

Fraudes, exagérations et auteurs fantômes : la fiabilité de la littérature scientifique en question

Si la crédibilité de la science est régulièrement mise à mal par des affaires de fraude retentissantes, elle souffre également de mauvaises pratiques et d'inconduites plus discrètes, tout aussi préjudiciables. Des scientifiques consacrent leur activité de recherche à l'étude de la production académique avec des projets destinés à identifier les anomalies et erreurs dans la littérature, à en comprendre les causes et à étudier les mécanismes possibles de correction. C'est le cas du projet [NanoBubbles](#), auquel contribuent les auteurs de ce billet.

By **Frédérique Bordignon**^{1,2}, **Guillaume Cabanac**^{3,4}, **Cyril Labbé**⁵ and **Maud Bernisson**⁶

¹ Ecole des Ponts, 77455 Marne-la-Vallée, France

² LISIS, INRAE, Univ Gustave Eiffel, CNRS, 77455 Marne-la-Vallée, France

³ IRIT UMR 5505 CNRS, Université Toulouse 3 – Paul Sabatier, France

⁴ Institut universitaire de France (IUF), Paris, France

⁵ Université Grenoble-Alpes, CNRS, Grenoble INP, Laboratoire d'Informatique de Grenoble, Grenoble, France

⁶ Radboud University, Nijmegen, The Netherlands

Les affaires faisant grand bruit sont celles qui quittent la bulle de la communauté scientifique parce qu'elles sont relayées, voire révélées, par les médias, et parfois même parce qu'elles finissent devant les tribunaux. Le lecteur curieux pourra consulter le site [RetractionWatch](#) qui les recense et les explique. Mais citons ici quelques cas pour montrer la diversité des problèmes que les spécialistes déclinent en termes de falsification et fabrication de données, et de plagiat, des pratiques contraires à l'intégrité scientifique. En 2021, un neuroscientifique a révélé une falsification d'images utilisées comme preuves scientifiques dans un article annonçant la validation d'une hypothèse formulée 14 ans plus tôt pour expliquer la maladie d'Alzheimer. Cette révélation a ébranlé une voie de recherche considérée comme prometteuse jusqu'alors, après avoir attiré des millions de dollars d'investissement. En 2011, un psychologue néerlandais réputé a fini par avouer avoir truqué des expériences et donc fabriqué des données de toutes pièces, produisant des résultats spectaculaires mais bidonnés, souvent repris sans méfiance par la presse, relayant alors des idées fausses, comme par exemple le fait que les employeurs jugent les candidats à l'emploi plus compétents lorsqu'ils ont une voix masculine...

Experts et faussaires [\[1\]](#)

À propos des plagats, il est maintenant coutume de dire qu'ils ont peu de conséquence sur la validité de la science. Cependant, ils affectent durement les auteurs plagiés qui doivent souvent batailler pour faire reconnaître leur préjudice. Le plagiat concerne autant le vol d'idées que le vol de travaux et a toujours existé. On rapporte que [Ptolémée s'est attribué](#) au II^e siècle les mesures astronomiques réalisées 300 ans plus tôt par Hipparque à Rhodes, alors qu'il prétendait les avoir faites sur la côte égyptienne.

Les conséquences de la fraude scientifique sont très diverses : elles peuvent malheureusement passer inaperçues mais aussi conduire à la stigmatisation des fraudeurs, à la mutilation ou la [mort de patients](#). Elles peuvent aussi infléchir l'acceptation des politiques sanitaires (comme avec la

promotion de l'hydroxychloroquine dont l'efficacité n'a jamais été prouvée contre la Covid-19, ou la fausse relation causale entre vaccin contre la rougeole et autisme en 1998 qui a déclenché une chute brutale de la vaccination au Royaume-Uni)[2].

Si ce genre de fraudes est parfois révélé par des lanceurs d'alerte, témoins de la tricherie, d'autres sont mises au jour par de fins limiers qui scrutent les publications scientifiques. Certains en ont fait une spécialité, voire une activité professionnelle, à l'instar de la microbiologiste [Elisabeth Bik](#) qui détecte les manipulations d'images telles que les coupes histologiques.

Détecteurs de fraudes

Pour traquer ces fraudes, et devant l'ampleur de la tâche, Guillaume Cabanac et Cyril Labbé, développent des outils informatiques qui scannent les 6 millions d'articles scientifiques publiés par an. Les neuf détecteurs du *Problematic Paper Screener* (PPS) signalent 12 000 articles problématiques à ce jour. Certains « articles » sont totalement dénués de sens (générés automatiquement avec *SCIgen* ou *Mathgen*). Des articles de biologie mentionnent des séquences ADN erronées (que détecte *Seek&Blastn*) ou utilisent des lignées cellulaires contaminées. Des « expressions torturées » [3] résultant de couper-paraphraser-coller (produisant « nucleic corrosive » à la place de « nucleic acid », ou « counterfeit consciousness » pour « artificial intelligence », par exemple) caractérisent les deux-tiers des articles problématiques. Les faussaires emploient ces logiciels paraphraseurs pour masquer leurs plagiat, quitte à détruire le sens du texte original.

