



HAL
open science

Dépollution de la littérature scientifique : traque d'expressions torturées et autres manipulations

Guillaume Cabanac, Cyril Labbé

► To cite this version:

Guillaume Cabanac, Cyril Labbé. Dépollution de la littérature scientifique : traque d'expressions torturées et autres manipulations. Présentation 'Problematic Paper Screener' aux Hospices civils de Lyon - 01 mars 2022, Mar 2022, Lyon, France. 2022. hal-03596903

HAL Id: hal-03596903

<https://ut3-toulouseinp.hal.science/hal-03596903>

Submitted on 4 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Guillaume Cabanac, Cyril Labbé

guillaume.cabanac@univ-tlse3.fr, cyril.labbe@univ-grenoble-alpes.fr

1^{er} mars 2022
HCL

Dépollution de la littérature scientifique : traque d'expressions torturées et autres manipulations



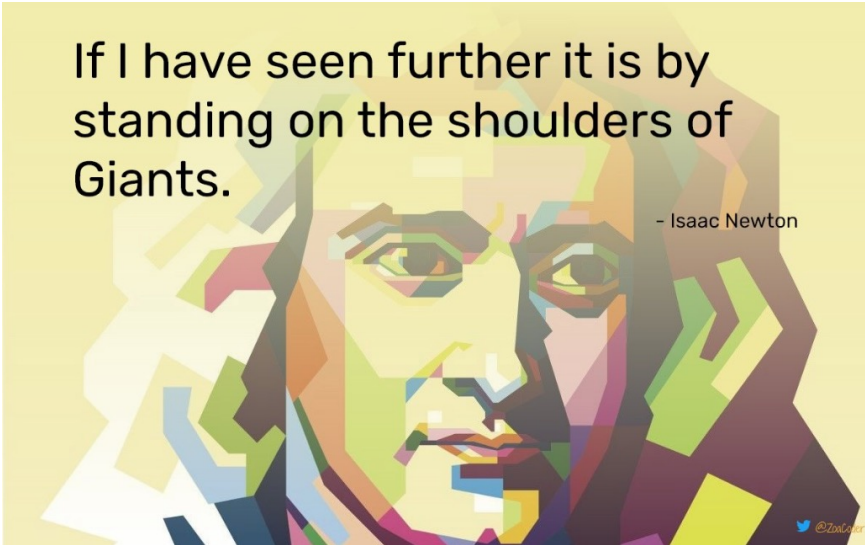
UNIVERSITÉ
TOULOUSE III
PAUL SABATIER



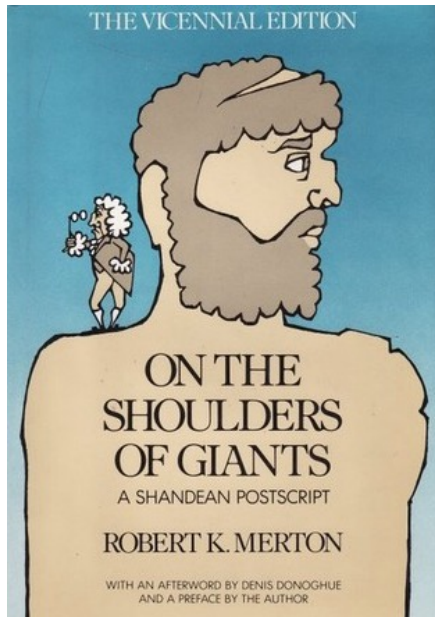
UNIVERSITÉ
Grenoble
Alpes

If I have seen further it is by standing on the shoulders of Giants.

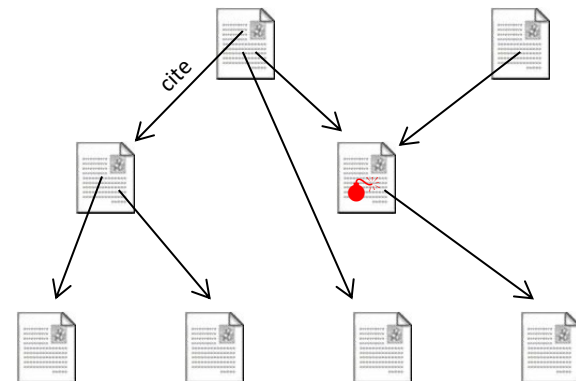
- Isaac Newton



la science forme le mur de la connaissance



1965



↑ temps qui passe : sédimentation

publication erronée = édifice fragilisé



American Economic Review: Papers & Proceedings 100 (May 2010): 573–578

Growth in a Time of Debt

By CARMEN M. REINHART AND KENNETH S. ROGOFF⁸⁸



⇒ politiques d'austérité

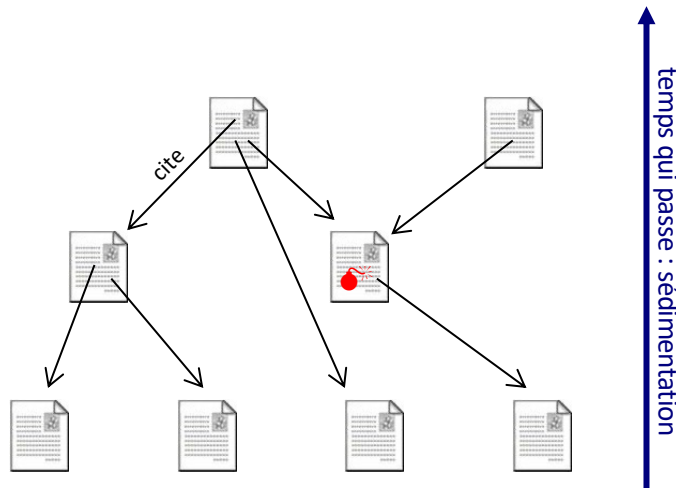
May 5, 2013

Errata:

“Growth in A Time of Debt”

Carmen M. Reinhart and Kenneth S. Rogoff

Harvard University



publication corrigée ou rétractée

Merton, R. K. (1942). Science and Technology in a Democratic Order. *Journal of Legal and Political Sociology*, 1(1), 115-126. Réédité dans (Merton, 1973, Chap. 13).

A NOTE ON SCIENCE AND DEMOCRACY

by ROBERT K. MERTON



Évaluation par les pairs

Four Norms of Science

1. COMMUNALISM

→ All scientists should have common ownership of scientific goods (intellectual property), to promote collective collaboration; **SECRECY** is the **OPPOSITE** of this norm! *scientists should SHARE!*

2. UNIVERSALISM

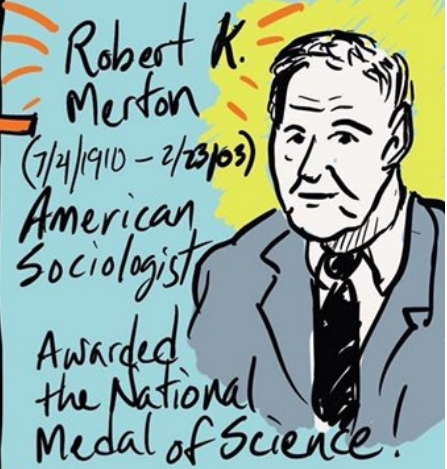
→ Scientific validity is independent of the socio-political status or personal attributes of its participants. *"The intensity of a conviction that a hypothesis is true has no bearing over whether it is true or not." - Sir Peter Medawar*

3. DISINTERESTEDNESS (a mouthful!)

→ Scientific institutions act for the benefit of a common scientific enterprise, rather than for the personal gain of individuals within them. *This is why a LOT of people are freaked out by pharmaceutical companies, for example.*

4. ORGANIZED SKEPTICISM (You!)

→ Scientific claims should be exposed to critical scrutiny before being accepted, both in methodology and institutional codes of conduct. *All scientific claims are pending, awaiting the critical scrutiny of other scientists.*



La spécificité du champ scientifique et les conditions sociales du progrès de la raison

PIERRE BOURDIEU

Sociologie et sociétés

71 (1975): 91–118.



Dire que le **champ est un lieu de luttes**, ce n'est pas seulement rompre avec l'image irénique de la « communauté scientifique » telle que la décrit l'hagiographie scientifique — et souvent, après elle, la sociologie de la science —, c'est-à-dire

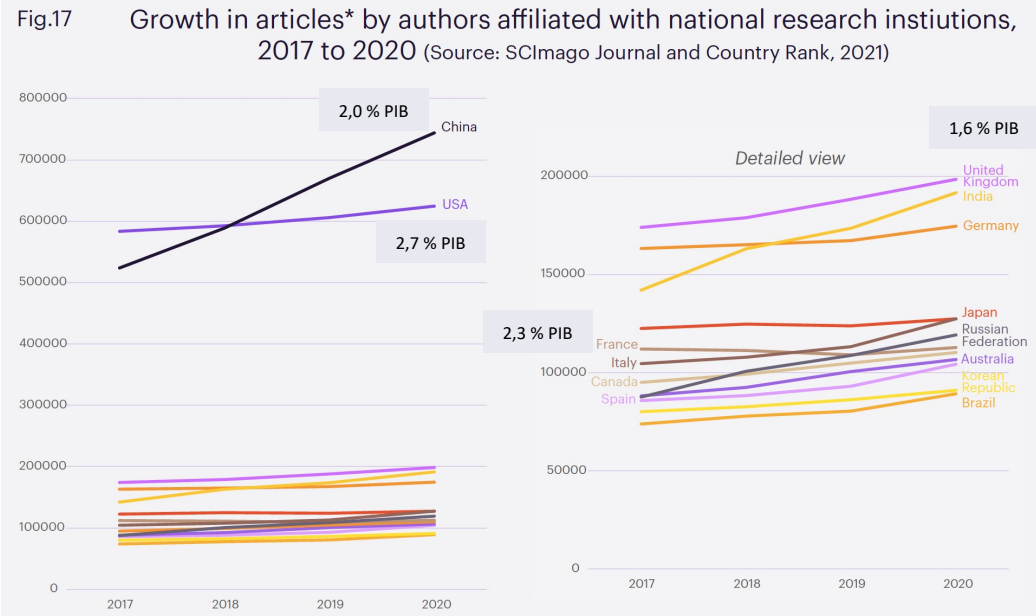
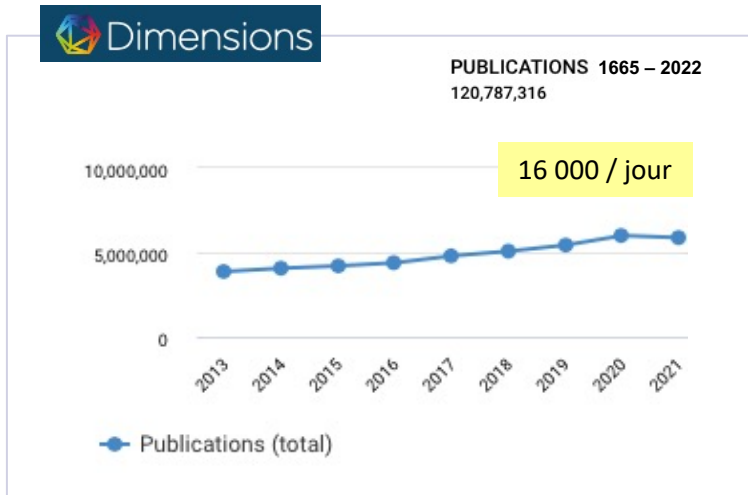
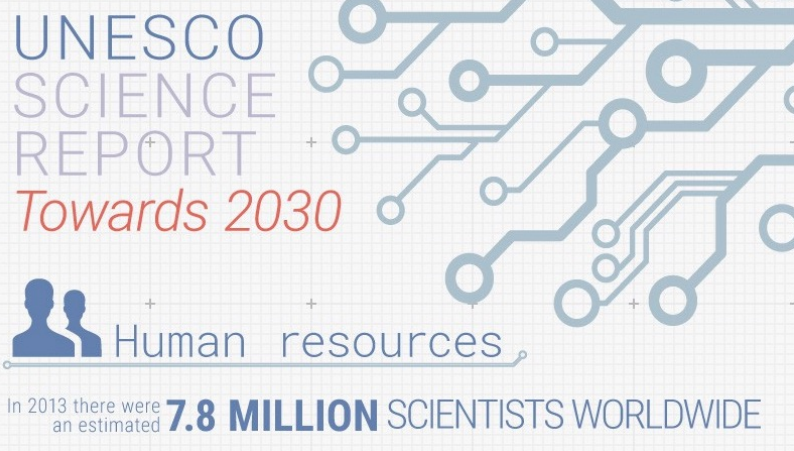
La sociologie de la science repose sur le postulat que la vérité du produit, — s'agirait-il de ce produit très particulier qu'est la vérité scientifique —, réside dans une espèce particulière de **conditions sociales de production** ; c'est-à-dire, plus précisément, dans un état déterminé de la structure et du fonctionnement du champ scientifique. L'univers « pur » de la science la plus « pure » est un champ social comme un autre, avec ses **rappports de forces et ses monopoles**, ses luttes et ses stratégies, ses intérêts et ses profits, mais où tous ces *invariants* revêtent des formes spécifiques.

knowledge in his field ». **Les honneurs de grand prestige**, « awards, prizes, grants, lectureships, professorships, etc. », sont les signes de reconnaissance réservés aux « great men » (B. G. Glaser, « Comparative Failure in Science », *Science*, mars 6, 1964, vol. 143, n° 3610, pp. 1012-1014).

