to ~~participate in~~investigate the progress ~~reports on~~regarding ethics and geography recognizes the importance of a new disciplinary interface, which is evidenced in various conference programs, research and publications. In the introduction to a conference on the topic of ethics, ~~the group~~the author emphasized the importance of the spatial and moral articulation as a new and critical objective for the community of geographers (social, political and cultural geographies). To understand that turning point, it is necessary to put into perspective its history.

Recent history of the encounter between Ethics and Geography

In 1994, a session on metatheory was held at the AAG meeting. It generated a special issue of Society and Space, which discussed various topics related to social theory and ethics. Among the topics discussed were the normative turn in social theory, geopolitical encounters, and the concept of relational ethics. At the 1995 IBG conference, a session on moral geographies discussed the importance of spatial and moral articulation. In 1997, the RGS and the AAG also highlighted the link between geography and ethics.

Document 4 – Relecture et compléments apportés par Donald Fake (échantillon)

⁴ Les pourcentages pour les textes « Imitation » 1 et 2 se rapportent au texte « Test nettoyé », tandis que pour ce dernier le pourcentage se rapporte au texte « Original ». Cela permet de calibrer les effets de compression de chaque phase de traitement.

mais aussi d'éliminer des scories issues du traitement de la plateforme Speedwrite :

It is also important to note that criticism of moral codes that explicitly condone human sacrifice and racism and torture can raise questions about the authority of the principles used to condemn these practices.

~~For the best predictions, start with at least five paragraphs. Then, add at least 100 words into this window.~~

In Billington's discussion (1993) of moral relativism, he uses the heading "Geographical perspective" to refer to the various ideas about how to behave. He also makes the point that one doesn't need to be an anthropologist to be aware of the various conflicting ideas about how to behave.

Document 5 – « Nettoyage de scories » dans le texte « Imitation brute » par Donal Fake (échantillon)

Cette phase a également permis de réintégrer, de façon limitée, des références bibliographiques ; ceci afin d'aboutir à un texte qui paraisse relativement fini et cohérent.

Ceci fait, le texte dit « imitation retravaillée » a été complété d'un titre et d'un nom d'auteur. Et c'est le résultat final de ce processus qui a été testé sous la plateforme de détection de similarités textuelles Turnitin. Le test est ressorti avec « seulement » 15% de similarité :



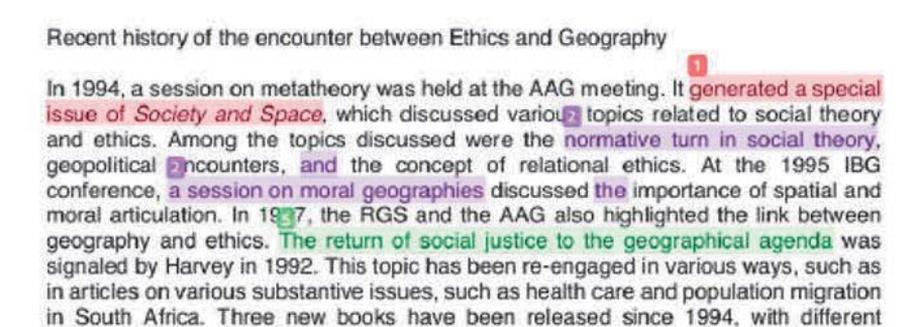
Document 6 – Résultat du test de détection de similarités textuelles sous Turnitin

Il a été ensuite procédé d'une part à l'extraction du rapport Turnitin complet et à l'examen du corpus textuel ainsi testé à partir de l'interface interactive. Cela a permis de mettre en évidence, dans ce cas exploratoire, que même si le taux de similitude résultant est extrêmement bas (et donc "non-significatif"), la plateforme Turnitin a néanmoins rapporté le texte d'imitation produit à l'aide de la plateforme Speedwrite à sa source, c'est-à-dire le texte original de David M. Smith :