Le PPS recense aussi des articles « aux pieds d'argile » reposant sur des références non fiables : rétractées ou critiquées par ailleurs. Des scientifiques préoccupés par de telles publications réévaluent les articles signalés par le PPS et postent des rapports de ré-évaluation sur [PubPeer](#). Des échanges avec les auteurs et les maisons d'édition s'ensuivent, amenant parfois à des corrections voire des rétractations. En automatisant la recherche de problèmes dans les 120 millions d'articles produits à ce jour, notre PPS facilite la réévaluation des résultats les plus contestables... une activité capitale pour accumuler des savoirs fiables.

Exagérations et fausses promesses

Bien loin des cas extrêmes cités précédemment, les écrits scientifiques peuvent souffrir d'embellissements trompeurs pour le lecteur. Les chercheurs ont inconsciemment tendance à survendre leurs résultats et à en exagérer les bénéfiques potentiels, en usant de tournures linguistiques trop positives, comme Frédérique Bordignon et ses collègues l'ont montré au tout début de la crise Covid avec une étude [4] sur l'explosion du recours aux adjectifs *effective*, *promising*, et *novel* (que la revue [ACS Catalysis avait déjà bannis](#) de ses titres et résumés en 2017) pour qualifier les résultats obtenus. Ces manipulations rhétoriques sont fréquentes, surtout dans le résumé d'articles qui doit donner envie au lecteur de s'intéresser à toute l'étude. Il existe une déformation plus grave de la réalité appelée *spin*, lorsque les auteurs vont au-delà de l'exagération linguistique et ne présentent que leurs meilleurs résultats, ceux qui sont les plus attractifs et sauront ensuite faire parler d'eux. Malheureusement, ces promesses se retrouveront sans nuance dans les communiqués de presse, puis les articles grand public, véhiculant de faux espoirs et fausses informations.

Auteurs fantômes

Maud Bernisson analyse aussi comment l'industrie pharmaceutique utilise les communications scientifiques pour diffuser un message marketing sur les bénéfiques d'un médicament nouvellement breveté, ou encore en introduisant un nouveau concept au bénéfice d'un médicament pour augmenter ses ventes (par exemple, le diagnostic de « pseudoaddiction » a été inventé pour démontrer que les patients qui souffrent devraient voir leur dose de médicaments à base d'opiacé augmenter (échantillon de l'étude : un patient [5])). Des agences de communication privées jouent le

rôle de prête-plume (*ghostwriters*) en rédigeant des articles dont les conclusions abondent dans le sens de l'entreprise, puis elles invitent des chercheurs complices à les signer en leur nom [6].

Enfin, parmi les cas extrêmes de *ghostwriting*, citons les *paper mills*, ces usines à papiers qui proposent des articles clé en main pour quelques centaines ou milliers d'euros. Malheureusement, les articles sont de piètre qualité, souvent fondés sur des données fabriquées, des images truquées ou plagiées, ou encore des textes dénués de sens, qui sont autant d'indices de leur provenance douteuse (et que le PPS traque).

Des auteurs virtuels

Désormais, ce sont les capacités d'outils d'intelligence artificielle (IA) qui permettent la génération d'images et de textes qui sonnent juste mais dont le fondement scientifique est questionnable. Sans complètement rejeter ce progrès, certains éditeurs se mobilisent déjà pour intégrer dans leur politique éditoriale l'interdiction de recourir à des outils d'IA pour la rédaction d'un article scientifique. La plupart du temps, ces politiques éditoriales n'autorisent pas l'inclusion de ChatGPT comme auteur. En effet, l'auteur d'une publication doit pouvoir répondre de son travail.

La question de qui peut, doit et ne peut pas être considéré comme auteur se pose sans cesse sous un éclairage nouveau, qu'il s'agisse de [l'ajout de son chat ou de son hamster en co-auteur](#), ou encore de l'utilisation du pseudonyme [Camille Noûs](#) à des fins revendicatives... mais il conviendra de trouver les solutions nécessaires à l'amélioration et au maintien de l'intégrité de la littérature scientifique. C'est ce à quoi s'attèlent les chercheurs du projet [NanoBubbles](#), financé par des fonds européens et mené en France et aux Pays-Bas par une équipe interdisciplinaire de physiciens, historiens, sociologues, philosophes, linguistes et informaticiens.

[1] Bessy, C., & Chateauraynaud, F. (2014). *Experts et faussaires : Pour une sociologie de la perception* (2e éd., augmentée d'une postface). Éditions Pétra.

[2] Voir aussi le documentaire "[La fabrique de l'ignorance](#)" sur Arte TV.

[3] Cabanac, G., Labbé, C., & Magazinov, A. (2021). *Tortured phrases: A dubious writing style emerging in science. Evidence of critical issues affecting established journals* (arXiv:2107.06751). arXiv. <https://doi.org/10.48550/arXiv.2107.06751>

[4] Bordignon, F., Ermakova, L., & Noel, M. (2021). Over-promotion and caution in abstracts of preprints during the COVID-19 crisis. *Learned Publishing*, leap.1411. <https://doi.org/10.1002/leap.1411>

[5] Michaels, D., & Michaels, D. (2020). *The Triumph of Doubt: Dark Money and the Science of Deception*. Oxford University Press.

[6] Sismondo, S., & Doucet, M. (2010). Publication Ethics and the Ghost Management of Medical Publication. *Bioethics*, 24(6), 273–283. <https://doi.org/10.1111/j.1467-8519.2008.01702.x>