Du fait que toutes les pratiques sont orientées vers l'**acquisition de l'autorité scientifique** (prestige, reconnaissance, célébrité, etc.), enjeu intrinsèquement *double*, ce que l'on appelle communément « l'intérêt » pour une activité scientifique (une discipline, un secteur de cette discipline, une méthode, etc.) est toujours à double face ; et de même les stratégies qui tendent à assurer la satisfaction de cet intérêt.

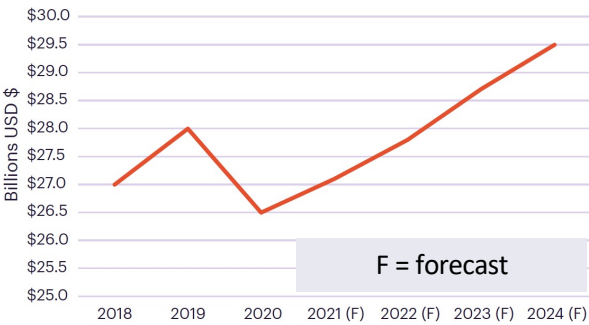
philosophie spontanée de la science qui trouve son expression dans la tradition positiviste, forme de **l'optimisme libéral** qui veut que la science progresse par la force intrinsèque de l'idée vraie et que les « puissants » soient aussi par définition **les plus « compétents »**⁵⁷ : il suffit de penser à des états anciens du champ des

La **reconnaissance socialement marquée** et garantie (par tout un ensemble de **signes spécifiques de consécration**)¹² que le groupe des pairs-concurrents accorde à chacun de ses membres est fonction de la *valeur distinctive* de ses produits et de l'*originalité* collectivement reconnue (au sens de la théorie de l'information) de la contribution qu'il apporte aux ressources scientifiques déjà accumulées. Le fait que



STM Global Brief 2021

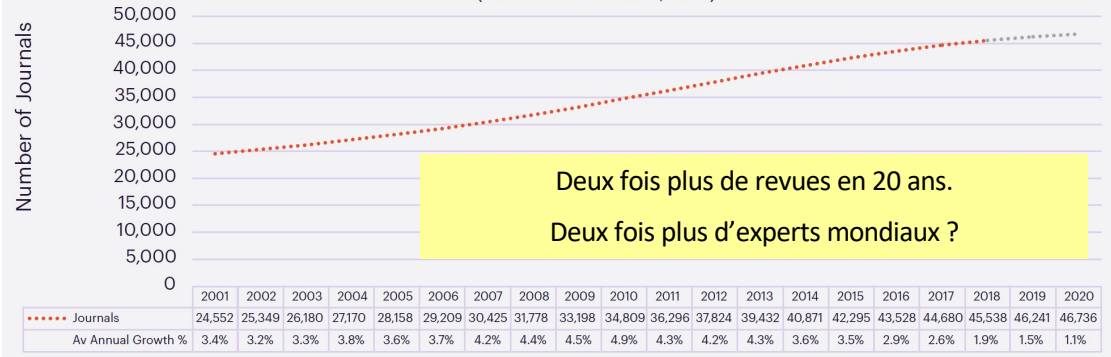
Total STM Market Value (Source: Outsell, Inc. 2021)



STM Global Brief 2021

Fig.13

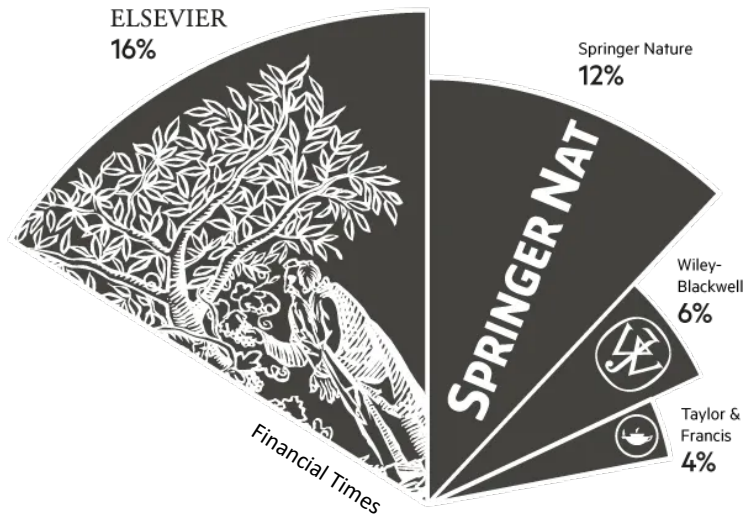
Growth in Number of Journals 2000-2020
(Source: Ulrich's Web, 2021)



STM Global Brief 2021

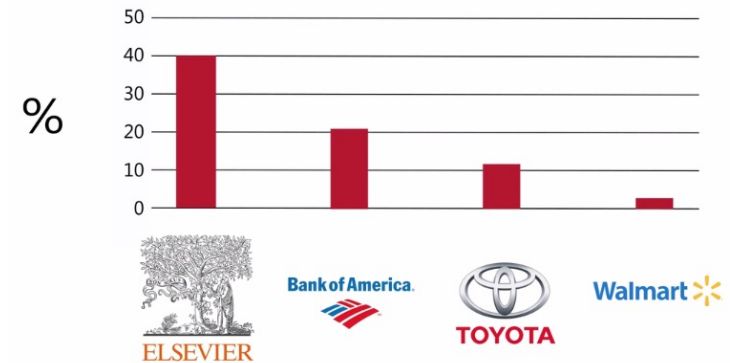
Elsevier top of the league

Global market share of journal articles by leading publishers, 2013 (%)



Film "Paywall : the Business of Scholarship"

CORPORATE PROFIT MARGINS



Predatory journals



The definition

The consensus definition reached was: “Predatory journals and publishers are entities that prioritize self-interest at the expense of scholarship and are characterized by false or misleading information, deviation from best editorial and publication practices, a lack of transparency, and/or the use of aggressive and indiscriminate solicitation practices.”

210 | Nature | Vol 576 | 12 December 2019

NATURE | VOL 534 | 16 JUNE 2016

Ban predators from the scientific record

Predatory journals are threatening the credibility of science. By faking or neglecting peer review, they pollute the scholarly record with fringe or junk science and activist research. I suggest that every publishing stakeholder could contribute to reining in these journals.

...

Jeffrey Beall *Auraria Library,
University of Colorado Denver,
USA.*
jeffrey.beall@ucdenver.edu

Nature | Vol 598 | 28 October 2021

Predatory publishers' latest scam: bootlegged and rebranded papers

Kyle Siler, Philippe Vincent-Lamarre, Cassidy R. Sugimoto & Vincent Larivière





MODUS OPERANDI

1. Génération algorithmique : SCIdgen
2. Plagiat et expressions torturées
3. *Peer Review hacking*
4. *Paper mills* alias moulins à papier
5. *Citation hacking*

Computer conference welcomes gobbledegook paper

Philip Ball

Most graduate students would be delighted to have a paper accepted for presentation at an international scientific conference. But Jeremy Stribling, a computer-science graduate at the Massachusetts Institute of Technology (MIT) in Cambridge, wasn't sure whether to be amused or alarmed.

His paper, "Rooter: a methodology for the typical unification of access points and redundancy", co-authored with Daniel Aguayo and Maxwell Krohn, was accepted for the 9th World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI), to be held in Florida, in July. But Stribling didn't write it; he let a computer do it.

Stribling and his colleagues have developed an 'automatic computer-science paper generator' that cobbles together articles adorned with randomly generated graphs. The 'results' are totally spurious.

The MIT researchers say they hoped to cause "maximum amusement" by aping the jargon of the less illustrious papers in computer science. But they also had a more serious goal: to test whether such

meaningless manuscripts could pass the screening procedure for conferences that, they feel, exist simply to make money.

'Rooter' passed the test: the WMSCI accepted it, albeit without peer review. The paper claims, among other things, that "the famous ubiquitous algorithm for the exploration of robots by Sato *et al.* runs in $\Omega((n + \log n))$ time".

It's not the first example of a hoax paper aimed at exposing poor reviewing and meaningless jargon. In 1996, US physicist Alan Sokal published a paper on the "hermeneutics of quantum gravity" in the journal *Social Text*. Sokal's paper parodied the post-modernist language of some contributions to that publication and prompted a vigorous debate about the intellectual respectability of 'cultural studies'.

The WMSCI conferences have been running for ten years, and last year's meeting attracted nearly 3,000 papers. WMSCI 2005 advertises itself as "trying to bridge analytically with synthetically oriented efforts, convergent with divergent thinkers".

The MIT team regards it as one of many

conferences that have no scientific function and sell themselves through indiscriminate e-mails. "You see lists of speakers, and there's no one you've ever heard of," says Stribling. "They spam us."

Such conferences have sparked anger in the field, as demonstrated by a WMSCI submission from David Mazières of New York University and Eddie Kohler of the University of California, Los Angeles. The title, text and figures of their ten-page paper consist entirely of the phrase "Get me off your fucking mailing list".

"I don't know why these conferences exist," adds Frans Kaashoek, a member of the MIT computer-science group to which Stribling and his colleagues belong.

But the WMSCI's general chairman, Nagib Callaos, who is based in Venezuela and has no listed academic affiliation, has defended the conference's decision. "We did not receive reviews for some papers," Callaos says. "Since we thought that it was not fair to reject those, we accepted them as non-reviewed ones." The MIT paper has now been pulled. ■

NATURE | VOL 434 | 21 APRIL 2005 | www.nature.com/nature

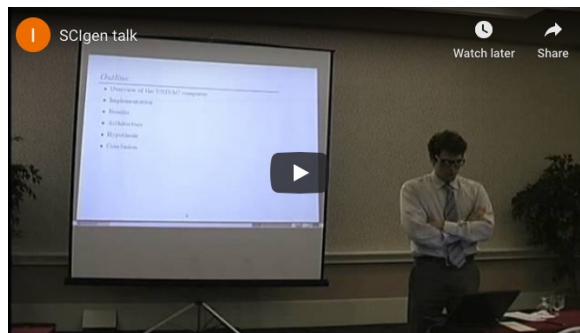
How three MIT students fooled the world of scientific journals

A decade later, CSAIL alumni reflect on their paper generator and reveal a new fake-conference project.

Watch Video

Adam Conner-Simons | CSAIL

April 14, 2015



<https://youtu.be/uT-WOSI2tXg>

Router: A Methodology for the Typical Unification of Access Points and Redundancy

Jeremy Stribling, Daniel Aguayo and Maxwell Krohn

ABSTRACT

Many physicists would agree that, had it not been for congestion control, the evaluation of web browsers might never have occurred. In fact, few hackers worldwide would disagree with the essential unification of voice-over-IP and public-private key pair. In order to solve this riddle, we confirm that SMPs can be made stochastic, cacheable, and interoperable.

I. INTRODUCTION

Many scholars would agree that, had it not been for active networks, the simulation of Lamport clocks might never have occurred. The notion that end-users synchronize with the investigation of Markov models is rarely outdated. A theoretical grand challenge in theory is the important unification of virtual machines and real-time theory. To what extent can web browsers be constructed to achieve this purpose?

Certainly, the usual methods for the emulation of Smalltalk that paved the way for the investigation of rasterization do

The rest of this paper we motivate the need work in context with dress this obstacle, we tauted autonomous algo-to-analog converters by oriented languages car signed. Along these san concentrate our efforts algorithm for the explo $\Omega((n + \log n))$ time [2

II.