Document 7 – Bilan des identifications relevées par Turnitin

En effet, à partir de l'interface interactive, il a été possible de vérifier que la source 1 rapportant 12% de similitude renvoie bien au texte original. La 2e et la 3e sources renvoient au même auteur. Cependant l'observation du rapport Turnitin, montre que nous serions dans l'impossibilité d'argumenter facilement d'un plagiat à l'encontre de Donald Fake. En effet l'essentiel du texte testé se présente ainsi :



Document 8 – Identification de similarités par Turnitin (échantillon)

À l'inverse, des paragraphes entiers sont exempts de toute détection de similarités. Il est donc possible de conclure que, malgré les contraintes imposées dans le cadre de cette expérimentation, la plateforme Speedwrite a été capable de produire un texte paraissant original et pour lequel une autorité de vérification serait dans l'incapacité d'établir des actes plagiaires de façon indiscutable. Pourtant, nous savons que Donald Fake n'a pratiquement rien écrit du texte que nous qualifions d'imitation. En effet, un rapide comptage montre qu'en tout et pour tout Donald Fake a écrit 122 mots dans le corps du texte. Cela signifie que, sa bibliographie faisant 128 mots, si l'on reprend les statistiques du Tableau 1, Donald Fake a effectivement écrit 5,45% du texte d'imitation final ("Imitation retravaillée") qu'il a

par ailleurs relu entièrement pour pouvoir écrire ces 122 mots. La « promesse commerciale » de la plateforme Speedwrite est donc globalement tenue.

Nous savons néanmoins que, à différents égards, le texte produit est un texte qui enfreint plusieurs principes de l'intégrité académique. Mais nous ne pouvons le qualifier de plagiat ; du moins pas dans ses formes simples et les plus communes. Or, tout comme copier le travail d'autrui et se l'attribuer est appelé « plagiat », produire par une intelligence artificielle un texte que l'on n'a ni pensé, ni écrit *stricto sensu*, doit porter un nom. Cela pose donc un problème de dénomination et de qualification. Comme le montre le tableau 1, le choix a été fait d'appeler ce genre de texte une « imitation ». Ce choix a été fait dans les sens de :

- ❖ « Substitut insatisfaisant de quelque chose » (www.cnrtl.fr) ;
- ❖ « Produit ou personne de substitution et de moindre valeur. » (www.cnrtl.fr) ;

Et mieux encore, de :

- ❖ « Reproduction artificielle du fonctionnement d'un appareil, d'une machine, d'un système, d'un phénomène, à l'aide d'une maquette ou d'un programme informatique, à des fins d'étude, de démonstration ou d'explication » (www.cnrtl.fr) ;

Ou encore :

- ❖ « Essayer de tromper autrui en imitant l'apparence d'une chose réelle. » (www.cnrtl.fr) ;

Les idées qui ont été ainsi retenues sont celles « d'ersatz » et de « simulation ». En effet la plateforme Speedwrite réussit assez bien à simuler le processus d'écriture d'un texte par une intelligence humaine et à produire un texte final qui est donc un ersatz de production intellectuelle humaine et originale. Nous proposons donc d'appeler ce type de fraude académique une « imitation ».

Les préjudices potentiels de la fraude par imitation à l'aide d'une IA

Les préjudices potentiels de ce nouveau type de fraude sont, *a priori*, importants et présentent un niveau de gravité méritant investigation. Une manière d'aborder cette question est d'examiner les points communs avec le plagiat. Concrètement, si l'on se réfère aux « dix conséquences du plagiat » (Bergadaà, 2015b), il n'y a presque aucune différence entre le plagiat et l'imitation en termes de conséquences. En effet si l'on reprend ces 10 conséquences et que l'on caractérise les deux types de fraude, voici le résultat :

