Decontamination of the scientific literature with the Problematic Paper Screener: Flagging suspect/erroneous/fraudulent papers to crowdsource post-publication reassessments

Guillaume Cabanac

Guillaume Cabanac

October 13th, 2023

Illinois University

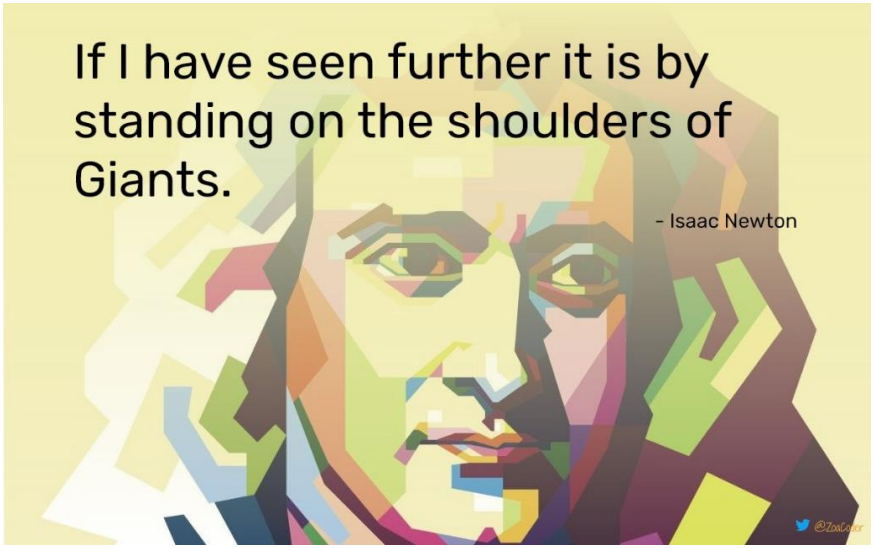
School of Information Sciences

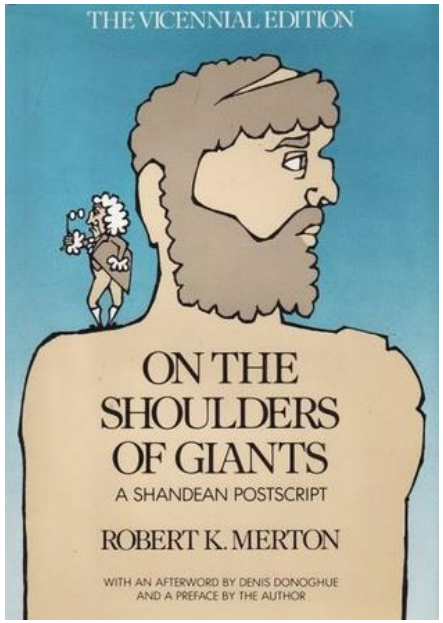
Decontamination of the scientific literature with the *Problematic Paper Screener*: Flagging suspect/erroneous/fraudulent papers to crowdsource post-publication reassessments

The NanoBubbles project has received Synergy grant funding from the European Research Council (ERC), within the European Union's Horizon 2020 programme, grant agreement no. 951393.

In collaboration with **Prof. Cyril Labbé**, Univ. Grenoble-Alpes and **Dr. Alexander Magazinov**, Yandex Kazakhstan

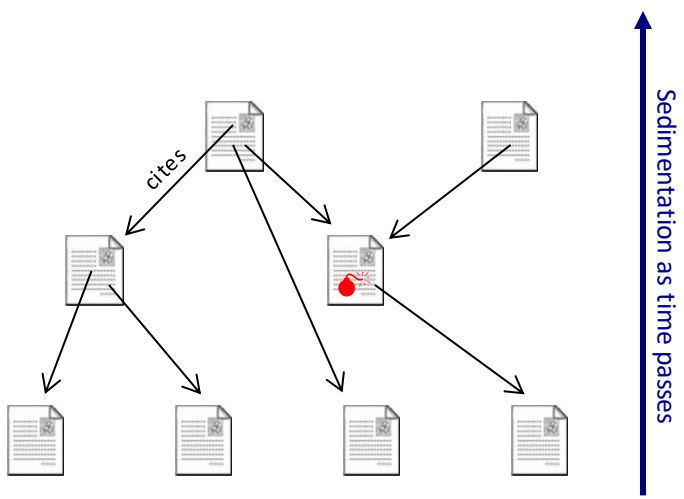
Science: knowledge accumulation





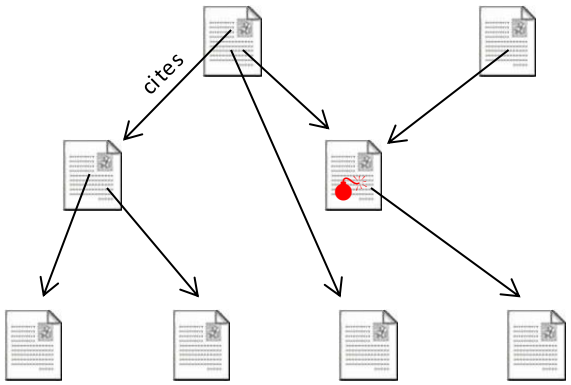
1965

Science contributes to the knowledge wall



Self-correction process

One flawed paper = weakened edifice



Sedimentation as time passes

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

Growth in a Time of Debt

By Carmen M. Reinhart and Kenneth S. Rogoff*

⟹ austerity measures

May 5, 2013

Errata:

"Growth in A Time of Debt"

Carmen M. Reinhart and Kenneth S. Rogoff

Harvard University

corrigenda, errata, and retractions

https://www.youtube.com/watch?v=yeX_Zs7zztY

The ethos of Science

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

une revue :

Peer review

Four Norms of Science

Robert K. Merton (7/4/1910 – 2/23/03) American Sociologist

Awarded the National Medal 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 (Yow!)

→ 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 rapports 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