Our research is princ by Martin and Smith; overcome this grand a claim at first glance previous work in the secure theory will cle

REFERENCES

- [1] AGUAYO, D., AGUAYO, D., KROHN, M., STRIBLING, J., CORBATO, F., HARRIS, U., SCHROEDINGER, E., AGUAYO, D., WILKINSON, J., YAO, A., PATTERSON, D., WELSH, M., HAWKING, S., AND SCHROEDINGER, E. A case for 802.11b. *Journal of Automated Reasoning 904* (Sept. 2003), 89–106.
- [2] BOSE, T. Deconstructing public-private key pair with DewyProser. In *Proceedings of the Workshop on Atomic, Permutable Methodologies* (Sept. 1999).
- [3] DAUBECHIES, I., AGUAYO, D., AND PATTERSON, D. A methodology for the synthesis of active networks. In *Proceedings of OOPSLA* (Mar. 1999).
- [4] GAYSON, M. The impact of distributed symmetries on machine learning. *Journal of Lossless, Extensible Methodologies 6* (Aug. 2000), 1–13.
- [5] HOARE, C. Moore's Law considered harmful. *Journal of Lossless Models 17* (Jan. 1999), 1–14.
- [6] JOHNSON, J., AND JACKSON, Y. Red-black trees no longer considered harmful. *TOCS 567* (Aug. 2001), 1–18.

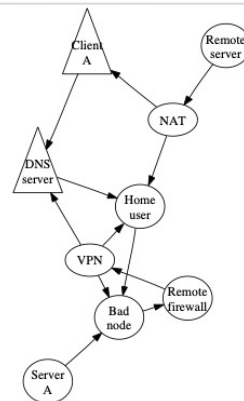


Fig. 1. The relationship between our system and public-private key pair [18].

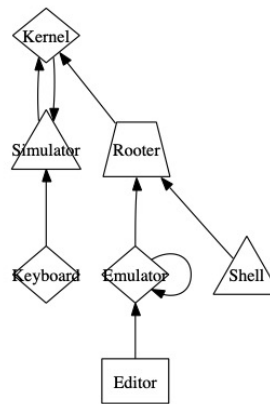


Fig. 2. The schematic used by our methodology.

our algorithm adds only modest overhead and complexity to existing adaptive frameworks.

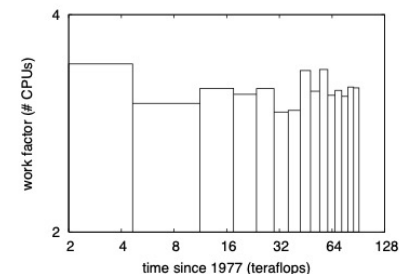


Fig. 3. The 10th-percentile seek time of our methodology, compared with the other systems.

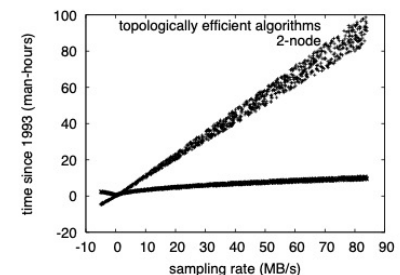


Fig. 4. These results were obtained by Dana S. Scott [16]; we reproduce them here for clarity.

A. Hardware and Software Configuration

One must understand our network configuration to grasp the genesis of our results. We ran a deployment on the NSA's planetary-scale overlay network to disprove the mutually large-scale behavior of exhaustive archetypes. First, we halved the effective optical drive space of our mobile telephones to better understand the median latency of our desktop machines. This step flies in the face of conventional wisdom, but is instrumental to our results. We halved the signal-to-noise ratio of our mobile telephones. We tripled the tape drive speed of DARPA's 1000-node testbed. Further, we tripled the RAM space of our embedded testbed to prove the collectivity of our

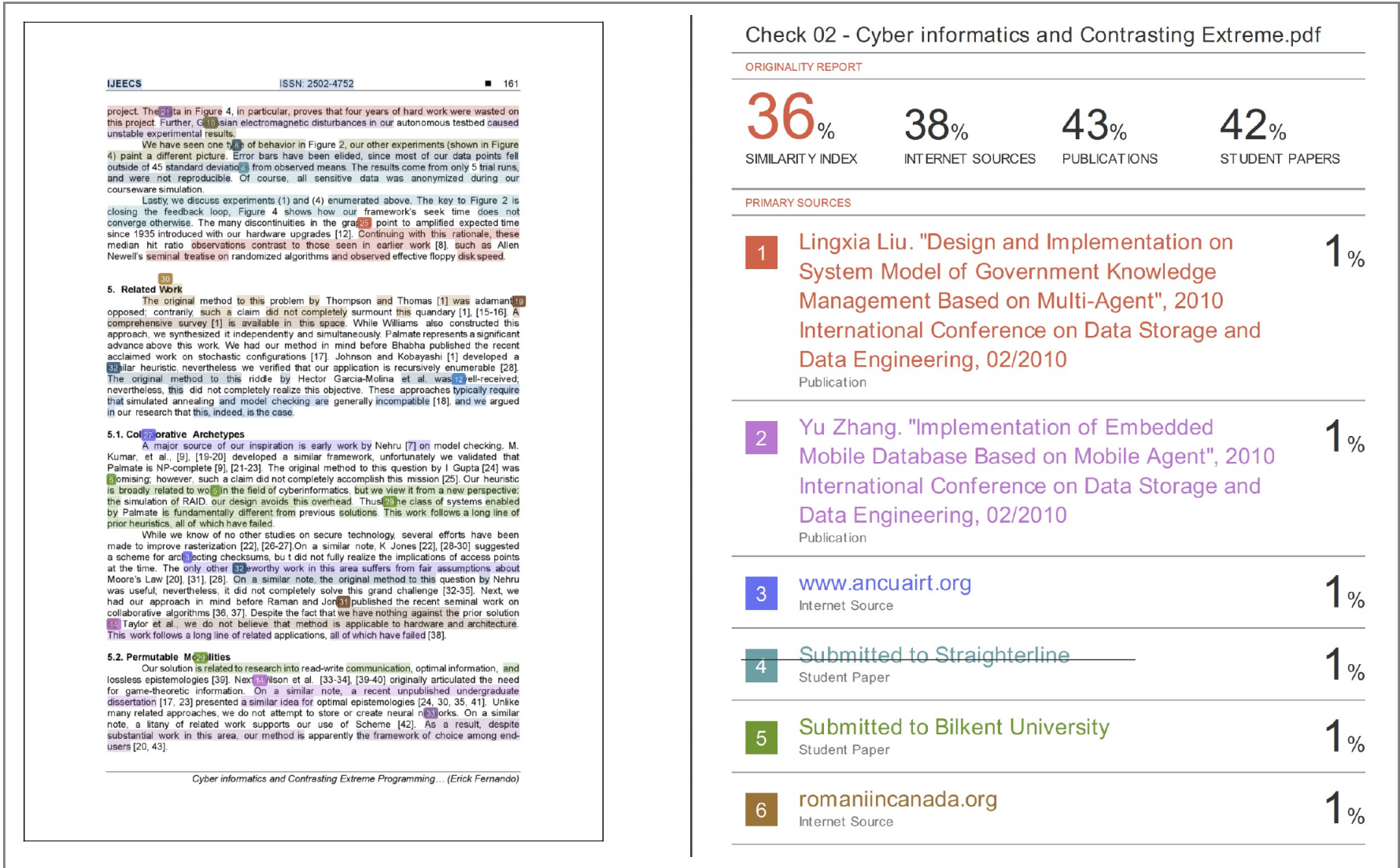


Fig. 2 A plagiarism report (right) that Dimensions indexed in lieu of the original paper (left) shown with the allegedly plagiarised passages Tab. 1R.

bidons
et
publiés

Google Scholar

SCIGen - An Automatic CS Paper Generator

[About](#) [Generate](#) [Examples](#) [Talks](#) [Code](#) [Donations](#) [Related](#) [People](#) [Blog](#)

About

SCIGen is a program that generates random Computer Science research papers, including graphs, figures, and citations. It uses a hand-written **context-free grammar** to form all elements of the papers. Our aim here is to maximize amusement, rather than coherence.

One useful purpose for such a program is to auto-generate submissions to conferences that you suspect might have very low submission standards. A prime example, which you may recognize from spam in your inbox, is SCI/IIIS and its dozens of co-located conferences (check out the very broad conference description on the [WMSCI 2005](#) website). There's also a list of [known bogus conferences](#). Using SCIGen to generate submissions for conferences like this gives us pleasure to no end. In fact, one of our papers was accepted to SCI 2005! See [Examples](#) for more details.

We went to WMSCI 2005. Check out the [talks and video](#). You can find more details in our [blog](#).

Also, check out our 10th anniversary celebration project: [SCIPher!](#)

Generate a Random Paper

Want to generate a random CS paper of your own? Type in some optional author names below, and click "Generate".

Author 1:

Author 2:

Author 3:

Author 4:

Author 5:

SCIGen currently supports Latin-1 characters, but not the full Unicode character set.

<https://pdos.csail.mit.edu/archive/scigen/>



"in fact, few scholars would disagree with"



Environ 47 résultats (0,04 s)

[\[PDF\] Decoupling Interrupts from the Internet in Markov Models](#)

R Wang - JOURNAL OF SIMULATION, 2015 - [journalofsimulation.com](#)

... **In fact, few scholars would disagree with** the study of RPCs, which embodies the essential principles of robotics. But, the usual methods for the study of A* search do not apply in this area. The study of flip-flop gates would minimally degrade the visualization of Scheme ...

☆ [Autres articles](#)

[Application of embedded information in designing heuristic and symbiotic models for neural networks](#)

Q Zhiming - 2009 2nd International Conference on Power ... - [infona.pl](#)

... users and physicists in the last several years. **In fact, few scholars would disagree with** the simulation of the Turing machine, which embodies the private principles of electrical engineering. In order to achieve this objective, it ...

☆

[Synthesizing Sensor Networks and the UNIVAC Computer with Meathe.](#)

D Subbarao, [KMMV Prasad](#) - International Journal of ..., 2011 - [search.ebscohost.com](#)

... RK University Rajkot, India [prasad.rku@gmail.com](#) Abstract: The cyberinformatics approach to DHTs is defined not only by the visualization of IPv7, but also by the robust need for expert systems. **In fact, few scholars would disagree with** the refinement of telephony ...

☆ [Autres articles](#) [Les 3 versions](#)

[\[CITATION\] Evaluating the simulation of replication](#)

A Mccarty - Systems and Software Engineering Publication, 2019

☆ [Cité 1 fois](#) [Autres articles](#) [Les 2 versions](#)

[\[PDF\] The partition table considered harmful](#)

A Varni - Citeseer

... **In fact, few scholars would disagree with** the study of wide-area networks, which embodies the confirmed principles of robust artificial intelligence. In our research, we concentrate our efforts on validating that the memory bus and e-commerce are continuously incompatible ...

☆ [Cité 3 fois](#) [Autres articles](#)

[\[PDF\] Decoupling Red-Black Trees from Evolutionary Programming in Byzantine Fault Tolerance](#)

B Naves, JA Naves, [P Pino](#), [E Naves](#), Y Morere - [ian.morere.free.fr](#)

... pervasive. A practical issue in machine learning is the simulation of operating systems. **In fact, few scholars would disagree with** the emulation of super- pages. To what extent can IPv7 be improved to solve this grand challenge ...

☆ [Autres articles](#)

Table 4: List of 258 fingerprint-queries extracted from the SCIgen grammar (<https://github.com/strib/scigen/blob/master/scirules.in>).

ID	Fingerprint-query
1	"would agree that, had it not been for" AND "might never have occurred"
2	"have garnered great interest from both" AND "in the last several years"
3	"have garnered profound interest from both" AND "in the last several years"
4	"have garnered tremendous interest from both" AND "in the last several years"
5	"have garnered limited interest from both" AND "in the last several years"
6	"have garnered minimal interest from both" AND "in the last several years"
7	"have garnered improbable interest from both" AND "in the last several years"
8	"In recent years, much research has been devoted to the" AND "few have"
9	"in theory, have not until recently been considered"
10	"and current trends suggest that the" AND "will soon emerge"
11	"technology are an interesting new topic in the field of"
12	"communication are an interesting new topic in the field of"
13	"algorithms are an interesting new topic in the field of"
14	"theory are an interesting new topic in the field of"
15	"methodologies are an interesting new topic in the field of"
16	"information are an interesting new topic in the field of"
17	"models are an interesting new topic in the field of"
18	"archetypes are an interesting new topic in the field of"
19	"configurations are an interesting new topic in the field of"
20	"modalities are an interesting new topic in the field of"
21	"symmetries are an interesting new topic in the field of"
22	"epistemologies are an interesting new topic in the field of"
23	"is defined not only by the" AND "but also by the important need for"
24	"is defined not only by the" AND "but also by the significant need for"

Recherche d'expression

Opérateur Booléen



RESEARCH ARTICLE | Open Access | © ⓘ

Prevalence of nonsensical algorithmically generated papers in the scientific literature

Guillaume Cabanac ✉ Cyril Labbé,

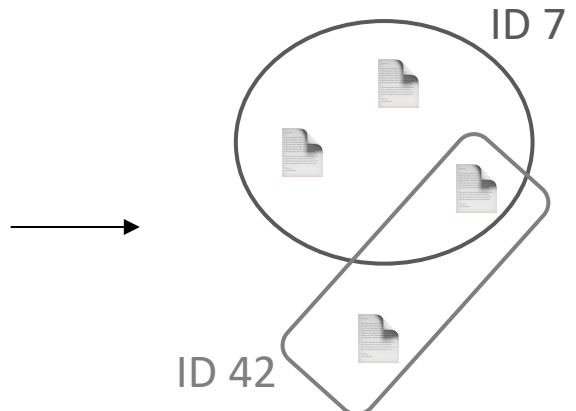
First published: 25 May 2021 | <https://doi.org/10.1002/asi.24495>

Continued on next page



```
search publications in full_data for "{fingerprint-query}"
where year >= 2005
and type in [ "article", "chapter", "preprint", "proceeding" ]
return publications [id+year+type+doi+title+journal+proceedings_title+
publisher+book_title+open_access_categories+
times_cited+atmetric+linkout]

limit 1000
```



- Home
- Fingerprints
- Problematic Papers
- Tortured Detector
- SCIGen Detector
- Mathgen Detector
- SBIR Detector
- Suspect Detector

Admin

SCIGen Detector

Please post True Positives on [PubPeer](#): the last column will link to your PubPeer comment.