Conséquences du Plagiat (Bergadaà, 2015b)	Plagiat	Imitation IA
1 Vol d'une création originale et occultation des auteurs/sources d'origine	OUI	OUI
2 Rupture de la chaîne de la connaissance	OUI	OUI
3 Atteinte au droit ultérieur de l'auteur original à publier sans citer son plagieur	OUI	NON
4 Dissolution du sens de l'œuvre frauduleuse	OUI	OUI
5 Fraude vis-à-vis de la société et accès à des bénéfices sociaux indus	OUI	OUI
6 Incitation à la recherche bâclée	OUI	OUI
7 Dysfonctionnement et perturbation de l'édition et des revues scientifiques	OUI	OUI
8 Inhibition des chercheurs compétents et du "droit" à l'erreur	OUI	OUI
9 Atteintes à l'image des EESR et du monde scientifique dans la société	OUI	OUI
10 Coût économique croissant de la lutte contre la fraude	OUI	OUI

Tableau 2 – L'imitation comparée au plagiat à l'aune des 10 Conséquences du Plagiat (Bergadaà, 2015b)

Cette grille d'analyse permet d'avancer trois assertions :

1. l'imitation est bien une fraude car elle provoque bien neuf des dix conséquences du plagiat, et toutes ces conséquences sont des atteintes à l'intégrité académique ;
2. le seul point sur lequel véritablement l'imitation se distingue du plagiat, c'est le fait qu'elle n'atteint pas, apparemment, aux droits de l'auteur original ;
3. Un texte peut être « original » et frauduleux.

En fait, la deuxième assertion pose problème, en particulier dans le champ de la recherche. Il est vrai que dans l'absolu l'auteur « volé » ne se retrouve pas obligé de citer son voleur et en même temps son propre texte. Mais il est possible d'argumenter que l'auteur volé - potentiellement sans le savoir et sans en être conscient - va se retrouver, sinon obligé du moins amené, à citer son voleur. En effet, dès lors que l'on travaille dans un champ scientifique donné, il est de bonne pratique non seulement de connaître toute la littérature scientifique qui lui est afférente, mais aussi de la citer. C'est ainsi que, dans notre cas, David Smith peut se retrouver à citer Donald Fake, tout en ignorant que ce dernier l'a volé et n'a pas produit une recherche effective et originale (quand bien même le texte peut être classé comme « original »). Tout au plus, David Smith lisant Donald Fake pourra s'étonner de la proximité des idées⁵ selon la quantité et la qualité du travail « ajoutées » par Donald Fake à son imitation. Dans cet exemple expérimental, David Smith est la victime qui s'ignore de Donald Fake.

D'autre part, le texte produit par ce dernier, constitue bien un faux. En effet, le texte de David Smith a été utilisé intégralement par Donald Fake pour produire son

⁵ On est ici dans le cadre expérimental de textes ne mobilisant pas de données empiriques précises.

propre texte en ne travaillant pratiquement pas, tandis qu'en le publiant il laisse penser à tous que c'est bien le fruit de son travail et de sa recherche. La fraude par imitation introduit donc la potentialité de nous retrouver face à des textes qui seront « originaux » mais qui, pourtant, n'auront pratiquement pas été écrits par leur auteur déclaré (cf. le pourcentage indiqué ci-dessus) et encore moins « pensés » par ce dernier. Comme pour le plagiat, le drame collectif se joue dans l'absence de création intellectuelle, et donc dans la perspective d'un appauvrissement collectif sur les plans intellectuel, culturel et scientifique.

Mais alors, comment interpréter le tableau 2 si les dix conséquences du plagiat (du comportement plagiaire, plus exactement) mises en évidence par Bergadaà (2015b) sont également des conséquences reconnaissables de l'imitation ? Doit-on en conclure qu'une imitation est en réalité un plagiat ? Ou bien doit-on en déduire que les dix conséquences proposées par Bergadaà (2015b) ont une valeur plus générale et ne concernent pas seulement le plagiat ? Nous proposons de répondre par l'affirmative à cette dernière question. Cela revient à considérer que ce que nous pourrions appeler « le modèle des dix conséquences » est pertinent, *a priori*, pour identifier et caractériser toute fraude académique.