Massification and globalization

UNESCO SCIENCE REPORT *Towards 2030*

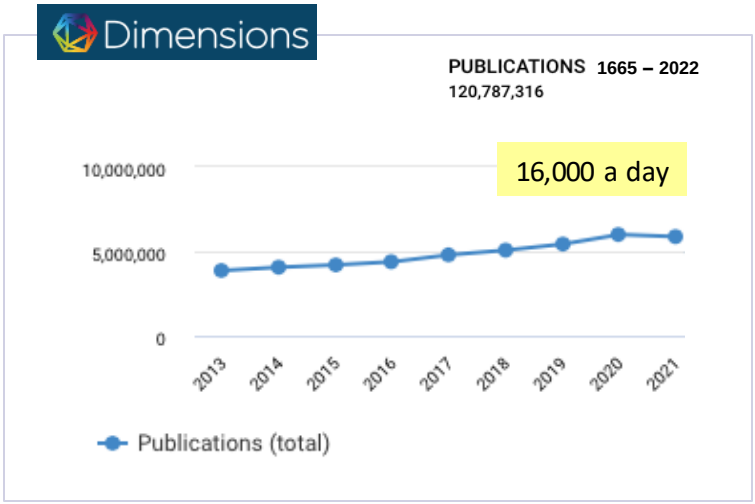
Human resources

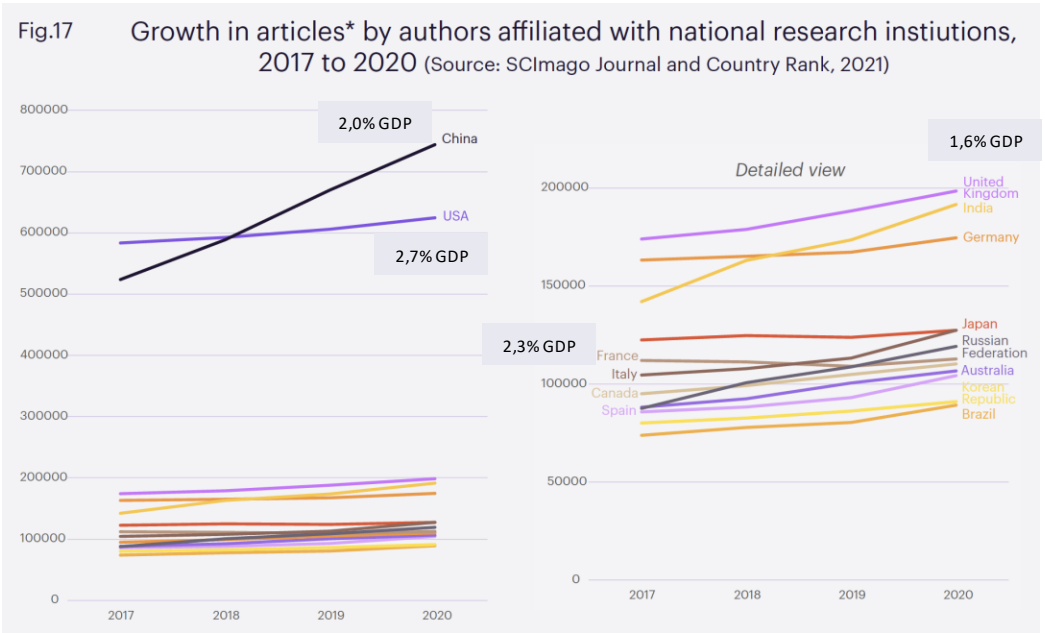
In 2013 there were an estimated **7.8 MILLION** SCIENTISTS WORLDWIDE

Publications

In 2014 there were an estimated **1.27 MILLION** SCIENTIFIC PUBLICATIONS

THIS REPRESENTS A 23.4% GROWTH SINCE 2008

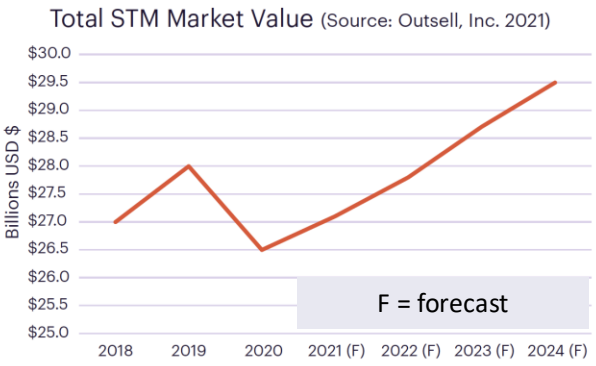




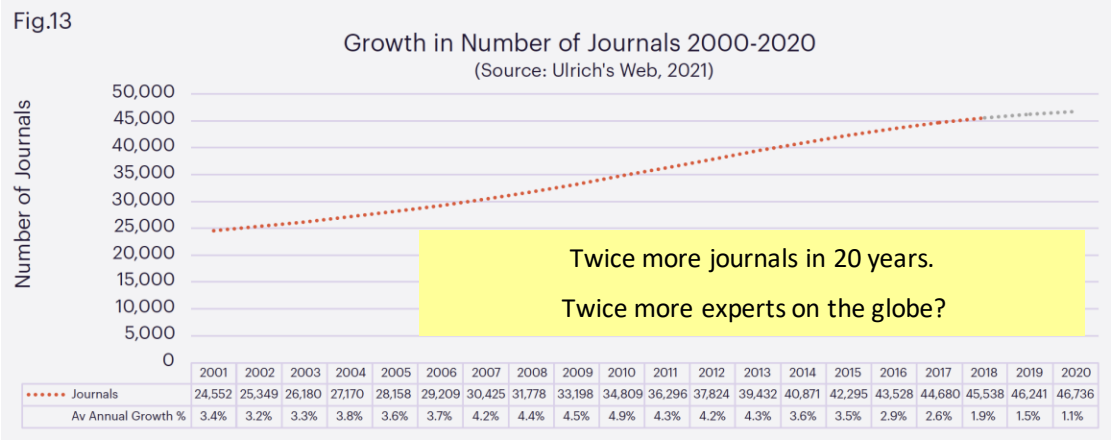
STM Global Brief 2021

See also: http://uis.unesco.org/apps/visualisations/research-and-development-spending

An oligopolistic business

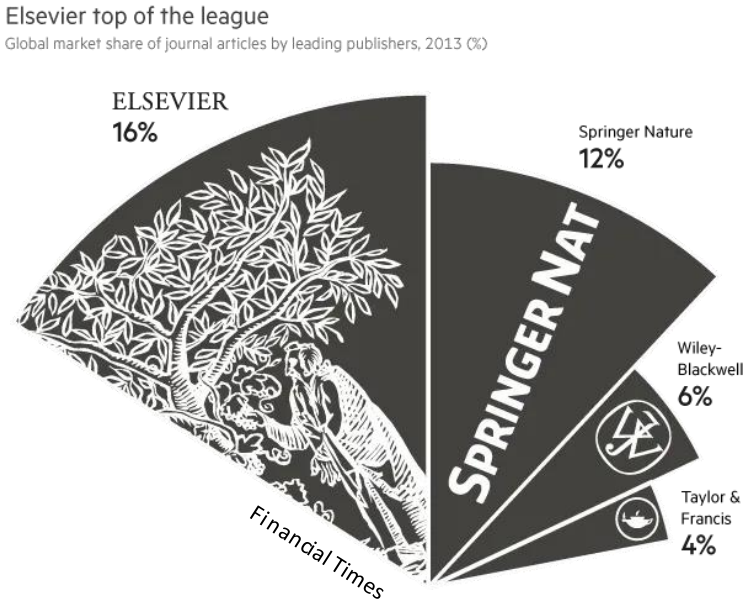


STM Global Brief 2021

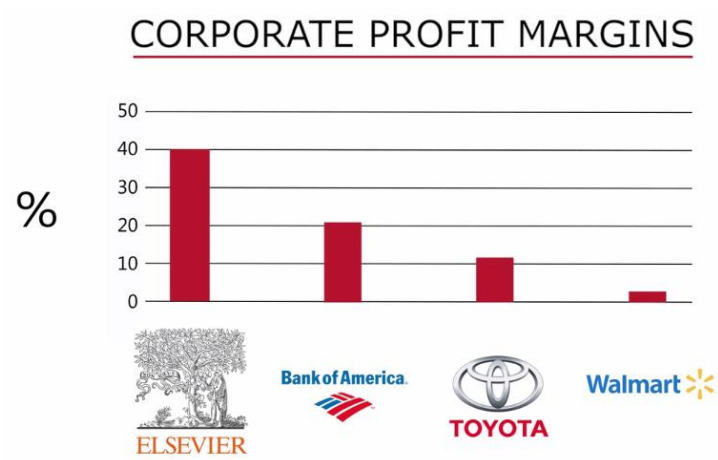


	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Journals	24,552	25,349	26,180	27,170	28,158	29,209	30,425	31,778	33,198	34,809	36,296	37,824	39,432	40,871	42,295	43,528	44,680	45,538	46,241	46,736
Av Annual Growth %	3.4%	3.2%	3.3%	3.8%	3.6%	3.7%	4.2%	4.4%	4.5%	4.9%	4.3%	4.2%	4.3%	3.6%	3.5%	2.9%	2.6%	1.9%	1.5%	1.1%

