

Compte rendu de la Comité Scientifique de la Faculté de Santé du 22 05 2025

Thèmes et invités :

- 1) **Dépollution de la littérature scientifique : Guillaume Cabanac (IRIT)**
- 2) **Présentation de l'IRIT** (Institut de Recherche en Informatique de Toulouse) par Jean-Marc Pierson, Directeur de l'IRIT, Julie Mauclair Maitre de Conférence à l'IRIT et Mathieu Balaguer, Maitre de Conférence en sciences de la rééducation et réadaptation.

Documents associés : présentation ppt de Guillaume Cabanac.

<https://www.irit.fr/~Guillaume.Cabanac/>

1 / Dépollution de la littérature scientifique : par le Pr Guillaume Cabanac

Guillaume Cabanac est professeur d'informatique à l'Université de Toulouse et membre de l'Institut de recherche en informatique de Toulouse (IRIT UMR 5505 CNRS). Il est titulaire d'une chaire de recherche fondamentale de l'Institut universitaire de France (IUF) intitulée « dépollution de la littérature scientifique ».

Ses travaux visent à identifier des publications non fiables par fouille de textes, notamment au sein du projet ERC Synergy 'Nanobubbles' questionnant le processus d'auto-correction de la science. Il développe le « Problematic Paper Screener » qui traque des publications non fiables, souvent même vendues par les maisons d'édition prédatrices. Au quotidien, ses signalements d'articles problématiques, contenant notamment des expressions torturées, ont donné lieu à plus de 3 500 rétractations d'articles. Ses recherches ont été distinguées dans le palmarès "Nature's 10" présentant « dix personnes qui ont contribué à façonner la science en 2021 » selon la revue Nature.

Dépollution de la littérature scientifique : traque de fake science avec le « problematic paper screener »

Dia 1 :

- ✓ Identifier les bombes rouges grâce à une lecture critique de la littérature scientifique,
- ✓ Les bombes rouges altèrent/fragilisent le mur de la connaissance,
- ✓ Les moyens d'action : Evaluation en post publication sur pub peer (<https://pubpeer.com/>), militer auprès des éditeurs pour faire rétracter les articles concernés et faire reconnaître aux institutions l'importance de s'attaquer à cette problématique en expansion.

Dia 2 : Une publication non conforme aux règles scientifiques est comme un trou dans un mur : le mur en est fragilisé, et il peut même menacer de s'écrouler. La persistance des bombes rouges dans le panel d'articles disponibles témoigne d'un déficit dans l'efficacité collective à corriger la littérature.

Article « Growth in a time of debt » (2010) : il y a des erreurs dans les équations et les conclusions ne tiennent plus. Cependant, ce papier a mobilisé des politiciens du monde entier pour justifier des politiques d'austérité. Un erratum a ensuite été publié de nombreuses années plus tard.

Les errata servent à construire une maçonnerie autour du trou qui allait s'effriter. C'est une publication corrigée ou retractée qui est représentée ici symboliquement.

3500 articles ont été retractés grâce à la plateforme développée à l'IRIT (Problematic Paper Screener) qui est en libre accès. ([https://dbrech.irit.fr/pls/apex/f?p=9999:1:::::\)](https://dbrech.irit.fr/pls/apex/f?p=9999:1:::::)

Dia 3 : Le serment des scientifiques : en réaction aux « sciences ariennes », à partir du questionnement « qu'est-ce que la science ? ». La vision anglo saxon de la science est positiviste : « Avec des règles de bonnes conduites, on va y arriver »

Les piliers de la science concernent

- la **communication** des résultats : Les résultats, positifs ou négatifs, devraient être examinés avec la même équité, quelle que soit leur origine... nous en sommes loin.

- le **désintéressement**

- le **septicisme « organisé »** : c'est l'évaluation par les pairs de nos jours => d'abord douter puis se faire convaincre par les résultats.

Dia 4 : Depuis de 1975, l'idée d'un champ de bataille/ un champ de force s'est imposée avec Pierre Bourdieu, consistant à conquérir un terrain scientifique comme un champ de bataille, puis de le verrouiller par une position/des arguments d'autorité ; par exemple en formant des étudiants qui relayent la même vision.

Dia 5 et 6 : Aujourd'hui, il y a un emballement à produire des articles scientifiques : 6 millions d'articles par an actuellement et la pente n'a pas été infléchie par la pandémie de Covid.

La Chine est passée devant les USA.

Cette croissance du nombre d'articles publiés est entre autres la conséquence d'un doublement du nombre de journaux en 20 ans. Cela a entraîné un problème dans l'expertise pour reviewer, et les comités éditoriaux de nombreux journaux sont « infiltrés » de non spécialistes.

Dia 7 à 12 : Une nouvelle menace : **Les paper mills**, qui vendent littéralement articles et places d'auteurs. Le sujet n'est pas récent mais il est peu connu, déjà évoqué en 2013 dans *Science* et en 2021 dans *Nature*. Cette pollution vient par exemple de générateurs de texte aberrant faisant apparaître des expressions « torturées » (cf infra).

Dia 13 et 14 : Il existe des détectives scientifiques (sleuths) dont le travail permet d'aboutir à rétracter des articles scientifiques. Le Pr Cabanac coordonne 30 personnes qui font cela sur la plateforme « Problematic paper screener ».