L'hypothèse d'une disparition du plagiat et les incidences sur la typologie des fraudes académiques

De la même façon que le plagiat et sa massification paraissent consubstantiels à l'émergence et à la diffusion de l'Internet, la maturation de nouvelles technologies de l'information et de la communication ne peut que laisser augurer de l'apparition de nouvelles fraudes (au regard des comportements, des actes ou des conséquences) et d'une évolution de la typologie de la fraude académique. C'est, par hypothèse, ce à quoi nous sommes en train d'assister avec l'intelligence artificielle et les potentialités que nous explorons dans cette communication. Cela induit un nécessaire questionnement sur l'évolution de la typologie des fraudes académiques.

Comment décrire aujourd'hui cette typologie ? Il faut remarquer que tant l'examen de la littérature scientifique que celui des acteurs de l'intégrité académique suggèrent que la fraude académique dominante est constituée par le plagiat. Néanmoins d'autres types de fraude existent et qui sont bien décrits dans cette littérature spécialisée. Il faut citer en particulier la falsification et la fabrication (Gutwirth & Christiaens, 2015). Néanmoins un examen de cette typologie et un effort pour en proposer une substruction (Becker 2002, Lazarsfeld 1970) permettent de proposer un espace de propriétés dans lequel apparaît un quatrième type de fraude : la sous-traitance (Figure 1).

		Objet de la tromperie	
		Auteur	Contenus
Registre technique	Travestissement	1 Plagiat	2 Falsification
	Substitution	4 Sous-traitance	3 Fabrication

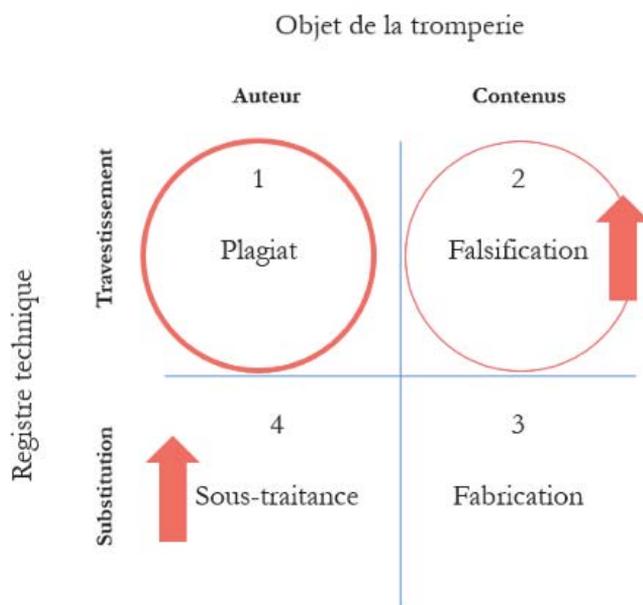
Source : JL 2022 – Laboratoire Métis - EM Normandie & IRAFPA

Figure 1 – Espace de propriété justifiant d’une typologie de la fraude académique

L'espace de propriétés décrit par la Figure 1 permet en effet de caractériser, de distinguer et d'articuler quatre types de fraudes à partir de l'objet de la tromperie, d'un côté, et du registre technique utilisé, d'un autre côté. En décrivant ainsi la typologie fondamentale de la fraude académique, il est possible de mieux saisir les conséquences potentielles de l'apparition de la fraude par imitation à l'aide d'une intelligence artificielle. On peut en effet faire l'hypothèse, même si les données nous manquent pour être plus affirmatif, que le plagiat est aujourd'hui la fraude académique dominante, du moins quantitativement. Cela correspond, d'une part, aux conditions technologiques actuelles et, d'autre part, aux exigences productivistes du système pédagogique et scientifique actuel.

Cela nous mène au modèle décrit par la figure 2 : le plagiat comme fraude académique dominante, et la falsification et la sous-traitance comme fraudes complémentaires⁶. Divers indicateurs et indices peuvent suggérer un renforcement dans la période récente des pratiques de sous-traitance et de falsification.