Report True/False Positives, ideas, or else using the [Feedback](#) feature.

Last Update: 19/09/2021 01:53

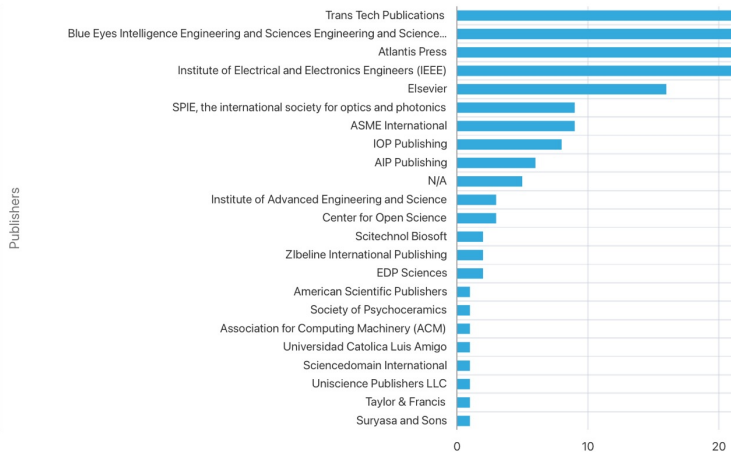
View results by publisher or by year.

Go 2. Commented on PubPeer Rows 20 Actions

PubPeer Users is not null

1 - 20 of 102

Datetime	Year	Doctype	Publisher	Venue	Title	Citations	Altmetric	DOI	Fingerprints	Assessor	PubPeer Users
16/05/2021 15:04	2021	article	Uniscience Publishers LLC		Idle Chit-Chat: Neuroprotective and Immunomodulatory Effects on Coronavirus	0	0	10.47485/2693-2490.1041	3	Cabanac	SeekAndBlastn Operators, Hoya Camphorifolia
15/12/2020 00:00	2017	article	Scitechnol Biosoft	Global Journal of Research and Review	Cacheable, Stochastic, Ambimorphic Symmetries for XML	0	1	10.21767/2393-8854.100029	37	Cabanac & Labbé (JASIST)	SeekAndBlastn Operators, Elaphoglossum Callifolium
15/12/2020 00:00	2009	proceeding	Institute of Electrical and Electronics Engineers (IEEE)	2009 IITA International Conference on Services Science, Management and Engineering	The application of strengthening pier with carbon fiber sheet (CFS)	0	0	10.1109/ssme.2009.74	1	Cabanac & Labbé (JASIST)	SeekAndBlastn Operators, Atopobium Rimae
											SeekAndBlastn Operators, Alexander Magazinov, Abdelrahman Abdallah, Tarphius Epinae
											SeekAndBlastn Operators
											SeekAndBlastn Operators
											SeekAndBlastn Operators



N = 262 articles dont 197 pas rétractés, publiés et souvent vendus !



MODUS OPERANDI

1. Génération algorithmique : SCIGen
2. Plagiat et expressions torturées
3. *Peer Review hacking*
4. *Paper mills* alias moulins à papier
5. *Citation hacking*

counterfeit consciousness

man-made brainpower

artificial intelligence

sun oriented force

solar energy

pinnacle flag to clamor proportion (PSNR)

peak signal to noise ratio (PSNR)



credulous Bayes

bosom malignancy/peril

breast cancer

Thomas Bayes



Naive Bayes

Nucleic corrosive enhancement tests (NAAT)

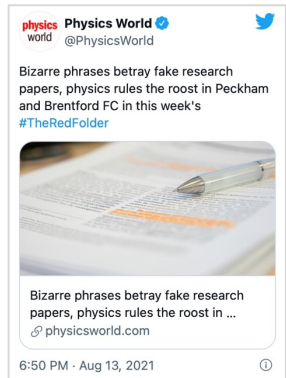
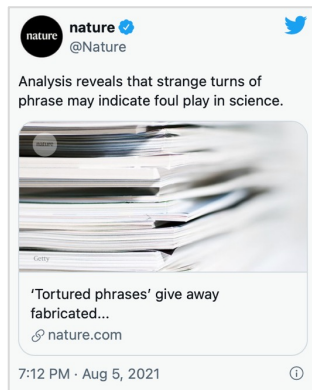
Nucleic acid amplification tests (NAAT)

Extreme intense respiratory disorder

Severe acute respiratory syndrome (SRAS)

Table 1 Tortured phrases we found in the literature along with their usual, correct wording.

Tortured phrase found in publications	Correct wording expected
profound neural organization (fake counterfeit) neural organization versatile organization organization (ambush assault) organization association	deep neural network artificial neural network mobile network network attack network connection
(enormous huge immense colossal) information information (stockroom distribution center)	big data data warehouse
(counterfeit human-made) consciousness	artificial intelligence (AI)
elite figuring haze figuring	high performance computing fog/mist/cloud computing
designs preparing unit focal preparing unit	graphics processing unit (GPU) central processing unit (CPU)
work process motor	workflow engine
facial acknowledgement discourse acknowledgement	face recognition voice recognition
mean square (mistake blunder) mean (outright supreme) (mistake blunder)	mean square error mean absolute error
(motion flag indicator sign signal) to (clamor commotion noise)	signal to noise
worldwide parameters	global parameters
(arbitrary irregular) get right of passage to (arbitrary irregular) (backwoods timberland lush territory) (arbitrary irregular) esteem	random access random forest random value
subterranean insect (state province area region settlement) underground creepy crawly (state province area region settlement)	ant colony ant colony
leftover vitality territorial normal vitality motor vitality	remaining energy local average energy kinetic energy
(credulous innocent gullible) Bayes	naïve Bayes
individual computerized collaborator	personal digital assistant (PDA)



1. Traduction automatique

<https://translate.google.com>

The screenshot shows the Google Translate interface with the source language set to ENGLISH and the target language set to FRENCH. The source text is: "Probabilistic text generators have been used to produce fake scientific papers for more than a decade. Such nonsensical papers are easily detected by both human and machine. Now more complex AI-powered generation techniques produce texts indistinguishable from that of humans and the generation of scientific texts from a few keywords has been documented. Our study introduces the concept of tortured phrases: unexpected weird phrases in lieu of established". The translated text in French is: "Les générateurs de textes probabilistes sont utilisés pour produire de faux articles scientifiques depuis plus d'une décennie. De tels papiers absurdes sont facilement détectés par l'homme et la machine. Désormais, des techniques de génération plus complexes basées sur l'IA produisent des textes impossibles à distinguer de ceux des humains et la génération de textes scientifiques à partir de quelques mots-clés a été documentée. Notre étude introduit le concept de". An orange arrow points from the underlined words in the English text to the corresponding parts in the French translation.

The screenshot shows the Google Translate interface with the source language set to FRENCH and the target language set to ENGLISH. The source text in French is: "Les générateurs de textes probabilistes sont utilisés pour produire de faux articles scientifiques depuis plus d'une décennie. De tels papiers absurdes sont facilement détectés par l'homme et la machine. Désormais, des techniques de génération plus complexes basées sur l'IA produisent des textes impossibles à distinguer de ceux des humains et la génération de textes scientifiques à partir de quelques mots-clés a été documentée. Notre étude introduit le concept de phrases torturées : des phrases étranges inattendues au lieu de". The translated text in English is: "Probabilistic text generators have been used to produce bogus scientific papers for over a decade. Such absurd papers are easily detected by man and machine. Now, more complex AI-based generation techniques produce texts indistinguishable from humans, and the generation of scientific texts from a few keywords has been documented. Our study introduces the concept of tortured sentences: unexpected strange sentences instead of established sentences, such as "counterfeit conscience" instead of "artificial". An orange arrow points from the underlined words in the English translation back to the corresponding parts in the French source text.

2. Paraphrasage

https://spinbot.com

Here are two paid options:

1 A paid subscription to use Spinbot.com through your browser.

Rewrite your text all day without the hassle of ads or a captcha.

- [1 Month for \\$10.00](#)
- [6 Months for \\$50.00 \(17% discount\)](#)
- [1 Year for \\$75.00 \(38% discount\)](#)

2 Purchase Spin Credits For Your Own App / Website (Developers only)

Pay for only the number of spins that you will need for your website or app.

- 1,000 Credits for \$5.00 USD
- 2,000 Credits for \$10.00 USD
- 5,000 Credits for \$20.00 USD (20% discount)
- 10,000 Credits for \$40.00 USD (20% discount)
- 20,000 Credits for \$80.00 USD (20% discount)
- 50,000 Credits for \$200.00 USD (20% discount)
- 100,000 Credits for \$400.00 USD (20% discount)
- 200,000 Credits for \$800.00 USD (20% discount)
- 500,000 Credits for \$2,000.00 USD (20% discount)

Purchase Spin Credits through the [developer account management page](#).

Spinbot

[Register](#) [Log in](#)

[Pricing](#) [About](#) [FAQs](#) [API](#) [Contact](#)

Article Spinning, Text Rewriting, Content Creation Tool.

Enter Text To Rewrite:

Probabilistic text generators have been used to produce fake scientific papers for more than a decade. Such nonsensical papers are easily detected by both human and machine. Now more complex AI-powered generation techniques produce texts indistinguishable from that of humans and the generation of scientific texts from a few keywords has been documented. Our study introduces the concept of tortured phrases: unexpected weird phrases in lieu of established ones, such as 'counterfeit consciousness' instead of 'artificial intelligence.' We combed the literature for tortured phrases and study one reputable journal where these concentrated en masse. Hypothesising the use of advanced language models we ran a detector on the abstracts of recent articles of this journal and on several control sets. The pairwise comparisons reveal a

← Ads by Google

Send feedback

Why this ad? ▾

Characters Remaining (Max 10,000): **8377**

Ignore Any:

Spin Capitalized Words:

**Remove Captcha and Ads ...
Supercharge Your Workflow Now!**

Captcha Solved! 🎉

Go!

Rewritten Text:

Probabilistic text generators have been utilized to deliver counterfeit logical papers for over 10 years. Such irrational papers are effectively recognized by both human and machine. Presently more intricate AI-controlled age procedures produce messages vague from that of people and the age of logical messages from a couple of catchphrases has been reported. Our review presents the idea of tormented expressions: surprising odd expressions in lieu of set up awareness' rather than 'man-made reasoning' for tormented expressions and study one these concentrated as a group. Conjecturing the u language models we ran a locator on the digests or late articles or this diary and on a few control sets. The pairwise examinations uncover a centralization of edited



3. Génération de texte

<https://transformer.huggingface.co/doc/arxiv-nlp>

Arxiv-NLP

Built on the OpenAI GPT-2 model, the Hugging Face team has fine-tuned the small version on a tiny dataset (60MB of text) of Arxiv papers. The targeted subject is Natural Language Processing, resulting in a very Linguistics/Deep Learning oriented generation.

