

Guillaume Cabanac, traqueur de fake science

NUMÉRIQUE 

La revue scientifique *Nature*, l'une des plus anciennes et réputées au monde, l'a surnommé *deception sleuth*. Littéralement : « fin limier de la supercherie ». Ses talents pour détecter les publications non fiables valent à Guillaume Cabanac, 39 ans, enseignant-chercheur de l'Institut de recherche en informatique de Toulouse¹, de figurer au palmarès des dix personnes qui ont le plus marqué la science en 2021 selon la revue².

En collaboration avec l'informaticien grenoblois Cyril Labbé et le mathématicien russe Alexander Magazinov, il a conçu et développé le logiciel Problematic Paper Screener³ (« Radar à papiers problématiques »). Chaque nuit, le détecteur Tortured de cet outil passe au crible les 120 millions de publications indexées par la base de données bibliographiques Dimensions pour y débusquer des expressions torturées. Cette appellation désigne des assemblages incongrus de mots, symptomatiques de méconduites et fraudes qui, hélas, passent le filtre de certaines revues défailtantes voire frauduleuses, et gagnent une petite fraction de la littérature scientifique.

Détectives invisibles et expressions torturées

Pour comprendre pourquoi, il faut se mettre dans la peau des fraudeurs... Pour eux, un simple copier-coller afin de récupérer des paragraphes d'articles sérieux, briques élémentaires pour en créer un faux, n'est plus possible du fait des détecteurs de plagiat. À moins de... paraphraser le texte avec des synonymes, grâce à des logiciels dédiés. Quitte à obtenir les fameuses expressions torturées. Coller le résultat sans relire et voici que fleurissent des déceptions du rein au lieu d'insuffisances rénales. Des périls de la poitrine au lieu de cancers du sein. Corrosif nucléaire remplace acide nucléaire. Et conscience contrefaite supplante intelligence artificielle !

Le trio a choisi une approche de réévaluation post-publication participative. En clair : un collège invisible d'une cinquantaine de « détectives », des chercheurs et des amateurs, identifient de nouvelles expressions torturées que Guillaume Cabanac injecte dans le logiciel. Puis, des articles qui en contiennent, pointés comme problématiques, sont signalés sur PubPeer.org, plateforme administrée par deux chercheurs du CNRS, où des scientifiques, du monde entier et de toutes disciplines, apportent