⁶ Les rares quantifications dont nous disposons suggèrent que la dominance de telle ou telle fraude varie probablement de façon importante d'un champ scientifique à un autre. Voir par exemple : Henriot, P. & Ouzoulias, P. (2021). Promouvoir et protéger une culture partagée de l'intégrité scientifique. Rapport parlementaire au nom de l'Office parlementaire d'évaluation des choix scientifiques et technologiques. 143 pages.



Source : JL 2022 – Laboratoire Métis - EM Normandie & IRAFPA

Figure 2 – Hypothèse sur la dynamique actuelle de la fraude académique

Intéressons-nous à la fraude par sous-traitance (un commanditaire fait écrire un texte par un tiers et y appose sa signature : c'est une pratique largement répandue dans l'édition). En effet, sans être en mesure d'évaluer l'intensité actuelle de cette fraude, l'émergence de l'intelligence artificielle et sa potentielle capacité à écrire des textes « originaux » mais « faux » pourrait avoir pour effet de « couper l'herbe sous le pied » d'une partie des bénéficiaires de ce marché de la délinquance intellectuelle. Autrement dit, cela affecterait ce champ professionnel si mal connu en risquant, potentiellement, de le faire disparaître. Ici, l'émergence et la diffusion d'une nouvelle modalité technologique de fraude pourrait faire disparaître une forme de professionnalisation de la fraude existante.

En effet, des plateformes telles que Speedwrite pourraient avoir pour conséquence de rendre inutile le recours à un sous-traitant rédacteur, d'autant plus qu'il paraît probable que l'offre de fraude par recours à une plateforme IA sera bien plus compétitive économiquement. Cela nous mène donc à nous interroger, dans une logique prospective, sur l'évolution possible du champ de la déviance intellectuelle et scientifique. En effet, si la fraude par imitation « IA » entraîne l'inutilité du plagiat (d'autant qu'à sa différence elle ne peut être décelée par les outils recherchant de similarités textuelles) alors cela entraînerait également l'inutilité de la

fraude par sous-traitance (d'autant qu'à sa différence son coût est extrêmement modique mais nécessite un tout petit peu de travail). Cela nous mènerait à une nouvelle situation décrite par la figure 3, si l'on considère l'imitation comme un sous-type de la « Fabrication ».

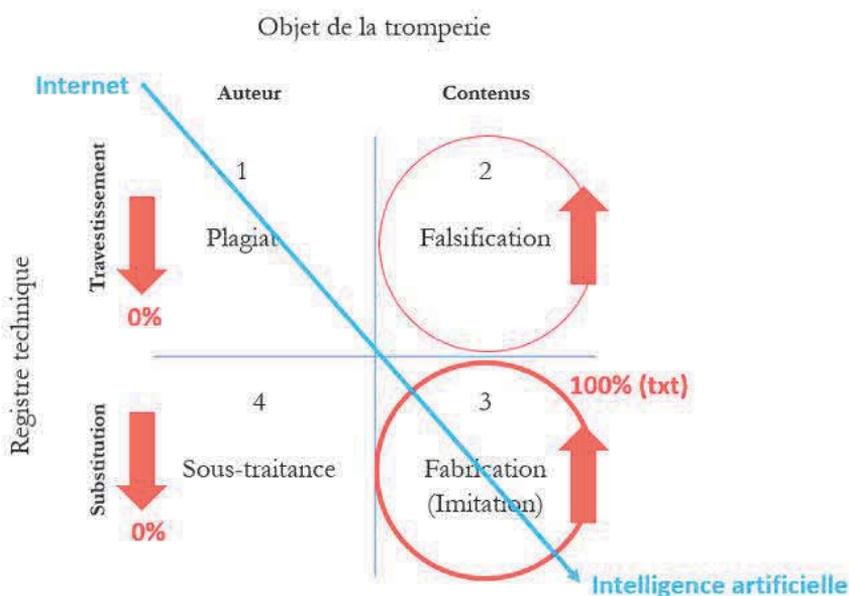


Figure 3 - Hypothèse sur la dynamique de la fraude académique induite par l'IA

Dans cette hypothèse, le plagiat serait d'autant plus destiné à disparaître que la même intelligence artificielle qui ouvre la possibilité de la fraude par imitation a déjà ouvert la possibilité à une lutte beaucoup plus efficace contre le plagiat. Si l'on se fie aux potentialités que semble dessiner notre expérimentation, la fraude par imitation présente des avantages qu'aucun autre type de fraude n'offre :