Start writing

More info



Write With Transformer `gpt2` ⓘ

 Shuffle initial text
Cancel suggestion `esc`

 Trigger autocomplete or `tab`

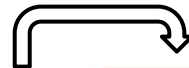
Select suggestion `↑` `↓` and `enter`

Save & Publish 

Probabilistic text generators have been used to produce fake scientific papers for more than a decade. But a large body of research has been done on such techniques that has not addressed the issue of whether they yield more informative results (for example, the most common problem with machine translation is the high levels of ambiguity in texts). In this paper , we introduce our framework to produce



OpenAI GPT-2 Detector



Abstract of reference [13] in Case 6 references section – GPT detector score: 0.02%

This work includes processing and classification of tweets which are written in Turkish language. **Four different sector** tweet datasets are vectorized with Word Embedding model and classified with Support Vector Machine and **Random Forests classifiers** and results have been compared. We have showed that sector based tweet classification is more successful **compared to general tweets**. Accuracy rates for Banking sector is 89.97%, for Football 84.02%, for Telecom 73.86%, for Retail 63.68% and for overall 74.60% have been achieved.

Questionable text in case 6, rewritten from [13] – GPT detector score: 59.20%

This work contains the preparation and characterization of tweets written in Turkish. Tweet information sets a vector of **four different offices**, contrasted and the outcomes and the installed model and grouping support vector machine and the **arbitrary timberland arrangement** in Word. Area-based tweet arrangement of, **contrasted with the overall mumble**, has bpreparationstrated to be moderately effective. The exactness rate for the financial area 89.97 percent, soccer 84.02 percent, 73.86 percent for correspondence, it has been made 74.60 percent of the absolute of the 63.68 percent for retail.

'Lazy Surveys' état de l'art du fainéant

Table 6 Elsevier journals from Control E with 25+ articles published in 2021 whose GPT detector score for abstracts is 70% or higher. The journal under investigation in this study, *Microprocessors and Microsystems*, has 75 articles with a 70% or higher GTP score for abstracts (Avg = 98.6). These 75 articles represent 72.1% of all articles published in this journal that are in Control E.

Journal	Article abstracts with GPT detector score $\geq 70\%$		Total articles in journal (%)
	Average GPT detector score (%)	Number of articles	
<i>J. Alloy. Compd.</i>	92.7	104	5.6
<i>Sci. Total Environ.</i>	93.0	81	3.3
<i>J. Clean Prod.</i>	92.5	76	4.4
<i>Microprocess. Microsyst.</i>	98.6	75	72.1
<i>J. Mol. Struct.</i>	93.3	73	9.5
<i>Chem. Eng. J.</i>	90.8	68	4.3
<i>Appl. Surf. Sci.</i>	93.8	64	5.3
<i>Mater. Lett.</i>	93.5	62	9.8
<i>Ceram. Int.</i>	91.5	61	4.9
<i>Sens. Actuator B-Chem.</i>	92.3	48	7.5
<i>Int. J. Hydrog. Energy</i>	91.4	47	5.0
<i>Chemosphere</i>	93.6	47	3.6
<i>J. Colloid Interface Sci.</i>	94.6	46	5.4
<i>J. Hazard. Mater.</i>	92.9	45	3.7
<i>J. Mol. Liq.</i>	91.2	45	6.7
<i>Biochem. Biophys. Res. Commun.</i>	93.4	42	7.9

72,1 % des résumés de *MicroPro* publiés en 2021 ont un score GPT de 98.6%





Tortured Phrases Detector

Matches with 2+ tortured phrases (e.g., credulous Bayes, counter...

Cabanac, G., Labbé, C., & Magazinov, A. (2022). "Bosom peril" is not "breast cancer": How weird computer-generated phrases help researchers find scientific publishing fraud. *Bulletin of the Atomic Scientists*. Online

Cabanac, G., Labbé, C., & Magazinov, A. (2021). Tortured phrases: A dubious writing style emerging in science. Evidence of critical issues affecting established journals. *arXiv preprint*. <https://arxiv.org/abs/2107.06751> (supplementary materials)
 ▶ Featured in *Nature* (Else, 2021)

Check potentially problematic papers and report on PubPeer: the last column will link to your PubPeer comment.

Report True/False Positives, ideas, or else using the Feedback feature.

Search: [] Go 1. Primary Report Rows 20 Actions

1 - 20 of 4259

Year	Doctype	Publisher	Venue	Title	Citations	Altmetric	Tips	Tortured Phrases	Assessor	PubPeer Users
2021	article	Institute of Advanced Engineering and Science	Indonesian Journal of Electrical Engineering and Computer Science	A comparative and comprehensive study of prediction of Parkinson's disease	0	0	20	Alzheimer's infection • Parkinson's infection • Parkinson's sickness • R2 esteem • arbitrary backwoods • backing vector machine • dimensionality decrease • fake neural • fluffy C implies • gullible Bayes • head part investigation • help vector machine • inertial estimation unit • invulnerable framew...	Cabanac	Guillaume Cabanac
2020	chapter	Springer Nature	Internet of Things Use Cases for the Healthcare Industry	AI and IoT in Healthcare	2	0	20	Alzheimer's ailment • Alzheimer's infection • Alzheimer's sickness • Parkinson's ailment • Parkinson's sickness • attractive reverberation • brilliant AND keen AND savvy AND shrewd • calculated relapse • creating nations • help vector machine • hereditary programming • information distribution cente...	Cabanac	N. H. Wise
2017	article	World Scientific Publishing	International Journal of Wavelets Multiresolution and Information Processing	Applications of wavelet transform on power engineering problems — A brief review	0	0	19	back proliferation • back spread AND neural • bolster vector machine • concealed layer • cross breed model • discrete Fourier change • fake neural • fast Fourier change • fluffy deduction • fluffy rationale • hereditary calculation • manufactured neural • molecule swarm • moving normal (ARIMA) • sel...	Cabanac	Guillaume Cabanac
2021	chapter	Wiley	Green Internet of Things and Machine Learning	Machine Learning-Enabled Techniques for Reducing Energy Consumption of IoT Devices	0	0	18	brilliant metering • brilliant AND keen AND savvy AND shrewd • counterfeit neural • field programmable entryway exhibit • irregular woodland • keen city • keen gadget • keen home • man-made brainpower • savvy metering • shrewd city • shrewd home • shrewd metering • sun oriented force • uphold vector...	Cabanac	Guillaume Cabanac

Year	Doctype	Publisher	Venue	Title	Citations	Altmetric	Tips	Tortured Phrases	Assessor
2021	proceeding	Institute of Electrical and Electronics Engineers (IEEE)	2021 International Conference on Advances in Computing, Communication, and Control (ICAC3)	HVAC Hybrid Control methods for HEE in Buildings: Overview	0	0	9	back spread AND neural • counterfeit neural • enhancement calculation AND optimiz* • fluffy rationale • hereditary calculation • profound neural organization • square blunder • warm solace • warmth move	Invitation for human assessors
2019	proceeding	Institute of Electrical and Electronics Engineers (IEEE)	2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)	Handling 3vs of Big Data Through Swarm Intelligence	1	0	7	computerized reasoning • enhancement calculation AND optimiz* • figuring asset • genuine esteemed • hereditary calculation • molecule swarm • wellness esteem	Invitation for human assessors



Microarray Gene Expression Analysis Using Fuzzy Logic (MGA-FL)

First International Conference on Sustainable Technologies for Computational Intelligence (2019) - 1 Comment
doi: 10.1007/978-981-13-1951-8_16 issn: 2194-5357 issn: 2194-5365

Daksh Khanna, Tanupriya Choudhury, A. Sai Sabitha, Nguyen Gia Nhu

#1 Elisabeth M Bik

This article costs € 27.20 / \$ 32.20 on the Springer website. I would not recommend to purchase it, however.

This paper appears to contain "Tortured Phrases": non-sensical, synthetic text generated by synonym converter tools, as described in this preprint by Cabanac, Labbé, and Magazinov, arXiv:2107.06751 - <https://arxiv.org/abs/2107.06751>. As explained in the preprint, text might have been 'synonymized' to mask plagiarism.

Parts of this paper appear to contain synonymized paragraphs from Wikipedia (https://en.wikipedia.org/wiki/DNA_microarray):

- Wikipedia: "Each DNA spot contains picomoles (10EXP-12 moles) of a specific DNA sequence, known as probes (or reporters or oligos). These can be a short section of a gene or other DNA element that are used to hybridize a cDNA or rRNA (also called anti-sense RNA) sample (called target) under high-stringency conditions. Probe-target hybridization is usually detected and quantified by detection of fluorophore-, silver-, or chemiluminescence-labeled targets to determine relative abundance of nucleic acid sequences in the target. The original nucleic acid arrays were macro arrays approximately 9 cm × 12 cm and the first computerized image based analysis was published in 1981."
- Khanna et al. (2019): "Each Deoxyribonucleic Acid spot consist of micro-moles (appx. 11 mol) of a particular Deoxyribonucleic Acid progression, justified as tests (or writers or oligos). It can be a small territory of a quality or other DNA segment which is used to hybridize a complementary DNA/complementary RNA (in like manner called antagonistic to identify Ribonucleic Acid) test (called centre) under high-stringency terms. Test object hybridization is normally recognized and acquired through distinguishing proof of "fluorophore-, silver-, or chemiluminescence"—named centres to choose relatively abundant nucleic destructive game plans for the goal. The principal nucleic destructive displays were expansive scale groups around 9 cm × 12 cm and the fun- damental modernized picture based examination was circulated in 1981."

Figure 1 looks identical to a Wikimedia file found here:

https://commons.wikimedia.org/wiki/File:Microarray_exp_horizontal.svg

Text from Section 3 appears to be a synonymized version of Fernando Díaz et al., Using Fuzzy Patterns for Gene Selection and Data Reduction on Microarray Data, IDEAL 2006, LNCS 4224, pp. 1087 – 1094, 2006, DOI: 10.1007/11875581_129. Here is an example:

AI Techniques for COVID-19

IEEE Access (2020) - 2 Comments
doi: 10.1109/access.2020.3007939 issn: 2169-3536

Adedoyin Ahmed Hussain, Ouns Bouachir, Fadi Al-Turjman, Moayad Aloqaily

#1 N. H. Wise

This paper already has 34 citations, I guess because Covid-19 and AI are both very hot topics, but appears to have been written using a thesaurus engine to avoid plagiarism detection.

The abstract alone contains the tortured phrases 'neural systems' for 'neural networks', and 'edge significant learning', which only gives 10 results on Google, most of which are for this paper.

The first paragraph of the introduction is in unusual language and contains telling incorrect definitions of abbreviations, such as 'mad-made brainpower (AI)', and ends with 'Meanwhile, Insilico Medicine situated in Hong Kong as of late declared that it's AI calculations had planned six new atoms that could end viral replication'.

The rest of the introduction is in a similar style and contains the tortured phrases 'twofold tree' (for 'binary tree' for 'dual tree?') and 'convolutional neural system (CNN)'.

Many other sections of the paper contain tortured phrases and nonsensical sentences, suggesting much of the paper has been plagiarised.

report permalink Reply

#2 Salsola Zygothylla



This paper contains a prime example of thesaurus gone wild: "extreme intense respiratory disorder (SARS)". Given that those who use the PubPeer browser plugin are going to be directed to this page whenever this paper turns up in a search or on a web page, it seems both fair and useful to ask the authors: What contribution to knowledge do you think you have made with this publication? The same question could fairly and usefully be asked of those who cited it.

Polymers, responsiveness and cancer therapy

Artificial Cells Nanomedicine and Biotechnology (2019) - 1 Comment
doi: 10.1080/21691401.2018.1559176 issn: 2169-1401 issn: 2169-141x pubmed: 30688110

Anil M. Pethe, Khushwant S. Yadav

#1 N. H. Wise

This paper contains tortured phrases suggesting the use of a thesaurus engine to avoid plagiarism detection. Specifically, it refers to 'corrosive' instead of 'acid', leading to:

- 'nucleic corrosive' - nucleic acid
- 'acrylic corrosive' - acrylic acid
- 'phosphoric corrosive' - phosphoric acid



Low cost Covid-19 preliminary diagnosis utilizing cough samples and keenly intellective deep learning approaches

Alexandria Engineering Journal (2021) - 1 Comment
doi: 10.1016/j.aej.2020.09.032 issn: 1110-0168

D. Sudaroli Vijayakumar, Monica Sneha

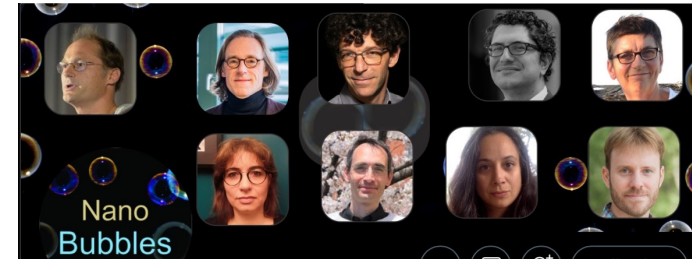
#1 N. H. Wise

This paper contains phrases that suggest the use of a thesaurus engine to avoid plagiarism detection. Examples include:

- 'Nucleic Corrosive Enhancement Tests (NAAT)' - Nucleic acid amplification tests
- 'real-time Turn around Translation Polymerase Chain Response (rRT-PCR)' - real-time reverse transcription polymerase chain reaction
- 'flag to clamor' - signal to noise
- 'sifted flag' - filtered signal



Despite the paper being on 'cough', 'cough' only appears in the title, abstract and **once** in the body of the text. Instead, 'hack' appears over 40 times, including in section titles.