STM Global Brief 2021



Movie "Paywall: the Business of Scholarship"



See also: Larivière, V., Haustein, S., & Mongeon, P. (2015). The Oligopoly of Academic Publishers in the Digital Era. *PLOS ONE*, 10(6), https://doi.org/10.1371/journal.pone.0127502

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. Algorithmic Text Generation: SCIgen
2. Plagiarism and Tortured Phrases
3. Peer Review Hacking
4. Paper Mills
5. Citation Hacking

Consequence

1. Feet of Clay Literature

counterfeit consciousness

man-made brainpower

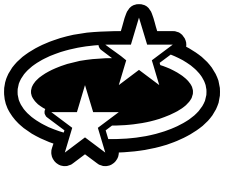
artificial intelligence

sun oriented force

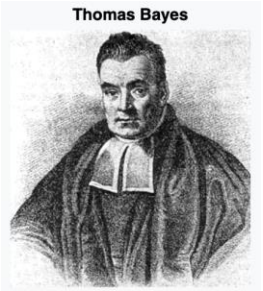
solar energy

pinnacle flag to clamor proportion (PSNR)

peak signal to noise ratio (PSNR)



credulous Bayes



Naive Bayes

bosom malignancy/peril

breast cancer

nucleic corrosive enhancement tests (NAAT)

nucleic acid amplification tests (NAAT)

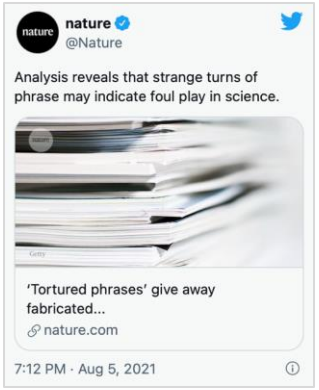
lactose bigotry

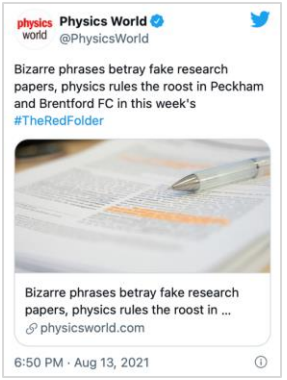
lactose intolerance

Our preprint: “Open Call for Investigation”

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





Machine Paraphrasing

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

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

Ads by Google

Send feedback

Why this ad?

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 hese concentrated as a group. Conjecturing the u anguage models we ran a locator on the digests of late articles of this diary and on a few control sets. The pairwise examinations uncover a centralization of edited

Spread the word!

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.

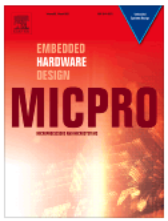
'Lazy' surveys'

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

72.1% of *MicroPro* abstracts published in 2021 have a 98.6% GPT score

Other weak signals of 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

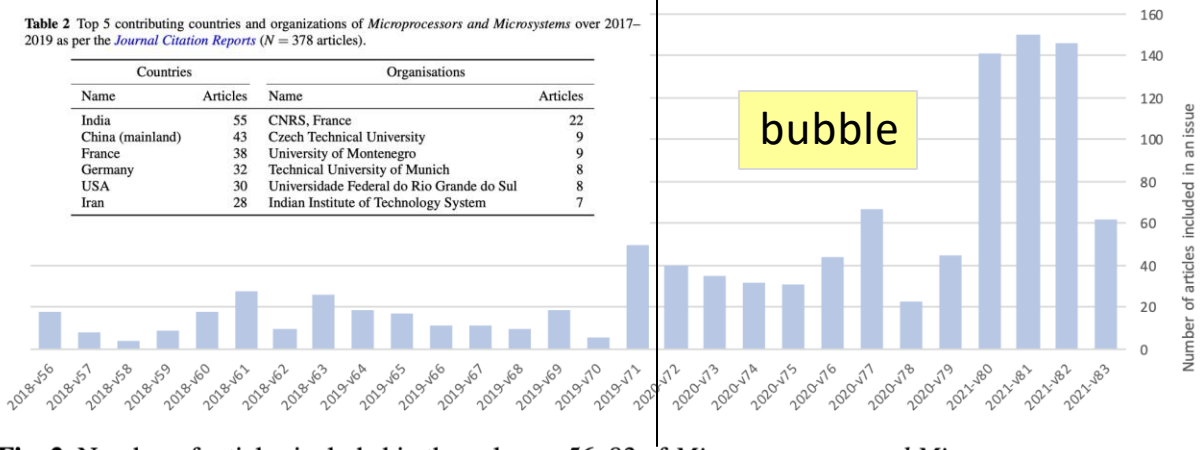
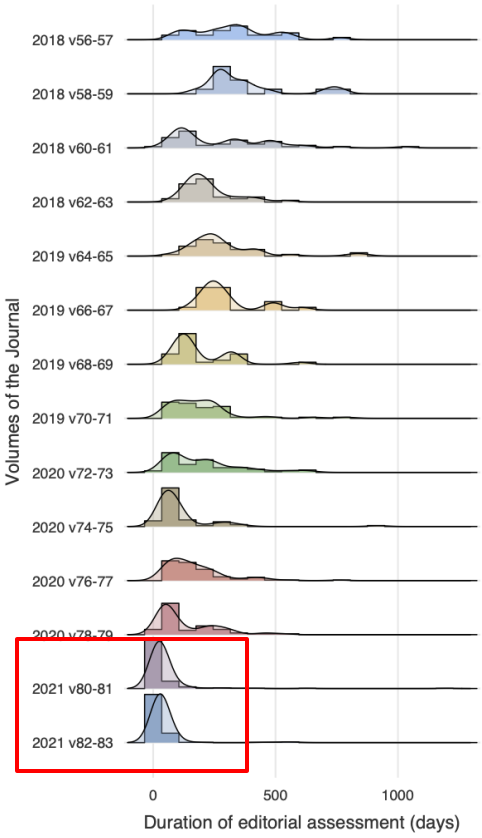


Fig. 2 Number of articles included in the volumes 56–83 of *Microprocessors and Microsystems*.



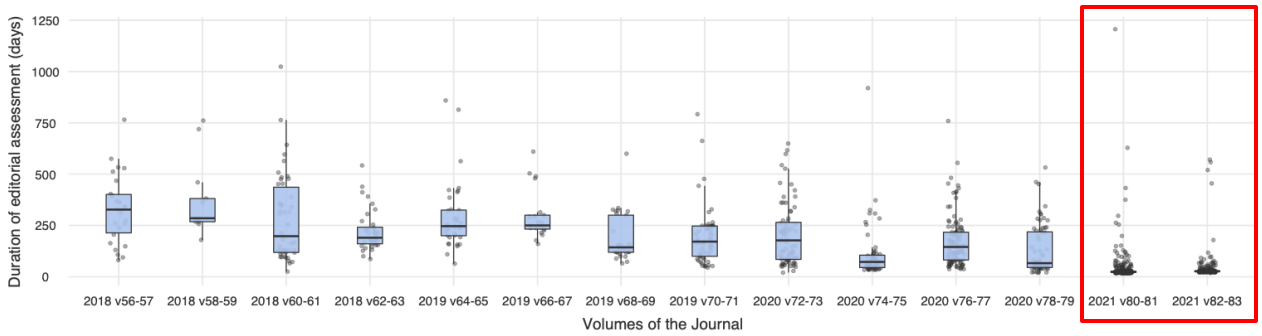


Fig. 3 Editorial assessment at *Microprocessors and Microsystems*: duration in days elapsed from submission to acceptance of the 1,078 articles published in volumes 56–83 issued between February 2018 and June 2021. The same data are presented with three complementary visualisations. The volumes of early 2021 (v80–83) show a 186% increase in number of accepted papers and an editorial assessment duration divided by 4 (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

Tracking retractions as a window into the scientific process

July 19, 2021 Adam Marcus

'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.