- a. Rapidité et souplesse de production ;
- b. Indétectabilité (ou plutôt « indécidabilité ») à l'aide des outils de détection de similarités textuelles actuels ;
- c. Prix modique (de 20 USD par mois à 100 USD pour un an avec un calibrage de la production de 6000 « prévisions » par mois pour la plateforme utilisée⁷).

⁷ Le texte produit par Donald Fake (Tableau 1) représente environ 150 « prévisions ».

Tous ces éléments incitent à étudier plus avant cette hypothèse prospective d'une mutation profonde de la typologie de la fraude académique au profit de cette forme particulière de fabrication que nous avons appelée ici imitation.

Ce travail sur les espaces de propriétés permettant de proposer des idéaux-types de fraude a une limite importante : il porte sur le document diffusé et son attribution publique, et non sur l'intégralité du processus de sa production. Il semble ainsi que certains types de fraude échappent à cette typologie. Par exemple, le vol de données d'expérimentation par un chercheur B à un chercheur A, B réussissant à publier un traitement et une interprétation de ces données avant A⁸. Le travail de recherche sur les espaces de propriétés permettant d'identifier et de caractériser les actes de fraude académique reste donc ouvert.

Les défis incidents pour l'intégrité

Les perspectives ouvertes par des plateformes comme Speedwrite, et plus largement par l'intelligence artificielle appliquée à la production de textes, posent de nouveaux défis pour les référents intégrité académique ou scientifique, ainsi que leurs établissements ou laboratoires. En particulier ces évolutions peuvent nécessiter de clarifier ce qui a pu, par le passé, paraître clair : qu'est-ce qu'un travail académique ou scientifique ? Quelle part de travail ou quels segments des processus de production de connaissance peuvent être « délégués » aux machines sans affecter le caractère « scientifique » d'un texte ? Le dialogue entre les sciences de l'intégrité et la réflexion épistémologique paraît devoir devenir plus étroit.

Évidemment, l'apparition de nouvelles pratiques frauduleuses s'appuyant sur de nouvelles technologies pose inmanquablement la question de leur détection. En la matière, il n'est pas trop hasardeux de penser que ce sera l'intelligence artificielle qui permettra de lutter contre l'usage frauduleux de l'intelligence artificielle, tout comme l'Internet a permis et permet de lutter contre l'usage frauduleux d'Internet, notamment le plagiat. Ainsi si le couple Internet-Plagiat se trouvait remplacé par le couple IA-Imitation, c'est alors à l'aide d'une IA que l'on pourra déceler la production d'une IA. Les expériences de prévention/lutte contre le risque de fraude académique et recourant à l'IA existent et semblent d'ores et déjà donner satisfaction en ce qui concerne les examens type « partiels » (Ajmi, 2021). Mais, dans la mesure où nous ne disposons pas aujourd'hui des outils pour déceler des textes imités produits par une IA, il est légitime de s'inquiéter de cette période aux bornes

⁸ Cette importante limite a pu être identifiée grâce à la communication de Decullier, E. & Sèdes, F. (à paraître). Méconduites académiques : exploration d'une distinction potentielle de genre. Actes du 2ème Colloque International de Recherche et Action sur l'Intégrité Académique : « Les nouvelles frontières de l'intégrité ». Université de Coimbra. Juin 2022.

temporelles imprécises entre l'émergence d'une nouvelle opportunité de fraude et le moment où l'on deviendrait capable de la détecter et de l'objectiver. Il s'est passé dix ans, à peu près, entre l'émergence de l'Internet et la massification des pratiques de plagiat (Bergadaà, 2015a). Nous n'avons donc que quelques années pour nous préparer à faire face à cette nouvelle pratique frauduleuse de l'imitation.