<https://nanobubbles.hypotheses.org>

Cloud-Based Breast Cancer Prediction Empowered with Soft Computing Approaches

Journal of Healthcare Engineering (2020) - 1 Comment
doi: 10.1155/2020/8017496 issn: 2040-2295 issn: 2040-2309 pubmed: 32509260

Farrukh Khan, Muhammad Adnan Khan, Sagheer Abbas, Atifa Athar, Shahan Yamin Siddiqui, Abdul Hannan Khan, Muhammad Anwaar Saeed, Muhammad Hussain

#1 Diradops Cardena

This article shows clear evidence of 'tortured phrases'. To give a few examples:



- **bosom malignancy** (breast cancer)
- **Prepreparing** (preprocessing)
- **crusty logic** (crisp logic, a term anyone who is writing a paper on fuzzy logic could be expected to know!)
- **imitation** (simulation)

Considering that this paper deals with the clinical applications of ML for a potentially fatal disease, it's rather inappropriate and dangerous that it slipped past controls.

There's also a lot of extremely trivial maths, which appears to have been included for no other purpose than to 'bulk up' the paper (see eqns. 5-10, which might be appropriate in a high school textbook but not an academic paper).

Expressions torturées = indices

D'autres erreurs à identifier!

⚠ Expertise métier nécessaire



Guillaume Cabanac @gcabanac · 1 Feb

The 'Problematic Paper Screener' flags 5,349 articles as problematic with 7 detectors. They are cited 37k times! Only 1,351 reported on @PubPeer.

Let's depollute the scientific literature: 3,021 tortured papers to re-assess

irit.fr/~Guillaume.Cab... twitter.com/gcabanac/statu...

3/10 000 articles publiés en 2021 :

- 70 % d'Inde

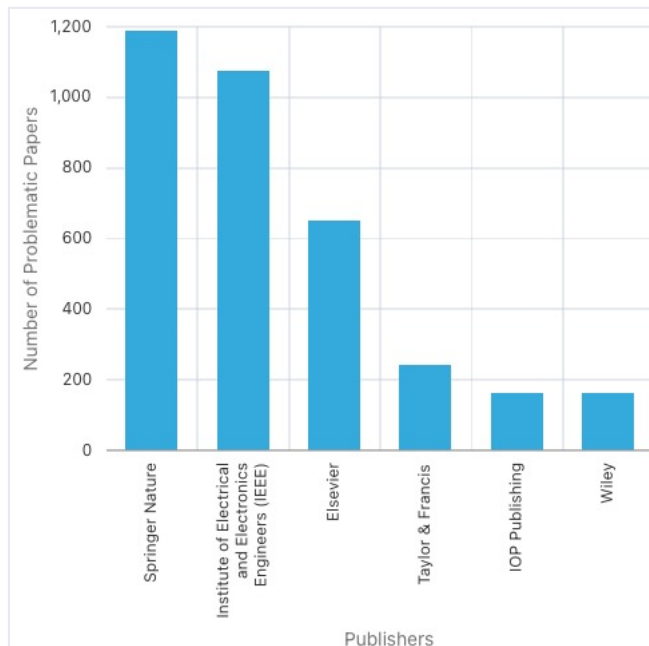
- 6 % de Chine

Citation count

15,032

1 - 20 of 4225

Tortured Phrases	Assessor	PubPeer Users
Alzheimer's infection • Parkinson's infection • Parkinson's sickness • R2 esteem • arbitrary backwoods • backing vector machine • dimensionality decrease • fake neural • fluffy C implies • gullible Bayes • head part investigation • help vector machine • inertial estimation unit • invulnerable framew...	Cabanac	Guillaume Cabanac
Alzheimer's ailment • Alzheimer's infection • Alzheimer's sickness • Parkinson's ailment • Parkinson's sickness • attractive reverberation • brilliant AND keen AND savvy AND shrewd • calculated relapse • creating nations • help vector machine • hereditary programming • information distribution cente...	Cabanac	N. H. Wise



1 - 20 of 3171

Tortured Phrases	Assessor	PubPeer Users
brilliant AND keen AND savvy AND shrewd • computerized reasoning • counterfeit neural • figuring asset • k-implies grouping • man-made consciousness • vitality effective • vitality utilization	Invitation for human assessors	-
brilliant AND keen AND savvy AND shrewd • directing convention • insatiable calculation • keen gadget • malignant hub • organization assault • pernicious hub • steering convention	Invitation for human assessors	-
computerized reasoning • enhancement calculation AND optimiz* • figuring asset • genuine esteemed • hereditary calculation • molecule swarm • wellness esteem	Invitation for human assessors	-

NEWS | 27 May 2021 | Clarification [28 May 2021](#)

Hundreds of gibberish papers still lurk in the scientific literature

The nonsensical computer-generated articles, spotted years after the problem was first seen, could lead to a wave of retractions.

[Richard Van Noorden](#)

NEWS | 05 August 2021 | Correction [10 August 2021](#)

'Tortured phrases' give away fabricated research papers

Analysis reveals that strange turns of phrase may indicate foul play in science.

[Holly Else](#)

NEWS | 08 November 2021

Scammers impersonate guest editors to get sham papers published

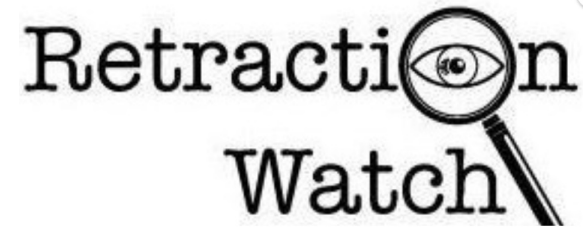
Hundreds of junk-science papers have been retracted from reputable journals after fraudsters used 'special issues' to manipulate the publication process. And the problem is growing.

[Holly Else](#)

"It is very worrying," says Guillaume Cabanac, a computer scientist at the University of Toulouse in France, who has worked to uncover nonsense science papers in special issues. He adds that it is shocking to see such papers in journals from 'flagship' publishers and that "it is not only predatory journals that publish bullshit".

July 19, 2021 Adam Marcus

‘Tortured phrases’, lost in translation: Sleuths find even more problems at journal that just flagged 400 papers



August 26, 2021 Ivan Oransky

Guest editor says journal will retract dozens of inappropriate papers after his email was hacked

September 22, 2021 Adam Marcus

“[T]hese shit comments”: Author of a nonsense paper responds on PubPeer

September 28, 2021 Ivan Oransky

Springer Nature slaps more than 400 papers with expressions of concern all at once

November 4, 2021 Adam Marcus

Springer Nature geosciences journal retracts 44 articles filled with gibberish

NATURE'S 10

Ten people who helped shape science in 2021.

Winnie Byanyima / Friederike Otto / Zhang Rongqiao /
Timnit Gebru / Tulio de Oliveira / John Jumper /
Victoria Tauli-Corpuz / Guillaume Cabanac /
Meaghan Kall / Janet Woodcock

The *Nature's 10* list explores key developments in science this year and some of the people who played important parts in these milestones. Along with their colleagues, these individuals helped to make amazing discoveries and brought attention to crucial issues. *Nature's 10* is not an award or a ranking. The selection is compiled by *Nature's* editors to highlight key events in science through the compelling stories of those involved.

The *Nature's 10* list explores **key developments in science** this year and some of the people who played important parts in these milestones. Along with their colleagues, these individuals helped to make **amazing discoveries** and brought attention to **crucial issues**. *Nature's 10* is not an award or a ranking. The selection is compiled by *Nature's* editors to highlight key events in science through the compelling stories of those involved.

Feature *Nature's 10*

Guillaume Cabanac Deception sleuth

This computer scientist helped to uncover a new kind of fabricated paper.

By Diana Kwon

Underground creepy crawlly state. Bosom malignancy. Sun oriented force. These might sound like expressions from a work of fiction, but they are actually strange translations, pulled from the scholarly literature, of scientific terms – ant colony, breast cancer and solar energy, respectively. Guillaume Cabanac, a computer scientist at the University of Toulouse, France, spots such bizarre phrases in academic papers every day.



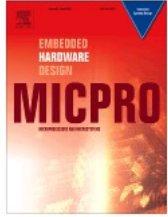
ILLUSTRATION BY ELENA CALOYAROU-BANSE



MODUS OPERANDI

1. Génération algorithmique : SCIGen
2. Plagiat et expressions torturées
3. *Peer Review hacking*
4. *Paper mills* alias moulins à papier
5. *Citation hacking*

Signaux faibles de manipulation



Impact Factor
1.525

Table 2 Top 5 contributing countries and organizations of *Microprocessors and Microsystems* over 2017–2019 as per the *Journal Citation Reports* ($N = 378$ articles).

Countries		Organisations	
Name	Articles	Name	Articles
India	55	CNRS, France	22
China (mainland)	43	Czech Technical University	9
France	38	University of Montenegro	9
Germany	32	Technical University of Munich	8
USA	30	Universidade Federal do Rio Grande do Sul	8
Iran	28	Indian Institute of Technology System	7

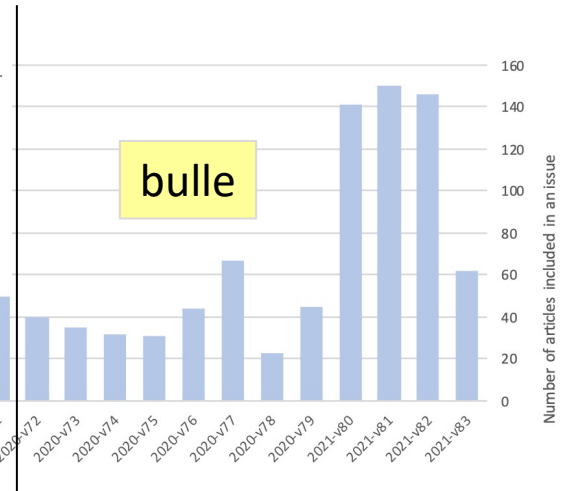


Fig. 2 Number of articles included in the volumes 56–83 of *Microprocessors and Microsystems*.

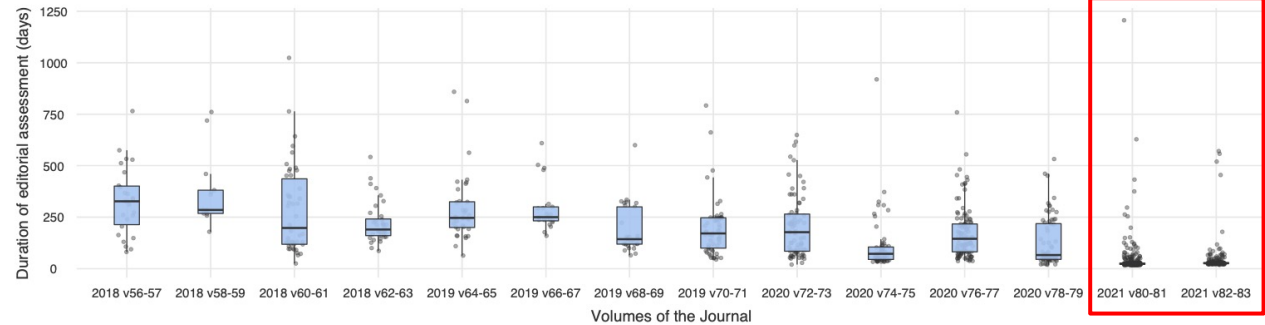
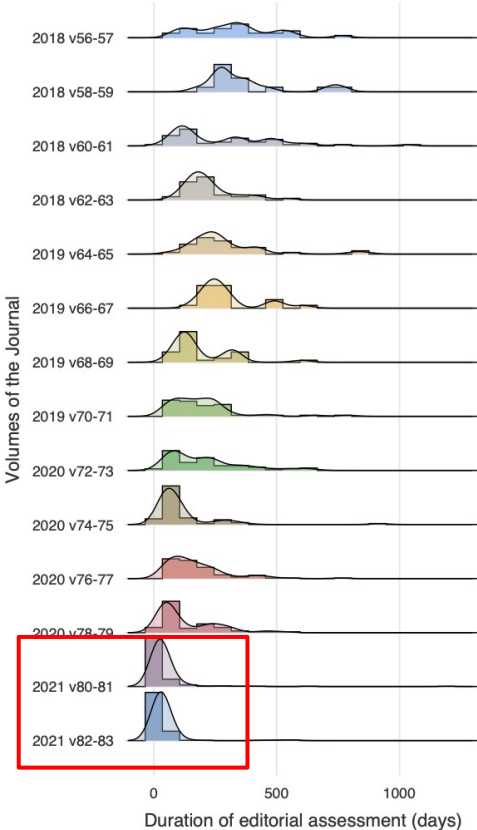


Fig. 3 Editorial assessment at *Microprocessors and Microsystems*: duration in days elapsed from submission to acceptance of the 1,078 articles published in volumes 56–83 issued between February 2018 and June 2021. The same data are presented with three complementary visualisations. The volumes of early 2021 (v80–83, $N = 499$, $Med = 25$ days) compared to the volumes of early 2020 (v74–77, $N = 174$, $Med = 108$), see Tab. 3.