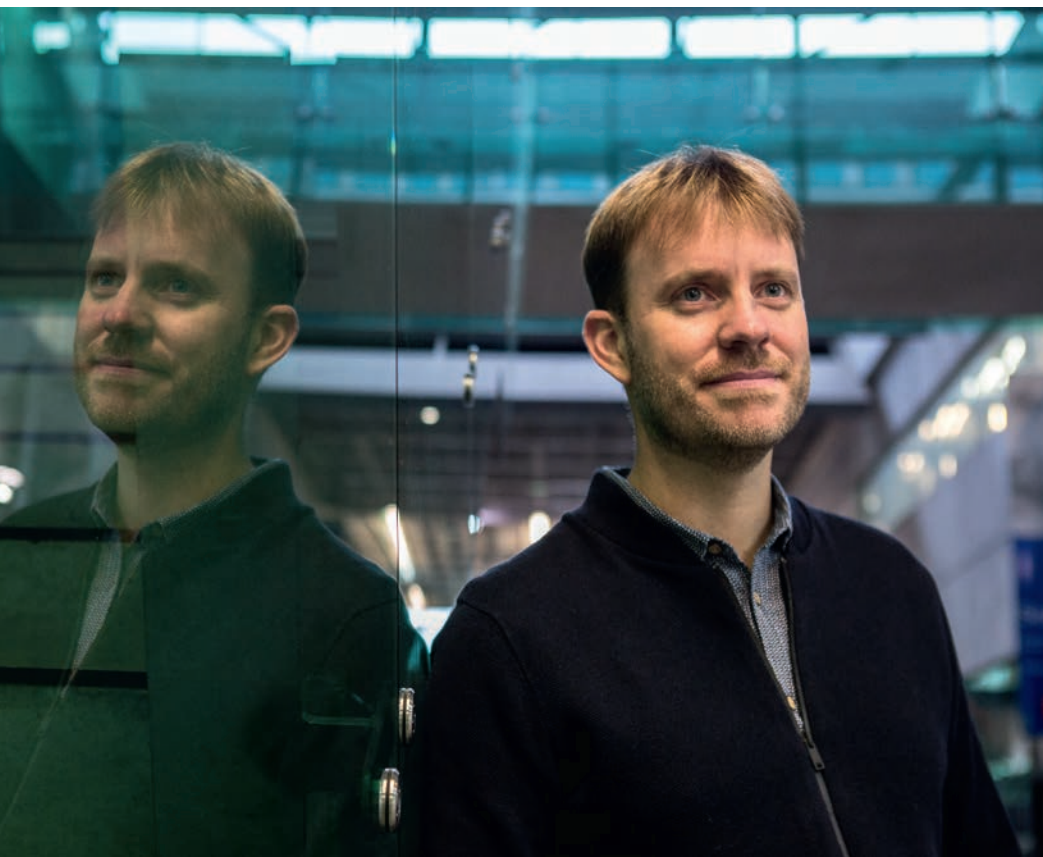
INFORMATIQUE « Corrosif nucléaire » pour « acide nucléaire » ou « conscience contrefaite » à la place d'« intelligence artificielle »... Ce sont les expressions torturées que Guillaume Cabanac chasse dans la littérature scientifique à l'aide d'un logiciel qu'il a cocréé. Objectif : identifier les publications frauduleuses.

PAR GAËL HAUTEMULLE

leur expertise en les commentant pour les réévaluer. En ce moment, le logiciel traque 550 expressions torturées dans les 16 000 nouvelles indexations quotidiennes de Dimensions. Sur près de 6 millions d'articles publiés en 2021, il en a ainsi pointé plus de 1 500 comptant au moins deux expressions torturées, critère garantissant qu'au moins 98 % d'entre eux contiennent, après réévaluation, d'authentiques plagiats. Et le phénomène va croissant, puisque, avec les mêmes critères, les chiffres tombent à 1 135 en 2020, 638 en 2019, et 416 en 2018. Par ailleurs, un autre des détecteurs de leur Radar à papiers problématiques, développé en collaboration avec Cyril Labbé et des chercheurs australiens, leur a permis de pointer plus de 700 articles d'oncologie, publiés entre 2014 et 2018, qui comportaient des erreurs de séquences génétiques et sont à ce jour pourtant déjà cités plus de 20 000 fois dans d'autres articles scientifiques⁴.

“ Je n'ai besoin que d'un crayon, d'un calepin, d'un ordinateur et d'une connexion Internet ! ”

1. Unité CNRS/ institut national polytechnique de Toulouse/ Université Toulouse-Paul Sabatier. 2. <https://www.nature.com/immersive/d41586-021-03621-0> 3. <https://www.irit.fr/~Guillaume.Cabanac/problematic-paper-screener> 4. <https://doi.org/10.26508/Isa.202101203> 5. Dans « A note on science and democracy », Robert K. Merton, 1942, *Journal of Legal and Political Sociology*.



© FREDERIC SCHIEBER

Son collègue Cyril Labbé s'amuse souvent en « *découvrant certaines publications "surréalistes". Mais, pour Guillaume, l'indignation finit souvent par l'emporter...* », raconte-t-il. Moins de trois « pollutions » tous les 10 000 articles, cela peut sembler peu. Mais « *accepterait-on que trois des 10 000 vols aériens quotidiens s'achèvent en crash ?* », insiste Guillaume Cabanac. *La science se construit, tel le mur d'une maison, avec des briques. L'usage d'éléments non fiables voire pourris – et réutilisés par la communauté scientifique – pourrait, à terme, menacer l'édifice !* »