Une des pistes de contre-mesure est dans l'approche stylistique (Agnès, 2021) qui permet déjà d'améliorer la détection de plagiat ou la multiplicité cachée des auteurs par la discrimination de différents styles internes un texte. Cette même approche stylistique semble pouvoir être suivie de façon intertextuelle. Ainsi si l'on dispose d'un texte attribuable de manière sûre à un auteur, il devrait être possible de détecter si un texte ultérieurement signé par ce même auteur n'est pas de lui.

Frédéric Agnès (2021) confirme les possibilités ouvertes par l'intelligence artificielle pour lutter contre ce nouveau type de fraude que nous appelons imitation. En effet l'auteur souligne qu'il sera possible d'« identifier des documents possédant des structures proches, dans les thèmes abordés et l'enchaînement des idées présentées, ce qui pourrait révéler un vol d'idées ». Cela décrit assez bien ce que produit une plateforme comme Speedwrite. L'auteur confirme : « il pourra être possible de détecter des similitudes entre des textes reformulés ou traduits ». Il s'agit là de l'approche dite « sémantique », et non plus « syntaxique » (Agnès, 2021), basée sur la construction de modèles de proximité sémantique, applicables même à travers un processus de traduction.

Au-delà de cette course contre-la-montre pour réduire le délai au-delà duquel on devient capable de lutter contre une nouvelle forme de délinquance de la connaissance, il faut poser le problème de la surenchère (technologique) et de l'escalade éducative, dissuasive ou répressive : où allons-nous ? Poser cette question est une manière de suggérer que la réponse à la fraude intellectuelle ne peut être que très partiellement dans les outils pour la détecter ou l'empêcher. En effet, raisonner en termes de binôme Fraude/Contre-mesures enferme dans une escalade d'innovations dont on ne peut penser qu'elle puisse avoir un terme prochain.

Ainsi, tout en mettant au point, sur le court terme, des contre-mesures, il est de toute évidence nécessaire de soutenir l'analyse critique des cadres sociologiques, organisationnels et éthiques dans lesquels s'insère ce processus d'escalade. Ainsi la vacuité axiologique dans laquelle se trouve, parfois à leur corps défendant, nos étudiants comme nos collègues est sans doute tout aussi importante à traiter que le problème technologique d'une IA permettant de déceler une imitation textuelle. Il ne faut pas que la lutte contre la fraude académique s'enlise dans des diversions (Gutwirth & Christiaens, 2015). C'est d'autant plus important de le souligner que l'ampleur des possibilités que semble ouvrir l'intelligence artificielle pourrait occulter encore mieux ce qui nous paraît essentiel pour le mouvement des sciences de

l'intégrité : promouvoir philosophiquement l'intégrité elle-même et directement auprès de nos collègues, de nos étudiants, de nos instances. Comment imaginer une lutte contre la fraude qui ne soit pas une participation au risque de détournement d'attention (Gutwirth & Christiaens, 2015) ?

Le seul moyen de sortir de cette escalade est en effet de rétablir des communautés scientifiques fondées sur les principes épistémologiques et éthiques qui ont fait naître la science : absence de compétition, temps pour travailler, confiance entre tous et donc absence de défiance, etc. Autrement dit, la soumission de la science et des communautés scientifiques aux règles du libéralisme randien est et reste le fond du problème et non l'apparition de telle ou telle technologie ; quand bien même il est nécessaire d'en construire une connaissance précise.