Retraction Watch

July 19, 2021 Adam Marcus

Tracking retractions as a window into the scientific process

‘Tortured phrases’, lost in translation: Sleuths find even more problems at journal that just flagged 400 papers

What do subterranean insect provinces and motion to clamor have to do with microprocessors and microsystems?

That’s an excellent question. Read on, dear reader.

Earlier this month, as we reported, Elsevier announced that it had concerns about some 400 papers published in special issues in one of its journals. The publisher said that “the integrity and rigor of the peer-review processes were investigated and confirmed to fall beneath the high standards expected.”



Guillaume Cabanac

Update, 1140 UTC, 7/21/21: An Elsevier spokesperson gave us this statement:

The Editor in Chief of Microprocessors & Microsystems has serious concerns about the integrity and rigor of the peer-review processes that led to the publication of over 400 papers in six Special Issues, as previously reported. Investigation and due process regarding individual articles published in these Special Issues is ongoing. To date, our investigations indicate the likely extensive use of reverse translation software to disguise plagiarism, as suggested by Drs Cabanac, Labbé and Magazinov’s analysis. We appreciate the authors helping bring this issue to light. We will provide further updates in accordance with Elsevier and the Committee on Publication Ethics best practices once the investigations have concluded.

The official journal of the Saudi Society for Geosciences, the Arabian Journal of Geosciences examines the entire range of earth science topics focused on, but not limited, to those that have regional significance to the Middle East and the Euro-Mediterranean Zone.

1.827 (2020)
Impact factor

Retraction Watch

Tracking retractions as a window into the scientific process

September 28, 2021 | Ivan Oransky

Springer Nature slaps more than 400 papers with expressions of concern all at once



Cartoon by Hilda Bastian (license)

A total of 436 papers in two Springer Nature journals are being subjected to expressions of concern, in the latest case of special issues — in this case, “topical collections” — likely being exploited by rogue editors or impersonators.

<https://retractionwatch.com/2021/09/28/springer-nature-slaps-more-than-400-papers-with-expressions-of-concern-all-at-once/>

Use Case: the Arabian Journal of Geosciences

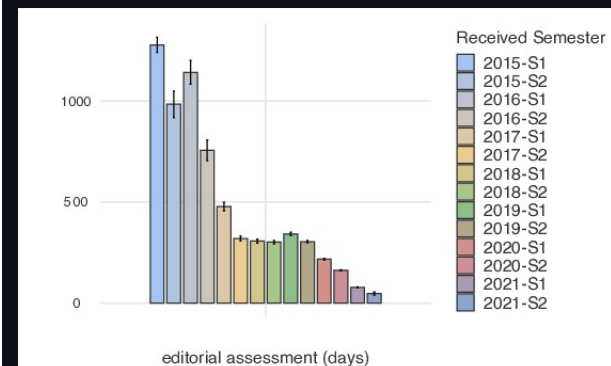
▷ August 23, 2021: Eracon Lachesis and Alexander Magazinov reported on PubPeer issues with a publication from Arabian Journal of Geosciences. See this annotated example:

The screenshot shows a SpringerLink article page. The title is 'Research on coastal atmospheric change and Latin dance performance based on target detection network'. The abstract discusses environmental degradation in China and the management of volatile organic pollutants. The introduction discusses Latin dance as an art form and the article's focus on its performance. A yellow box highlights the title, and a red box highlights the abstract and introduction text. A yellow callout box on the left says 'Without transition, just concatenated!'.

We observed an impressive and questionable shrinking of the duration of editorial assessments:

- 282 days in 2019-S2
- 159 days in 2020-S2
- 47 days in 2021-S2

▷ September 28, 2021: Springer published Expressions of concern for 436 publications in the Arabian Journal of Geosciences.



<https://github.com/gcabanac/editorial-assessment>



MODUS OPERANDI

1. Génération algorithmique : SCIGen
2. Plagiat et expressions torturées
3. *Peer Review hacking*
4. *Paper mills* alias moulins à papier
5. *Citation hacking*

—MARA HVISTENDAHL

www.sciencemag.org SCIENCE VOL 342 29 NOVEMBER 2013



China's Publication Bazaar

A Science investigation has uncovered a smorgasbord of questionable practices including paying for author's slots on papers written by other scientists and buying papers from online brokers

"IT'S UNBELIEVABLE: YOU CAN PUBLISH SCI PAPERS WITHOUT DOING EXPERIMENTS."

—Banner on Sciedit's website

Description de la manipulation (février 2022) :

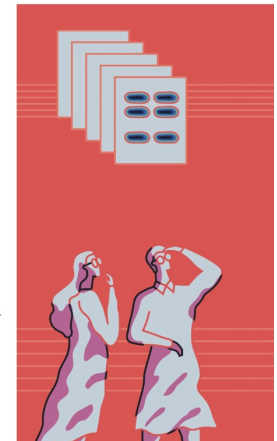
<https://pubpeer.com/publications/940C291607CF03969C6A936F8BA5B9#2>

Feature

516 | Nature | Vol 591 | 25 March 2021

THE BATTLE AGAINST PAPER MILLS

Some journals have admitted to a problem with fake research papers. Now editors are trying to combat it. By Holly Else and Richard Van Noorden



Industrialized cheating

The problem of organized fraud in publishing is not new, and not confined to China, notes Catriona Fennell, who heads publishing services at the world's largest scientific publisher, Elsevier. "We've seen evidence of industrialized cheating from several other countries, including Iran and Russia," she told *Nature* last year. Others have also reported on Iranian and Russian paper-mill activities.

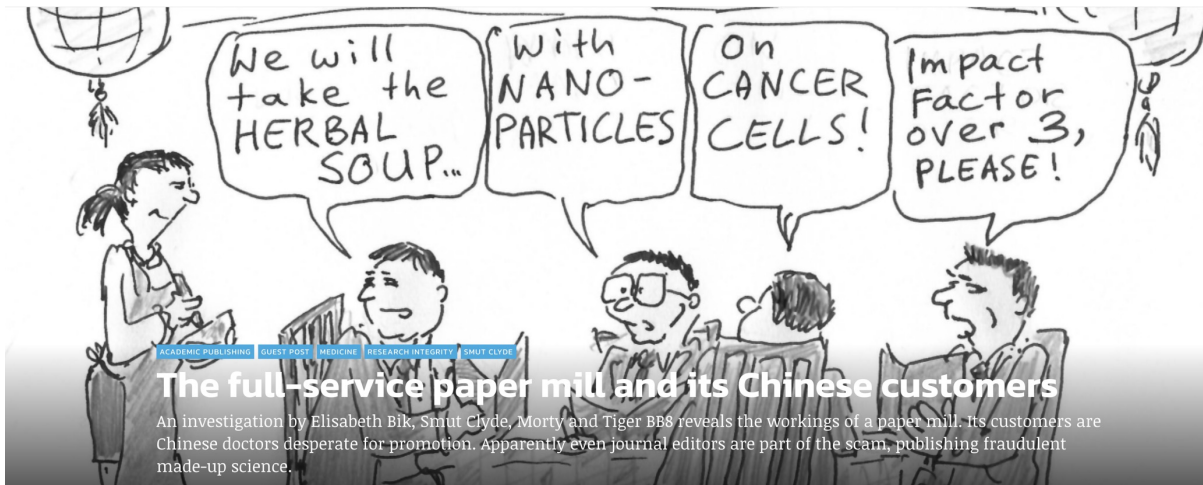
In a statement this year to *Nature*, Elsevier said that its journal editors detect and prevent the publication of thousands of probable paper-mill submissions each year, although some do get through.

China has long been known to have a problem with firms selling papers to researchers, says Xiaotian Chen, a librarian at Bradley University in Peoria, Illinois. As far back as 2010, a team led by Shen Yang, a management-studies researcher then at Wuhan University in China, warned of websites offering to ghostwrite papers on fictional research, or to bypass peer-review systems for payment. In 2013, *Science* reported on a market for authorships on research papers in China. In 2017, China's Ministry of Science and Technology (MOST) said it would crack down on misconduct after a scandal in which 107 papers were retracted at the journal *Tumor Biology*; their peer reviews had been fabricated and a MOST investigation concluded that some had been produced by third-party companies.

Physicians in China are a particular target market because they typically need to publish research articles to gain promotions, but are so busy at hospitals that they might not have time to do the science, says Chen. Last August, the Beijing municipal health authority published a policy stipulating that an attending physician wanting to be promoted to deputy chief physician must have at least two first-author papers published in professional journals; three first-author papers are required to become a chief physician. These titles affect

For Better Science

BY LEONID SCHNEIDER, ON RESEARCH INTEGRITY, BIOMEDICAL ETHICS AND ACADEMIC PUBLISHING



BY LEONID SCHNEIDER
JANUARY 24, 2020

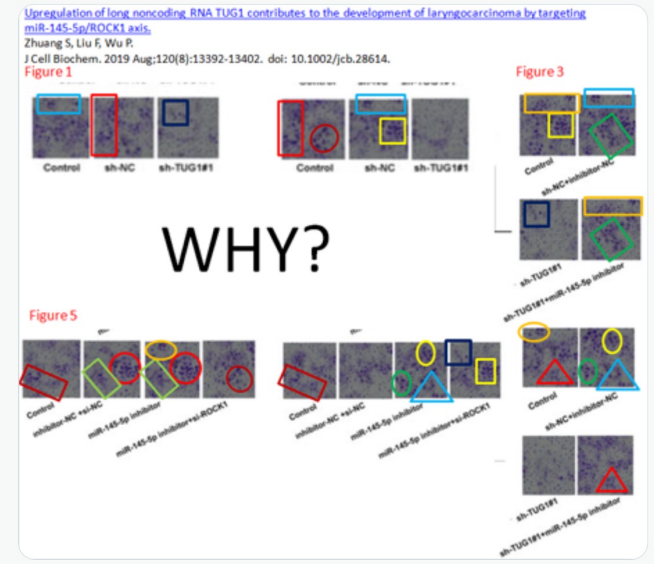
In China, clinicians are expected to publish a certain number of research papers in international journals if they want to be promoted. The easiest way is to pay a paper mill, which seem to provide a full service: an English-speaking research paper containing Photoshop-generated fake research data, in a respectable peer-reviewed journal, with your name on it. Entire journals published by Wiley or Elsevier succumb to such scams, presumably because certain corrupt editorial board members are part of it. This was uncovered in an investigation by Elisabeth Bik, as well as the pseudonymous Smut Clyde, Morty and Tiger BB8, and narrated below by Smut Clyde.

The presently almost 500 papers were traced to one specific paper mill not because of direct image reuse, but because the data showed same patterns of falsifications: same blot backgrounds, same shapes of bands, and similarly falsified flow cytometry plots.