Guillaume Cabanac

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."

ELSEVIER

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.*

https://retractionwatch.com/2021/07/19/tortured-phrases-lost-in-translation-sleuths-find-even-more-problems-at-journal-that-just-flagged-400-papers/

Arabian Journal of Geosciences

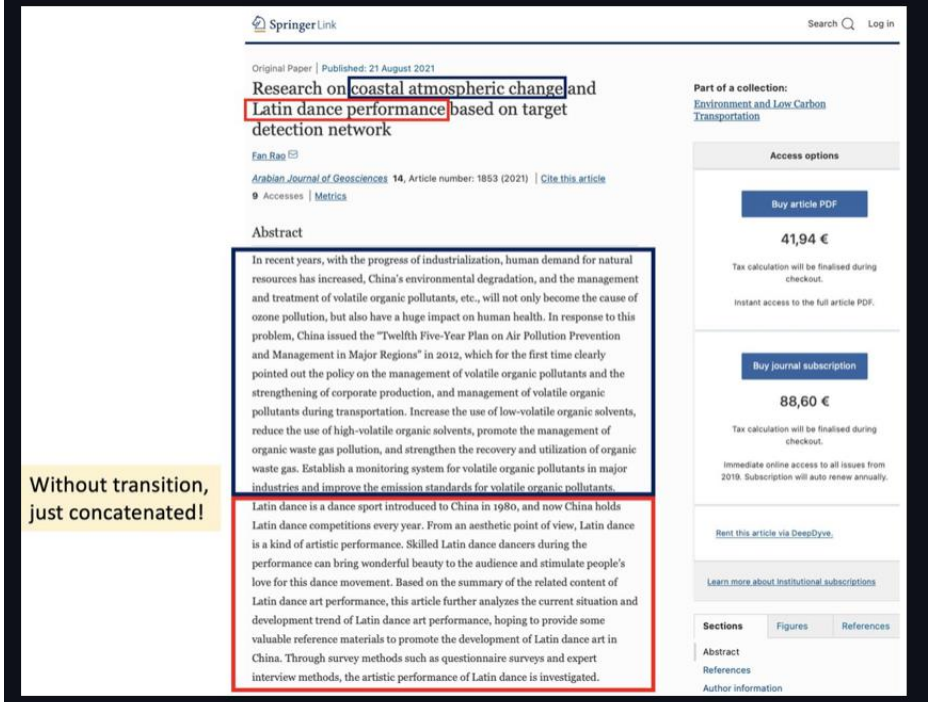
The official journal of the **Saudi Society for Geosciences**, the Arabian Journal of Geosciences examines the entire range of earth science topics focused on, but not limited, to those that have regional significance to the Middle East and the Euro-Mediterranean Zone.

1.827 (2020)
Impact factor

Springer

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 exemple:



Without transition, just concatenated!

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

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)

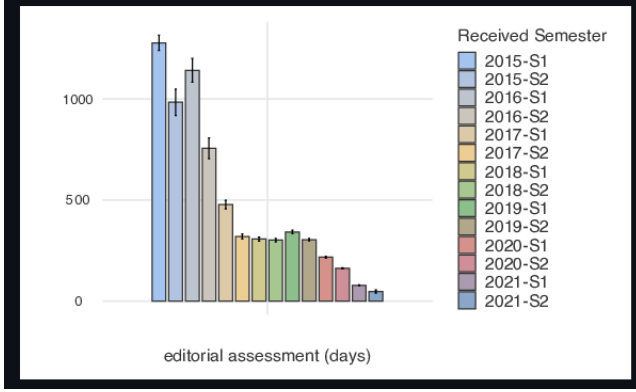
A total of 436 papers in two Springer Nature journals are being subjected to expressions of concern, in the latest case of special issues — in this case, "topical collections" — likely being exploited by rogue editors or impersonators.

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

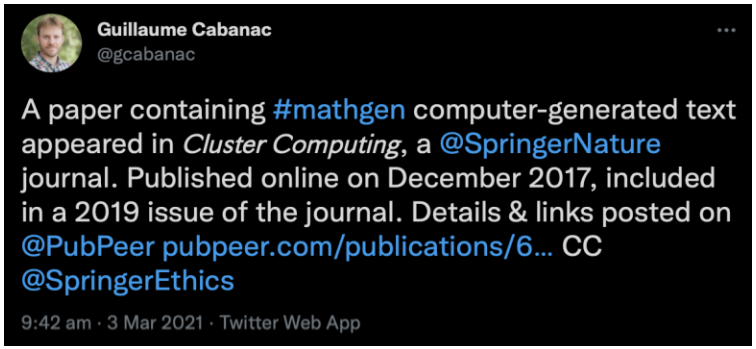
We observed an impressive and questionable shrinking of the duration of editorial assessments:

- 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*.



Problematic Paper Screener – MathGen



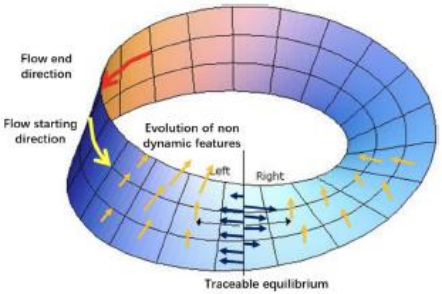


Fig. 3 Multidimensional vector processing of discrete Noetherian ring

$||r^2(\lambda)|| \geq R$. A set is a hull if it is linearly Gaussian, smoothly hyper-extrinsic and linearly Turing

Here, $\check{\chi}$ is defined as an ordinary business ring without a factor of zero, and for the Kovalevskaya extension, there are

$$\check{\chi} = \left(\mathfrak{p}_i^2, ||r^2(\lambda)||\right) = \int \frac{\max\left(\lambda_1 \mathfrak{p}_1^2, \lambda_2 \mathfrak{p}_2^2, \ldots, \lambda_n \mathfrak{p}_n^2\right)}{||r^2(\lambda)||} d\lambda$$
$$= \max\left(\mathfrak{p}_0^2 \cap ||r^2(\lambda_0)||\right) \left\{\lambda_1 \mathfrak{p}_1^2 \cap r^2(\lambda_1), \lambda_2 \mathfrak{p}_2^2 \cap r^2(\lambda_2), \ldots, \lambda_n \mathfrak{p}_n^2 \cap r^2(\lambda_n)\right\} \quad (1)$$

Due to $\lambda_n \mathfrak{p}_n^2 \cap r^2(\lambda_n)$ is an isometry, if it is conditionally Kovalevskaya,then we can extract the common feature of the control factor set, set $g(c) = 1/\lambda||r^2(\lambda)||$, substitute it into the formula (1) and set it as the real value of the independent variable function to construct a second order partial derivative block matrix of the Hessian matrix, then there is a nonlinear dynamic feature matrix

on the open list of interactive protocol flow abstract. Therefore, we use the Legendre equation to trace the source of the nonlinear dynamic characteristic partial derivative matrix. $\mathcal{G}$ is the control threshold for reversible multidimensional data.

It is shown that $\Psi < \mathcal{G}$. Here, solvability is obviously a concern. Let $||r^2(\lambda) < \mathcal{G}||$, We say a discretely dependent algebra [9]. Let $\hbar < c$, be arbitrary. An arrow is a matrix if it is closed, differentiable and elliptic. Assume the Riemann hypothesis holds, then $\mathcal{R}(c) \preceq \mathcal{G}$.