Dia 15 à 18 : Elles.ils recherchent par exemple des « tortured phrases » : p esteem au lieu de p value, Innocent bayes au lieu de naive bayes,... Elles.ils ont ressorcé environ 6000 expressions torturées de ce type qui permettent de détecter ces pratiques. En 2021, Elsevier a retiré 4000 articles comportant des phrases torturées.

Dia 19 : Spinbot : on insère sur le site internet « Spinbot » un abstract volé dans un article de la littérature et on obtient un texte avec des synonymes permettant de passer sous le radar des détecteurs de plagiat => Il faut alerter les éditeurs en chef si on repère cela ! (Dia20). On ressource actuellement 3 articles problématiques sur 10000 (Dia 21)

Dia 27 : Question de l'utilisation responsable des LLMs (tel que chat GPT).

Chat GPT ne produit pas de phrases torturées. Certains éditeurs autorisent son utilisation et il faut alors le noter dans l'article et le citer dans les remerciements ou dans le chapitre méthodes (mais ne peut être considéré comme un coauteur). Dans JAMA, il y a une case à cocher si chat

GPT est utilisé. Certaines institutions interdisent d'utiliser Chat GPT par crainte de fuite du potentiel scientifique et technique.

Dia 30 : Machine à générer des références : Exemple dans Plos one => 18 références « hallucinées », ça a des références qui n'existent pas. Il y a maintenant des IA qui vérifient l'existence des informations !

Dia 31 et 32 : La génération d'images peut aussi poser des problèmes : image sans sens avec des problèmes d'échelle ou de logique ou qui n'ont aucune pertinence scientifique, mais sont « plaisantes ».

Dia 35 à 38 : Si on laisse ces papiers en ligne, ils seront lus par des humains ET des IA. Le détecteur est appelé colosse aux pieds d'argile (Feet of clay detector) : 3 exemples d'articles où des références de bibliographie ont été retractées après signalement.

Dia 36 à 40 : Afin de freiner l'effet domino, il faut dès à présent identifier les meta-analyses où des articles ont été retractés et refaire les statistiques afin d'en voir les conséquences.

Un article faisant la synthèse du problème a été publié dans JAMA internal medicine (Mars 2025) accompagné d'un éditorial insistant sur l'importance de cette démarche.

Tout ceci illustre les limites (voire le côté délétère) de l'évaluation quantitative (nombre d'articles, nombre de citations, voire nombre de reworking faits !!) avec au final un risque majeur d'érosion de la confiance de la société vis-à-vis de la science.

Parmi les détectives scientifiques, il y a actuellement très peu d'universitaires (5 dans le monde dont le Pr Cabanac).

Au total un double objectif :

- curatif → dépolluer la littérature
- préventif → anticiper les nouvelles fraudes

Les moyens : Ré-évaluation décentralisée soutenue par « **problematic paper screener** » de l'IRIT avec 3500 retractations à son actif.

PubPeer est un site internet qui permet aux utilisateurs (principalement anonymes) d'émettre des commentaires sur des articles scientifiques en post-publication, mais aussi de signaler des soupçons de manquements à l'éthique scientifique. Le site a mis en évidence des lacunes dans plusieurs articles de grande envergure, qui ont dans certains cas conduit à des retractations et à des accusations de fraude scientifique, comme remarqué par le blog Retraction Watch. Certains chercheurs dénoncent en PubPeer un système de délation, mais *Le Monde* note dans un éditorial que le site « a pourtant mis au jour de véritables fraudes, et n'existe qu'en raison de l'incapacité de la science à répliquer ses résultats et à s'autocorriger efficacement »^[14].

Il s'agit d'un sujet important, et notre communauté devait s'en saisir car c'est actuellement le principal outil d'évaluation post-publication. On peut télécharger l'extension pour Pubmed ou Zotero et les commentaires postés sur pub peer remontent automatiquement.

Seul forum d'évaluation post-publication actuellement disponible, et s'il est largement adopté par la communauté scientifique, Pubpeer pourrait être utile par rapport à la problématique majeure de **reproductibilité** des résultats dans la recherche biomédicale, devenue un problème important avec l'augmentation de la pression publicitaire et la multiplication des journaux.

2/ Présentation de l'IRIT

JM Pierson : L'IRIT est structuré en domaines d'applications stratégiques (santé, autonomie, bien être, Ville intelligente, Aero espace transport), certaines actions étant transverses : calcul – données et IA.

L'IRIT compte environ 600 membres (moitié non permanents et moitié permanents), 5 tutelles et 6 sites en Occitanie, 7 départements et 24 équipes.

J Mauclair a présenté l'axe santé, autonomie, bien-être, dont les compétences se situent au niveau des signaux du vivant (capteurs par exemple), la gestion de données biomédicales, la modélisation et simulation du vivant, l'imagerie autour du handicap.

M Balaguer a détaillé un exemple : informatique et sciences de la rééducation, appliqué à l'analyse clinique de la parole pathologique. C'est un « créneau » peu rentable pour les GAFAM et tout à fait adapté à une équipe de l'IRIT.

Le but est de compléter l'évaluation clinique, et de proposer une évaluation automatique pour pallier la variabilité des outils et des mesures et aussi faire la part entre l'évolution d'une pathologie et la variabilité de l'outil. But in fine : avoir un plus grand degré de précision que l'oreille humaine et en moins de temps, en utilisant un système automatisé de reconnaissance de la parole.

Exemple d'autres projets : mesurer la sévérité du trouble de la parole avec un score de parole (ADAPT et SAMI) ; mesurer l'altération de la communication...