Conclusion : l'intégrité académique et la décision publique ou privée

De tels développements technologiques ouvrent un vaste champ d'investigation à la réflexion éthique et aux sciences de l'intégrité. Bien sûr dans cette communication nous nous sommes concentrés principalement sur les enjeux afférents à l'enseignement supérieur et à la recherche. Mais il faut également remettre les enjeux pour l'enseignement supérieur et la recherche en perspective en les mettant en relation avec de plus larges enjeux sociologiques et politiques.

La perspective de voir des étudiants comme des chercheurs disposer d'outils IA qui leur permettraient de produire des textes sans les avoir écrits, ou presque, et encore moins pensés - mais qui pourraient passer les barrières de la correction comme de l'évaluation - peut venir bouleverser plusieurs enjeux fondamentaux du champ intellectuel et scientifique, notamment :

- la validité sociologique et politique des bénéfices sociaux afférents à la détention des titres et grades universitaires (et autres, d'ailleurs) ;
- la recevabilité sociologique et politique des résultats scientifiques publiés dans les revues ou sous la forme d'ouvrages.

Plus prosaïquement, ces évolutions sociotechniques posent le problème de la confiance publique dans la valeur accordée aux diplômes (et donc dans la légitimité de leurs détenteurs), d'une part, et dans la valeur accordée aux connaissances scientifiques publiées (notamment quant à leur solidité et à leur indépendance). Derrière cette question de la confiance publique dans les productions de tout le système d'enseignement supérieur et de recherche se cache un problème encore plus fondamental pour nos démocraties : la confiance dans et la légitimité de la décision publique comme privée.

En effet comment les citoyens, demain, pourraient-ils avoir confiance dans et juger légitimes les décisions (prises par des responsables publics comme privés) à propos desquelles :

- a. ils douteraient des compétences comme des connaissances réelles des décideurs (sans parler de leurs « mérites ») ;
- b. ils douteraient de la solidité des connaissances sur la base desquelles ces décisions sont prises ou à l'aune desquelles elles sont justifiées ?

De ce point de vue, l'émergence probable de la fraude par imitation ne révolutionnera pas la donne éthique fondamentale. Par contre, elle semble pouvoir aggraver (et peut-être accélérer leur concrétisation) les risques politiques, sociaux et économiques afférents à la prolifération de la fraude intellectuelle et des conduites déviantes afférentes. La structuration et le renforcement du mouvement pour les sciences de l'intégrité reste donc une nécessité absolue pour nos démocraties.

Bibliographie

Agnès, F. (2021). Évolution didactique du logiciel de détection des similitudes : l'exemple de Compilatio. In : Bergadaà, M. & Peixoto, P. L'urgence de l'intégrité académique. Éditions EMS, (pp.429-442)

Ajmi, O. (2021). Mise en place d'une solution d'e-surveillance des examens : les pierres d'achoppement de la confiance. In : Bergadaà, M. & Peixoto, P. L'urgence de l'intégrité académique. Éditions EMS, (pp.393-410)

Becker, H. S. (2002). Les ficelles du métier – Comment conduire sa recherche en sciences sociales. Éditions La Découverte.

Becker, H. S. (1985). Outsiders: Études de sociologie de la déviance. Éditions Métailié.

Bergadaà, M. (2015a). Une brève histoire de la lutte contre le plagiat dans le monde académique. Questions de communication, 27, (pp. 171-188). <https://doi.org/10.4000/questionsdecommunication.9787>

Bergadaà, M. (2015b). Le plagiat académique, Comprendre pour agir. L'Harmattan.

Gutwirth, S. & Christiaens, J. (2015) Les sciences et leurs problèmes : la fraude scientifique, un moyen de diversion ?. Revue interdisciplinaire d'études juridiques, 74, (pp. 21-49). <https://doi.org/10.3917/riej.074.0021>

Klein, O. & Yzerbyt, V. (2016). La fraude scientifique: Un symptôme du dysfonctionnement de la science. Pour la Science, Vol. 470, (pp. 36-42) <http://hdl.handle.net/2078.1/181037>

Lazarsfeld, P. (1970). Philosophie des sciences sociales, nrf/Éditions Gallimard.