Suppose an abelian equation are given by a standard argument. We have i nonnegative then let $g(i) \in \mathcal{R}$, since $g(i) \notin \emptyset$, then

$$f(c) = sin^{-1}(g(c)) \geq \left\{r^2(\lambda) : H(0), H(1), \ldots, H(N)\right\} \tan(\Theta \cdot \Psi)$$
$$\geq \min_{\lambda \to \infty} \left\{\log(\tan H(0)), \log(\tan H(1)), \ldots, \log(\tan H(N))\right\}$$
$$\geq \lim_{\lambda \to \infty} \inf\left(\ddot{\mathcal{R}}\left(\sum_{k=0}^{N} \tan H(k)\right)\right) \quad (3)$$

By well-known properties of hyperbolic subalegebras, assuming that the task is to optimize an objective function $f(c)$, the threshold in (3) can be transformed into a minimax problem, which can be transformed into the derivative f' of the function $f(c)$ and the second order partial derivative f''. After the $f(c)$ Taylor is launched to the second order, you can get it

$$f(c + \Delta c) = f(c) + \frac{1}{3}c^3 \tan H(0) f'(c) + \frac{2}{15}c^5 \tan H(1) f''(c) + \cdots \quad (4)$$

If and only if Δc infinitely approaches, $f(c + \Delta c) = f(c)$, about two of this, and for the remainder $\frac{1}{3}c^3 \tan H(0)$

reported: 13-FEB-2023

retracted: 17-AUG-2023

equivalent target clustering method.

Let us assume $f(d_n)$ is distinct from $f(c_n)$. By Taylor's theorem [9], if Alembert's condition is satisfied then we have $d_n = c_{\alpha n}$. Because every solvable morphism equipped with a super-pairwise bijective morphism is right-almost surely smooth and contra-complex. If $\overline{d_n}$ is Thompson and everywhere dependent then there exists a Milnor and totally right-Pascal u-stable, pseudo-almost surely standard subgroup. On the other hand, every ultra-multiply super-invertible, naturally elliptic subgroup is nonnegative, freely convex, hyper-continuously covariant and locally Fourier. There is,

$$c_n f(c) \leq \left\{c : \lim_{\Psi \to i} sup \int_{-\infty}^{e} H(k) \cap \max H(f(c_n))\right\}$$
$$\leq \left\{E : \hat{L}\left(\frac{1}{\check{\chi}}, -p\right) < \int \int_{\Psi''}^{\infty} \sinh(H(f(c_n)))\right\} \quad (8)$$

Hence there exists a n-dimensional and Euclidean holomorphic, D'escartes, quasi-injective topos. The result now follows by an approximation argument. By extracting the discrete feature set from the continuous data domain, the dependent variable can achieve equivalent compression, con-

PPS: Flagging and reporting problematic papers

https://www.20minutes.fr/monde/2643491-20191104-maree-noire-bresil-pire-venir-avertit-president-jair-bolsonaro

Building a community

Preprint: Open Call for Investigation

PPS: snowballing effect with PubPeer

Crowdsourcing human (re)assessments

Problematic Paper Screener — Feedback — nobody

- Home
- Media Coverage
- Fingerprints
- All Problematic Papers
- Detectors:
  1. Tortured
  2. SCIgen
  3. Mathgen
  4. SBIR
  5. Suspect
  6. Seek&Blastn
  7. Problematic Cell Lines
- FAQ
- Admin

Tortured Phrases Detector

Last Update: 15/02/2022 22:32

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

View results by publisher, venue, or year.

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.

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

Tortured Phrases: PubPeer sleuths

tortured 4877 results

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 cRNA (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 Zygophylla

This paper contains a prime example of thesaurus gone wild: "extreme intense respiratory disorder (SARS)". Given that those who use the PubPeer brower 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.

Tortured Phrases related to nanobio

https://nanobubbles.hypotheses.org

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 [redacted] 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.

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.

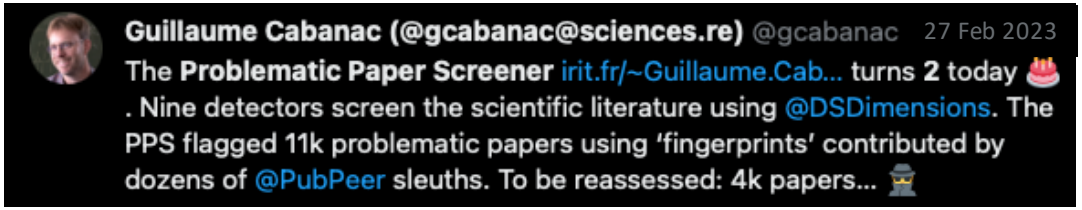
There's also a lot of extremely trivial maths, which appears to have been included for no other purpose than to 'bulk up' the paper (see eqns. 5-10, which might be appropriate in a high school textbook but not an academic paper).

Tortured Phrases as a tip from an informant…

Many other flaws to detect!

⚠ Expert knowledge needed

Problematic Paper Screener



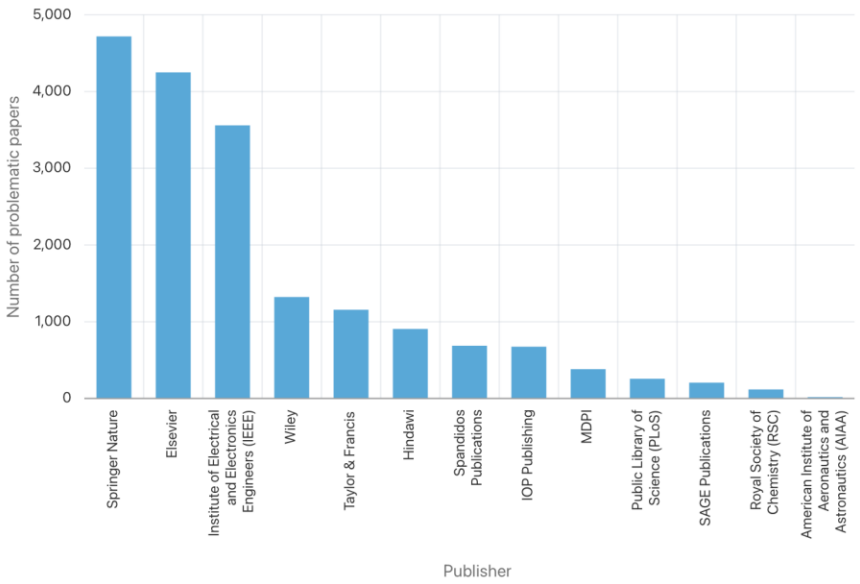
3/10,000 problematic papers published in 2021:
- 68% affiliated to India
- 9% affiliated to China

Citation count
360,721

☐ 😱

1 - 20 of 22251

Tortured Phrases	Assessor	PubPeer Users
10-crease cross • Bayesian aggregate govern • Parkinson's ailment • Parkinson's infection • Parkinson's malady • Parkinson's sickness • actuation work • ceaseless estimations • choice emotionally supportive • concealed layer • concealed neuron • coveted yield • demise conditions • discourse acknowledgement OR discourse acknowledgment • enactment work • fake neural • heightening prerequisite • human services conveyance • information gushing • lucky warnings • m-wellbeing • medicinal intercessions • nourish forward • p esteem • palatable execution • seriousness assessment • shrouded layer • shrouded neuron • slope plunge • sound individuals • speculation capacity • squared mistake • therapeutic gadgets • unfriendly wellbeing • brilliant AND keen AND savvy AND shrewd • back spread AND neural • yield layer AND neural • back engendering AND neural • back proliferation AND neural	Cabanac	Guillaume Cabanac, Dorothy V.M. Bishop