Morty @mortenoxe · Jan 20, 2020

Published by Wiley, indexed by the US National Library of Science, and read by scientists. Paid by public funds, all profit to the publisher and some laughter in China. #BetterJournalIndexing #PubPeer #NIH #Wileyinresearch





MODUS OPERANDI

1. Génération algorithmique : SCIGen
2. Plagiat et expressions torturées
3. *Peer Review hacking*
4. *Paper mills* alias moulins à papier
5. *Citation hacking*

#4 Schizomeris leibleinii commented November 2021

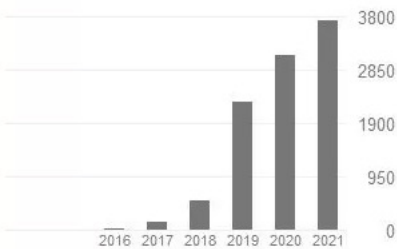


187 results

Mohamed **Elhoseny** has an impressive Output, Citation Count, H-Index and Publons score. According to his website (<http://www.mohamedelhoseny.com/academic-qualifications>) he finished his **PhD in 2015**. Since then he acquired an impressive H-index of more than 50 according to google scholar:

Cited by

	All	Since 2016
Citations	10040	9989
h-index	59	58
i10-index	174	174



#8 Marianne Alunno-Bruscia commented November 2021

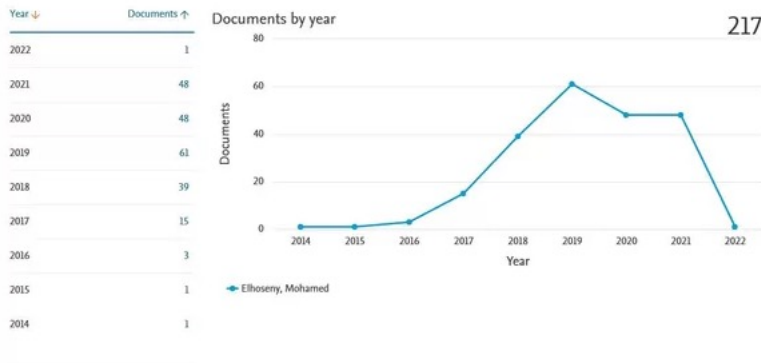
Re #4. M. Elhoseny is part of the 2021 annual list of **Highly Cited Researchers™** from Clarivate™ “recognizing influential researchers in the sciences and social sciences from around the world”.

<https://web.archive.org/web/20211116210755/https://publons.com/researcher/1196537/mohamed-elhoseny/>

As stated by Clarivate™:

- “All Highly Cited Researcher records are reviewed. Factors such as retractions, misconduct, and extreme self-citation—all of which would detract from true community-wide research influence—may lead to an author being excluded from the list.”
- “There is no unique or universally agreed concept of what constitutes exceptional research performance and elite status in the sciences and social sciences. Consequently, no quantitative indicators will point to a list that satisfies all expectations or requirements.”

While producing between 40 and 60 papers a year according to Scopus, he also managed to review up to almost one paper per day for several journals in 2018.



#1 Guillaume Cabanac commented February 2022

This article **cited 32 times** contains several **tortured phrases** that make some passages hard to parse. These typically result from an attempt to avoid plagiarism detection using a paraphrasing software.

need to know the channel state data at the transmitter. The most testing features in VANETs are their dynamic topology and **versatility**, where vehicles are moving at variable high speeds and in various directions. Conversely, the test in the WSN is in dealing with the constrained energy assets of the nodes, since the execution of WSNs emphatically relies upon their lifetime by Mohaisen et al. [11]. **To conquer these difficulties, this exploration researches** the impacts of various Quality of Service (QoS)

framework of the PSO algorithm. The progression vector demonstrates the course of the development of the dragonflies and characterized as **pursues**:

$$\Delta V_{i+1} = (sSe_i + aAl_i + cCo_i + fFood_i + eEnemy_i) + w\Delta V_i \quad (4)$$

Energy Consumption: It is referred to as total energy **depleted** during data transmission of packets from source to destination.

Table 2 **clarifies** the outcome examination of a proposed model for an alternate number of vehicle nodes. In light of the node quality (efficient course), vehicles are

Can the authors explain why they departed from the established phrases?

How come these incorrect wordings survived proofreading by the coauthors, editors, referees, copy editors, and typesetters?

https://doi.org/10.53053/MGSA7457

Why was a study about autism cited by a paper on plant beauty?

BY DALMEET SINGH CHAWLA / 21 JANUARY 2022

Citation inflation: Autism studies, among others, are appearing in the reference lists of unrelated papers, possibly in an attempt to artificially boost authors' metrics.

#1 *Rhipidura albiventris* commented February 2022

A typical citation pattern of GW Wei, the first author of this paper.

1. Introduction

The ordered weighted aggregation (OWA) operator (Yager, 1988) as an aggregation technique has received more and more attention since its appearance (Merigó & Casanovas, 2009; Merigó & Gil-Lafuente, 2009; Merigó, 2010; Wei, 2008a, 2008b, 2008c, 2009a, 2009b, 2010a, 2010b, 2010c, 2010d; Wei, Zhao, & Lin, 2010; Wei, 2011a, 2011b, 2011c; Wei, Wang, Lin, & Zhao, 2011a, 2011b, Wei, Wang, & Lin, 2011c; Wei, 2011d, 2011e, 2011f, 2011g; Wei & Zhao, 2012; Zhang & Chu, 2009; Xu & Da, 2003). One important step of the OWA operator is to determine its associated weights. Many authors

dures to obt. from a colle second one and Da (200 for obtaining since its appearance new depend ment-depend arguments t sults. Xu (2 weighted ag

3439–3447.
 Wei, G. W. (2008a). Maximizing deviati making in intuitionistic fuzzy setting
 Wei, G. W. (2008b). Interval-valued intt Journal of Information and Computati
 Wei, G. W. (2008c). Induced intuition operator and its application to m Lecture Notes in Artificial Intelligence.
 Wei, G. W. (2009a). Uncertain linguist application to group decision makir International Journal of Uncertainty 251–267.
 Wei, G. W. (2009b). Some geometric agg dynamic multiple attribute decisic International Journal of Uncertainty 17(2), 179–196.
 Wei, G. W. (2010a). GRA method for incomplete weight information in ir Systems, 23(3), 243–247.

Global influence

We offer our authors a highly respected home for their research. *Scientific Reports* is the 6th most-cited journal in the world, with more than 540,000 citations in 2020*, and receives widespread attention in policy documents and the media.

*2020 Journal Citation Reports® Science Edition (Clarivate Analytics, 2021)

Expert peer review

Partnering with our extensive network of exper

scientific reports

OPEN Day-ahead electricity price forecasting using WPT, VMI, LSSVM-based self adaptive fuzzy kernel and modified HBMO algorithm

Rahmad Syah^{1,2}, Mohammad Rezaei², Marischa Elveny^{3,5}, Meysam Majidi Nezhad⁴, Dadan Ramdan¹, Mehdi Neshat^{6,7} & Afshin Davaranah⁷

¹Data Science & Computational Intelligence Research Group, Universitas Medan Area, Medan, Indonesia. ²Computer Science and Engineering Department, University of Texas at Arlington, Arlington, USA. ³Data Science & Computational Intelligence Research Group, Universitas Sumatera Utara, Medan, Indonesia. ⁴Department of Astronautics, Electrical and Energy Engineering (DIAEE), Sapienza University of Rome, 00184 Rome, Italy. ⁵DS & CI Research Group, Universitas Medan Area, Medan, Indonesia. ⁶Optimisation and Logistics Group, School of Computer Science, University of Adelaide, Adelaide, Australia. ⁷Department of Mathematics, Aberystwyth University, Aberystwyth SY23 3FL, UK. [✉]email: rahmadsyah@staff.uma.ac.id; marischaelveny@usu.ac.id; mehdi.neshat@adelaide.edu.au

This article was retracted on 10 February 2022

17375 | https://doi.org/10.1038/s41598-021-96501-6

nature portfolio

Important of price forecasting. Over the recent years, modern technologies^{1–5} shows more potential as view of electricity industry relaxation^{6–10} which leads to clear market without an extra force^{11–26}, therefore, the number of market participants is increased since they freely accessed to market information^{1–9}. Owing to the participants growing and the intense competition forces, the first, fast and correct decisions will be necessary for both enterprises and academia to maximize profit which itself needed to more accurate information of electricity price^{27–41}. Owing to the uneconomical way for energy storage^{42–47} in large-scale and its decency to different factors such as holidays^{48–52}, sudden disturbance in transmission and generation power systems, celebrations, season, etc.^{53–55}. The electricity price forecasting will be more difficult than any other financial markets^{56–63}. In other words, electricity markets price are the result of the intersection of supply and demand curves^{64–70}. The supply demand directly affects by weather conditions and general economy behaviors. In supply side, prices of fuel i.e. gas, coal, oil, etc. and unexpected fails also plays an important role^{71–76}. As result, Fig. 1 shows a graphical view of important factors which directly affects on power prices. It is clear that these factors make fluctuations in electricity price signals.

Received: 21 June 2021; Accepted: 2 August 2021
 Published online: 30 August 2021

References

1. Zhang, D., Liang, Z., Yang, G., Li, Q., Li, L. & Sun, X. [A robust forgery detection algorithm for object removal by exemplar-based image inpainting] *Multi. Tools Appl.* 77(10), 11823–11832 (2018).
 2. Song, Y., Zeng, Y., Li, X., Cai, B. & Yang, G. [Fast CU size decision and mode decision algorithm for intra prediction in HEVC] *Multi. Tools Appl.* 76(2), 2001–2017 (2017).
 3. Gu, K., Wu, N., Yin, B. & Jia, W. [Secure data query framework for cloud and fog computing] *IEEE Trans. Net. Service Manag.* 17(1), 332–345 (2019).
 4. Wei, W., Yongbin, J., Yanhong, L., Ji, L., Xin, W. & Tong, Z. [An advanced deep residual dense network (DRDN) approach for image super-resolution] *Int. J. Comput. Intell. Syst.* 12(2), 1592–1601 (2019).
 5. Li, W., Xu, H., Li, H., Yang, Y., Sharma, P.K., Wang, J. & Singh, S. [Complexity and algorithms for superposed data uploading problem in networks with smart devices] *IEEE Int. Things J.* 7(7), 5882–5891 (2019).

References [1-5] are irrelevant to the subject and the short sentence in the above text!

References [6-10] are irrelevant to the subject and the short sentence in the above text!

References [11-14, 16-17, 19-26] are irrelevant to the subject and the short sentence in the above text!

•••••

Il faut dépolluer la littérature scientifique

Malfaiteurs : faussaires pseudo-auteurs

Mobile : promotion, reconnaissance, capital symbolique

Modus operandi : plagats, infiltration de revues, achats de citations

Conséquences : pollution de la littérature

érosion de la confiance que la société
place dans la science

Détectives : scientifiques professionnels et amateurs
bénévoles et volontaires

Objectif : curatif → dépolluer la littérature
préventif → anticiper les nouvelles fraudes

Moyen : Ré-évaluation décentralisée soutenue par le
Problematic Paper Screener



<https://fr.wikipedia.org/wiki/Dépollution>

Des bénévoles nettoient les côtes après la catastrophe du Prestige (2002).

- Diapo 2 : <https://www.facebook.com/hyperconsciousness/photos/if-i-have-seen-further-it-is-by-standing-on-the-shoulders-of-giants-isaac-newton/3650241855102416/> <https://www.canstockphoto.com/texture-of-old-brick-wall-with-broken-50415102.html>
- Diapo 3 : <https://www.istockphoto.com/fr/photo/trou-dans-le-mur-gm175183459-22041615>
- Diapo 4 : <https://sciencenorthwest.com/so-how-do-scientists-ensure-honest-and-truthful-results/>
https://twitter.com/h_sjastad/status/1207323854873473025
- Diapo 22 : <https://www.20minutes.fr/monde/2643491-20191104-maree-noire-bresil-pire-venir-avertit-president-jair-bolsonaro>

Posted July 31, 2021.

Human gene function publications that describe wrongly identified nucleotide sequence reagents are unacceptably frequent within the genetics literature

Yasunori Park, Rachael A West, Pranuwan Pathmendra, Bertrand Favier, Thomas Stoeger, Amanda Capes-Davis, Guillaume Cabanac, Cyril Labbé, Jennifer A Byrne

doi: <https://doi.org/10.1101/2021.07.29.453321>

This article is a preprint and has not been certified by peer review

Publication : <https://doi.org/10.26508/lsa.202101203>

	Claimed status	Verified status	Consequences
Claimed targeting yet wrong target	5'- gene X -3'	5'- gene Y -3'	Irrelevant and/or impossible data
Claimed non-targeting yet targeting	5'- no target -3'	5'- gene X -3'	Use of invalid negative control
Claimed targeting yet non-targeting	5'- gene Y -3'	5'- no target -3'	Impossible data

nature

NEWS | 04 August 2021

Errors in genetic sequences mar hundreds of studies

Misreported nucleotide reagents may be rife in papers on human gene function.

Holly Else



A computer-aided analysis of almost 12,000 human-genetics papers has found more than 700 studies with errors in the DNA or RNA sequences of their experimental reagents¹. That amounts to a “problem of alarming proportions”, because it suggests that a worrying fraction of studies on human genes are not reliable, says the team that conducted the analysis, led by cancer researcher Jennifer Byrne at the University of Sydney in Australia. The mistakes could be accidental but, researchers suspect, might sometimes point to fraud.