Dépollueurs contre moulins à papier

Cette ardeur au « nettoyage » trouve peut-être son origine dans les valeurs familiales et morales du chercheur. Un grand-père représentant syndical, animé par les lumières de la formation tout au long de la vie, le Toulousain aux origines modestes a toujours été attiré par la culture et croit en l'éthos de la science, notamment au désintéressement et au scepticisme organisé. Or, du fait de la massification des publications (qui a presque doublé en dix ans), « *ces deux garde-fous théoriques par le sociologue Robert K. Merton⁶ sont aujourd'hui piétinés par les "moulins à papier"⁶* », fabricants d'articles-patchwork au kilomètre qui volent et paraphrasent des publications fiables. Le tout dans un univers ultra-concurrentiel où les chercheurs sont souvent sous l'emprise du fameux *publish or perish* (« publier ou périr »)...

C'est en DUT d'informatique de gestion que la recherche a capté sérieusement l'attention de Guillaume

ETHOS DE LA SCIENCE

Normes morales encadrant l'activité des membres de la communauté scientifique d'après le sociologue des sciences américain Robert K. Merton.

INFODÉMIE

Contraction de « information » et « épidémie » pour désigner un déluge d'informations dont bon nombre se révèlent fausses voire malveillantes.

Cabanac. Son stage, en tant que programmeur auprès d'un doctorant de l'université Paul-Sabatier de Toulouse, s'avère une expérience ardue, stimulante et fondatrice. Il poursuit en master à l'université et décroche un financement de thèse du ministère sur l'activité d'annotation collective de textes numériques, dans le même laboratoire qui l'accueillera comme maître de conférences en 2009, neuf mois après sa soutenance.

Idées, solutions et programmation

Motivé par l'interdisciplinarité, le lien social et les relations humaines, il se rapproche de chercheurs en sociologie des sciences et, en 2016, soutient une

habilitation à diriger des recherches intitulée « Interroger le texte scientifique ». Sa recherche, qui demande très peu de moyens, fait pour lui écho à ses origines sociales : « *Je n'ai besoin que d'un crayon, d'un calepin, d'un ordinateur et d'une connexion Internet!* » Il se dit *full stack* (« polyvalent » en jargon d'informaticien) : chercheur, il a des idées. Ingénieur, il résout les problèmes. Technicien, il programme et met les mains dans le cambouis numérique. Enseignant, il estime que ses cours devant ses étudiants nourrissent sa recherche.

Aujourd'hui, Guillaume Cabanac est membre nommé au sein du Conseil scientifique de l'Institut des sciences humaines et sociales du CNRS. Il travaille sur le volet préventif de la dépollution au sein du projet franco-hollandais Nanobubbles⁷ à la problématique cruciale : « *Comment, quand et pourquoi la science échoue-t-elle à s'autocorriger ?* » Et le trio de détectives est sollicité par une dizaine de maisons d'édition comme ACM, Elsevier, IEEE, Institute of Physics, SAGE, et Springer Nature, pour intégrer le Radar à papiers problématiques en amont de la chaîne de publication.

Vu l'actuelle infodémie, la démarche du fin limier toulousain trouve de plus en plus d'écho. L'Office français de l'intégrité scientifique (Ofis) juge même à présent que « *les activités de commentaires post-publication* » relèvent « *des activités ordinaires du chercheur, au même titre que les activités traditionnelles d'évaluation par les pairs* »⁸, rappelle l'informaticien. Enthousiaste, il invite chacun à le rejoindre pour, ensemble, dépolluer la littérature scientifique. ||

6. Expression fréquente dans les publications scientifiques de référence comme *Nature* (par exemple <https://doi.org/10.1038/d41586-021-00733-5>) 7. Projet ERC Synergy <https://www.irit.fr/erc-nanobubbles/> 8. Note du 27 septembre 2021, <https://www.hceres.fr/en/node/30562040>