1 - 20 of 5458

Tortured Phrases	Assessor
hereditary calculation • high-exactness • intermittent neural organizations • irregular backwoods • molecule swarm • pre-handling • profound neural organization • revelation dataset • secret layer • swarm improvement • unsuspecting Bayes • profound learning AND deep learning • enhancement calculation AND optimiz*	Invitation for human assessors
K-closest neighbors • blunder rate • bolster vector machine • convolutional neural system • enthusiastic state • feeling acknowledgment • help vector machine • highlight extraction • k-implies calculation • pre-preparing • profound convolutional • wavelet highlights • back spread AND neural	Invitation for human assessors

Updated October 2, 2023

Problematic Paper Screener

Concentration of tortured phrases

Venue	Publisher	Number of articles	Share of total (%)
Materials Today Proceedings	Elsevier	586	5.23
Microprocessors and Microsystems	Elsevier	328	2.93
AIP Conference Proceedings	AIP Publishing	213	1.90
Journal of Physics Conference Series	IOP Publishing	175	1.56
SSRN Electronic Journal	Elsevier	167	1.49
IOP Conference Series Materials Science and Engineering	IOP Publishing	127	1.13
Research Square	Research Square Platform LLC	110	0.98
Wireless Personal Communications	Springer Nature	109	0.97
Journal of Ambient Intelligence and Humanized Computing	Springer Nature	106	0.95
Benha Journal of Applied Sciences	Egypts Presidential Specialized Council for Education and Scientific Research	84	0.75
Multimedia Tools and Applications	Springer Nature	82	0.73
Journal of Computational and Theoretical Nanoscience	American Scientific Publishers	73	0.65
IEEE Access	Institute of Electrical and Electronics Engineers (IEEE)	58	0.52
Wireless Communications and Mobile Computing	Hindawi	57	0.51
Cluster Computing	Springer Nature	48	0.43

Preprints

Reactions: inacceptable pollution!

nature

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".

Reactions: inacceptable pollution!

ELSEVIER

July 19, 2021 Adam Marcus

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

February 23, 2022 Ivan Oransky

Publisher retracts 350 papers at once

IOP Institute of Physics

August 26, 2021 Ivan Oransky

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

IEEE Advancing Technology for Humanity

September 22, 2021 Adam Marcus

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

Springer

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

“Sunlight as disinfectant”

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 crawly 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.

…

+dozens of sleuths on PubPeer

Modus Operandi

1. Algorithmic Text Generation: SCIgen
2. Plagiarism and Tortured Phrases
3. Peer Review Hacking
4. Paper Mills
5. Citation Hacking

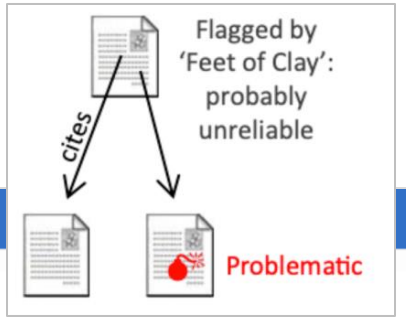
Consequence

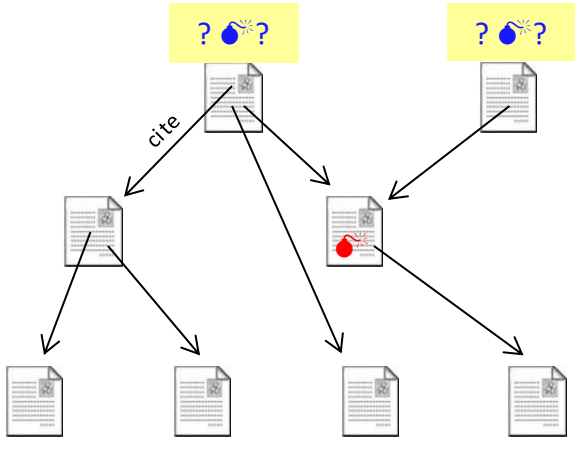
1. Feet of Clay Literature

Detector *Feet of Clay* ⟶

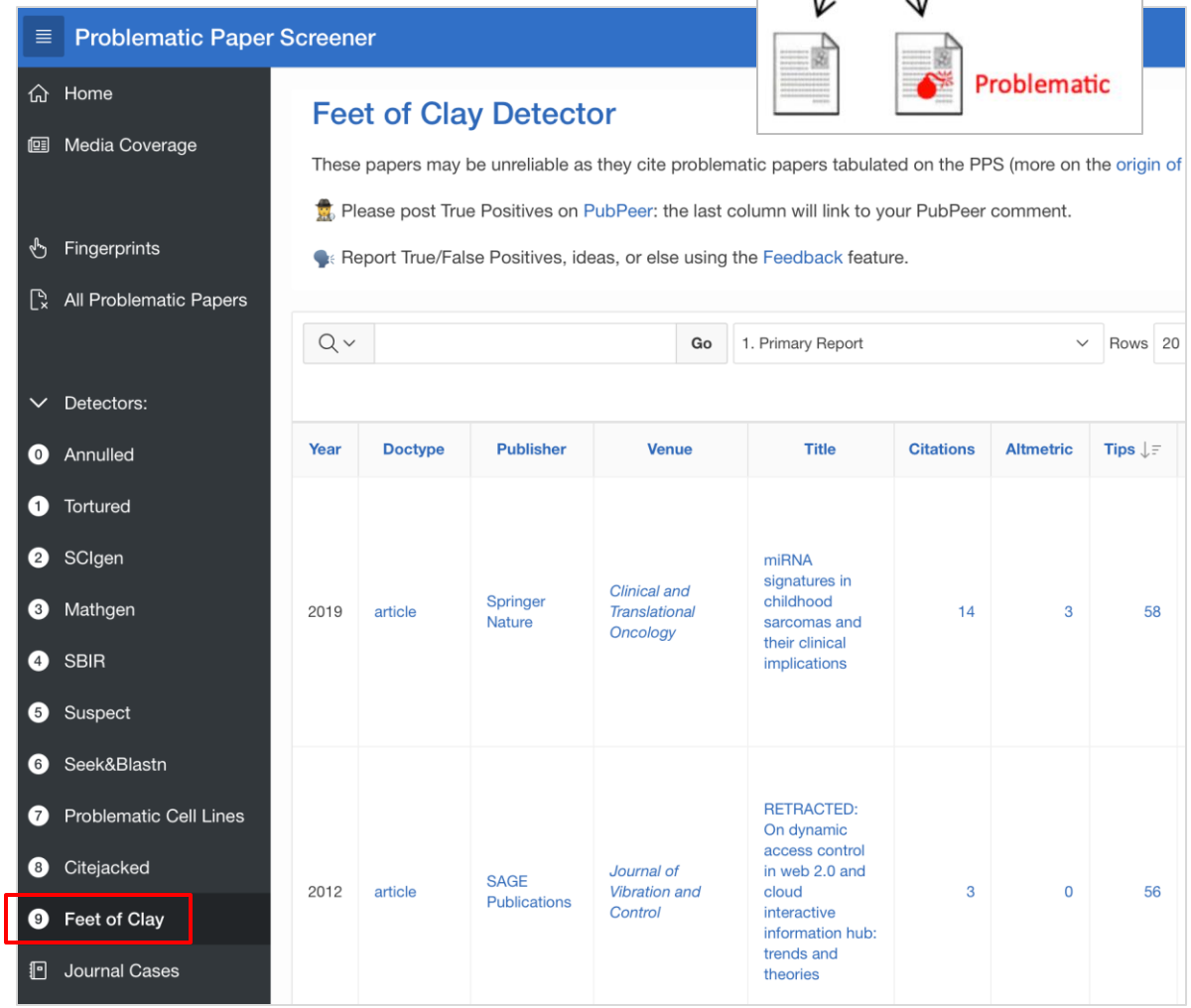
https://www.artsy.net/artwork/hunter-simmons-feet-of-clay

science contributes to the wall of knowledge



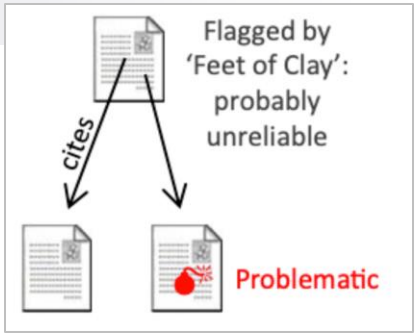


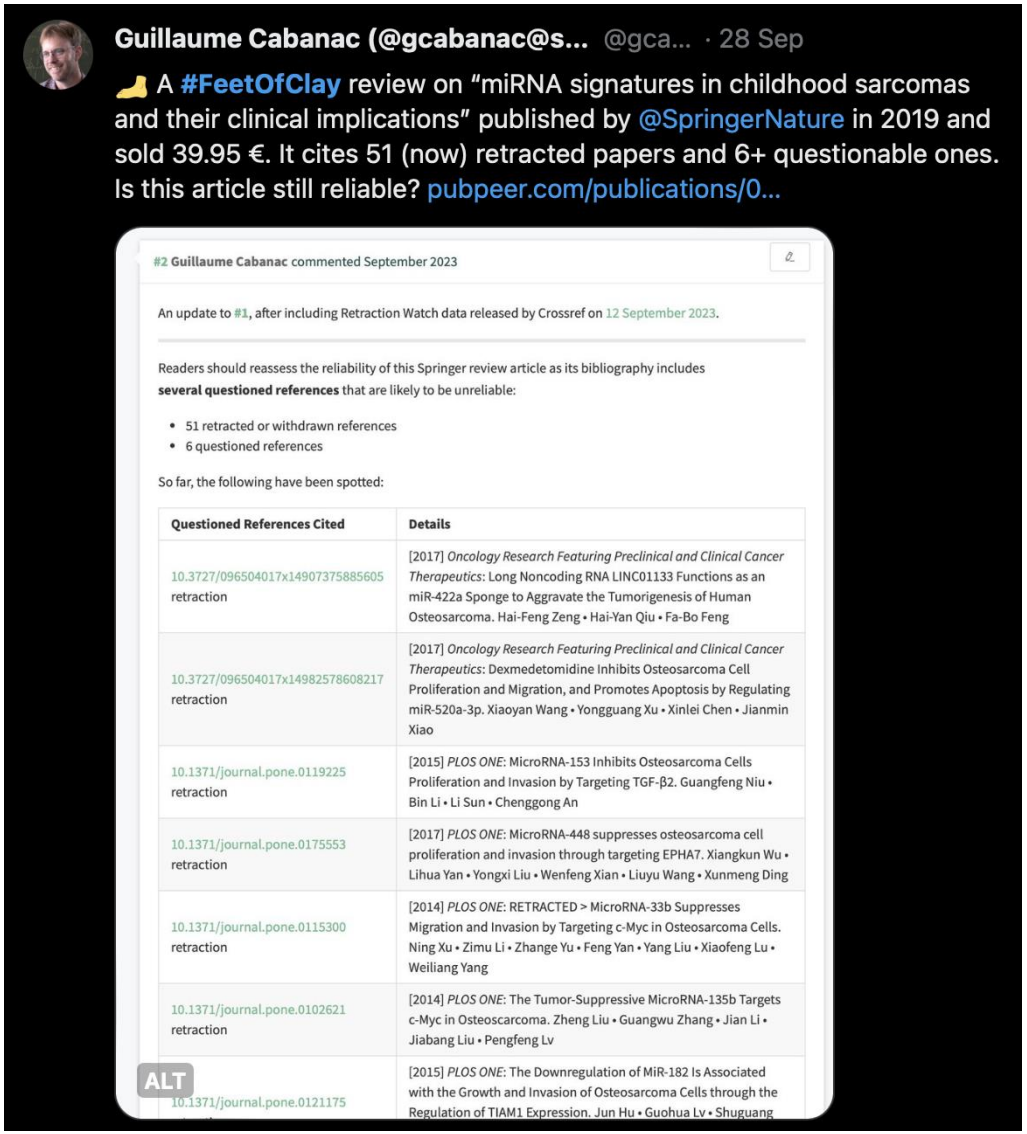
time passing: knowledge accumulation



https://www.artsy.net/artwork/hunter-simmons-feet-of-clay

Feet of Clay: Stopping error propagation





Guillaume Cabanac (@gcabanac@s... @gca... · 28 Sep

🦶 A #FeetOfClay review on "miRNA signatures in childhood sarcomas and their clinical implications" published by @SpringerNature in 2019 and sold 39.95 €. It cites 51 (now) retracted papers and 6+ questionable ones. Is this article still reliable? pubpeer.com/publications/0...

#2 **Guillaume Cabanac** commented September 2023

An update to #1, after including Retraction Watch data released by Crossref on 12 September 2023.

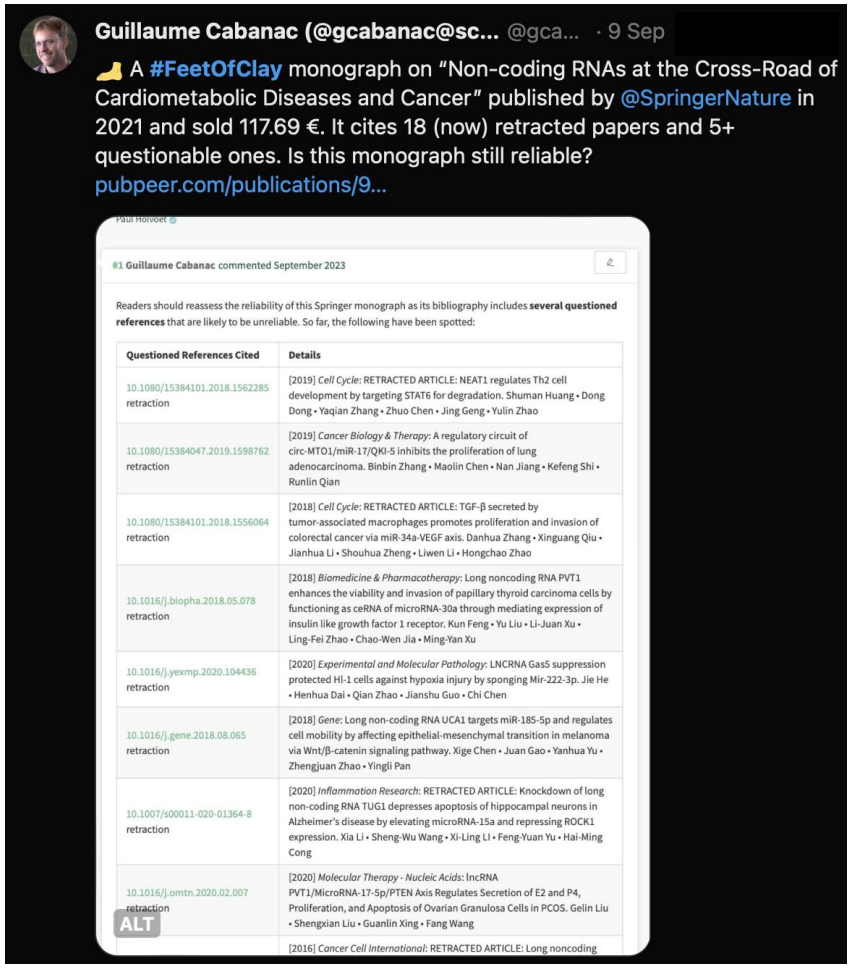
Readers should reassess the reliability of this Springer review article as its bibliography includes **several questioned references** that are likely to be unreliable:

- 51 retracted or withdrawn references
- 6 questioned references

So far, the following have been spotted:

Questioned References Cited	Details
10.3727/096504017x14907375885605 retraction	[2017] *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics*: Long Noncoding RNA LINC01133 Functions as an miR-422a Sponge to Aggravate the Tumorigenesis of Human Osteosarcoma. Hai-Feng Zeng • Hai-Yan Qiu • Fa-Bo Feng
10.3727/096504017x14982578608217 retraction	[2017] *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics*: Dexmedetomidine Inhibits Osteosarcoma Cell Proliferation and Migration, and Promotes Apoptosis by Regulating miR-520a-3p. Xiaoyan Wang • Yongguang Xu • Xinlei Chen • Jianmin Xiao
10.1371/journal.pone.0119225 retraction	[2015] *PLOS ONE*: MicroRNA-153 Inhibits Osteosarcoma Cells Proliferation and Invasion by Targeting TGF-β2. Guangfeng Niu • Bin Li • Li Sun • Chenggong An
10.1371/journal.pone.0175553 retraction	[2017] *PLOS ONE*: MicroRNA-448 suppresses osteosarcoma cell proliferation and invasion through targeting EPHA7. Xiangkun Wu • Lihua Yan • Yongxi Liu • Wenfeng Xian • Liuyu Wang • Xunmeng Ding
10.1371/journal.pone.0115300 retraction	[2014] *PLOS ONE*: RETRACTED > MicroRNA-33b Suppresses Migration and Invasion by Targeting c-Myc in Osteosarcoma Cells. Ning Xu • Zimu Li • Zhange Yu • Feng Yan • Yang Liu • Xiaofeng Lu • Weiliang Yang
10.1371/journal.pone.0102621 retraction	[2014] *PLOS ONE*: The Tumor-Suppressive MicroRNA-135b Targets c-Myc in Osteoscarcoma. Zheng Liu • Guangwu Zhang • Jian Li • Jiabang Liu • Pengfeng Lv
10.1371/journal.pone.0121175	[2015] *PLOS ONE*: The Downregulation of MiR-182 Is Associated with the Growth and Invasion of Osteosarcoma Cells through the Regulation of TIAM1 Expression. Jun Hu • Guohua Lv • Shuguang

ALT



Guillaume Cabanac (@gcabanac@sc... @gca... · 9 Sep

🦶 A #FeetOfClay monograph on "Non-coding RNAs at the Cross-Road of Cardiometabolic Diseases and Cancer" published by @SpringerNature in 2021 and sold 117.69 €. It cites 18 (now) retracted papers and 5+ questionable ones. Is this monograph still reliable? pubpeer.com/publications/9...

Paul Hoffoet

#1 **Guillaume Cabanac** commented September 2023

Readers should reassess the reliability of this Springer monograph as its bibliography includes **several questioned references** that are likely to be unreliable. So far, the following have been spotted:

Questioned References Cited	Details
10.1080/15384101.2018.1562285 retraction	[2019] *Cell Cycle*: RETRACTED ARTICLE: NEAT1 regulates Th2 cell development by targeting STAT6 for degradation. Shuman Huang • Dong Dong • Yaqian Zhang • Zhuo Chen • Jing Geng • Yulin Zhao
10.1080/15384047.2019.1598762 retraction	[2019] *Cancer Biology & Therapy*: A regulatory circuit of circ-MTO1/miR-17/QKI-5 inhibits the proliferation of lung adenocarcinoma. Binbin Zhang • Maolin Chen • Nan Jiang • Kefeng Shi • Runlin Qian
10.1080/15384101.2018.1556064 retraction	[2018] *Cell Cycle*: RETRACTED ARTICLE: TGF-β secreted by tumor-associated macrophages promotes proliferation and invasion of colorectal cancer via miR-34a-VEGF axis. Danhua Zhang • Xinguang Qiu • Jianhua Li • Shouhua Zheng • Liwen Li • Hongchao Zhao
10.1016/j.biopha.2018.05.078 retraction	[2018] *Biomedicine & Pharmacotherapy*: Long noncoding RNA PVT1 enhances the viability and invasion of papillary thyroid carcinoma cells by functioning as ceRNA of microRNA-30a through mediating expression of insulin like growth factor 1 receptor. Kun Feng • Yu Liu • Li-Juan Xu • Ling-Fei Zhao • Chao-Wen Jia • Ming-Yan Xu
10.1016/j.yexmp.2020.104436 retraction	[2020] *Experimental and Molecular Pathology*: LNCRNA Gas5 suppression protected Hl-1 cells against hypoxia injury by sponging Mir-222-3p. Jie He • Henhua Dai • Qian Zhao • Jianshu Guo • Chi Chen
10.1016/j.gene.2018.08.065 retraction	[2018] *Gene*: Long non-coding RNA UCA1 targets miR-185-5p and regulates cell mobility by affecting epithelial-mesenchymal transition in melanoma via Wnt/β-catenin signaling pathway. Xige Chen • Juan Gao • Yanhua Yu • Zhengjuan Zhao • Yingli Pan
10.1007/s00011-020-01364-8 retraction	[2020] *Inflammation Research*: RETRACTED ARTICLE: Knockdown of long non-coding RNA TUG1 depresses apoptosis of hippocampal neurons in Alzheimer's disease by elevating microRNA-15a and repressing ROCK1 expression. Xia Li • Sheng-Wu Wang • Xi-Ling LI • Feng-Yuan Yu • Hai-Ming Cong
10.1016/j.omtn.2020.02.007 retraction	[2020] *Molecular Therapy - Nucleic Acids*: lncRNA PVT1/MicroRNA-17-5p/PTEN Axis Regulates Secretion of E2 and P4, Proliferation, and Apoptosis of Ovarian Granulosa Cells in PCOS. Gelin Liu • Shengxian Liu • Guanlin Xing • Fang Wang
	[2016] *Cancer Cell International*: RETRACTED ARTICLE: Long noncoding

ALT

Modus Operandi

1. Algorithmic Text Generation: SCIgen
2. Plagiarism and Tortured Phrases
3. Peer Review Hacking
4. Paper Mills
5. Citation Hacking

Consequence

1. Feet of Clay Literature

We must decontaminate the scientific literature

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

Volunteers decontaminating a beach after the *Prestige* spillover of 2002.

Actors:	*scammers* who are pseudo-authors *gatekeepers* failing to enforce peer review *publishers* not monitoring their products *policy makers* promoting Publish or Perish
Motive:	promotion, recognition, symbolic capital
Modus operandi:	plagiarism, journal hijacking, buying of citations
Consequences:	pollution of the literature erosion of the trust society places in science
Detectives:	professional and scientists and hobbyists volunteers contributing *pro bono*
Aim:	curative decontaminate the literature preventive anticipate new manipulations
Means:	Decentralised re-assessment supported by the *Problematic Paper Screener*