

# The 'Problematic Paper Screener' automatically selects suspect publications for post-publication (re)assessment

Guillaume Cabanac, Cyril Labbé, Alexander Magazinov

### ▶ To cite this version:

Guillaume Cabanac, Cyril Labbé, Alexander Magazinov. The 'Problematic Paper Screener' automatically selects suspect publications for post-publication (re)assessment. 7th World Conference on Research Integrity (WCRI 2022), May 2022, Cape Town, South Africa. 10.48550/arXiv.2210.04895. hal-03829578

HAL Id: hal-03829578

https://hal.science/hal-03829578

Submitted on 31 Aug 2023

**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.



May 30<sup>th</sup>, 2022 WCRI, South Africa

# The Problematic Paper Screener

automatically selects suspect publications for postpublication (re)assessment















man-made brainpower

artificial intelligence

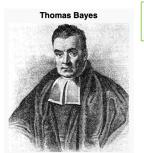
sun oriented force

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)

Extreme intense respiratory disorder

Severe acute respiratory syndrome (SRAS)

# 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 (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)













Register Log in

Pricing About FAQs API Contact

# Machine Paraphrasing

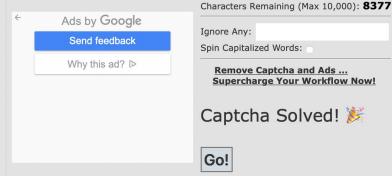
https://spinbot.com

# **Spinbot**

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



### Here are two paid options:

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)
- 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 <u>developer account management page</u>.

#### **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 than 'man-made reasoning expressions and study one for tormented hese concentrated as

a group. Conjecturing the uprana 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

Spread the word!

# PPS: Flagging and reporting problematic papers



# Our contribution: Combing the literature for nonsense

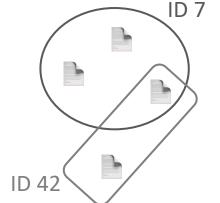


JASST ASSISTANCE OF THE COLUMN ASSISTANCE OF T							
RESEARCH ARTICLE   ① Open Access   ② ①							
Prevalence of nonsensical algorithmically generated papers in the scientific literature							
the scientific literature							
Guillaume Cabanac 📉 Cyril Labbé,							
First published: 25 May 2021   https://doi.org/10.1002/asi.24495							

Fingerprint - Tortured Phrase	Expected Text		Retrieved Papers ↓ <i>≡</i>		
"preformation" AND "cell" AND ("rat" OR "mouse" OR "ultimate" OR "destiny")	procedure AND ultima	ate/final density	6,961		
"highlight extraction"	Discosol	feature extraction		1,783	
"driving reason"		Phrasal leading cause			
"as pursues"	search	as follows		1,588	
"brilliant" AND "keen" AND "savvy" AND "shrewd"	"brilliant" AND "keen" AND "savvy" AND "shrewd"				
"substantial metals"		heavy metals		1,428	
"created nations"	"created nations"				
"crude information"		raw data		1,378	
"regenerative organs" AND "reproductive"		reproductive organs		1,141	
"vitality utilization"		energy use		1,043	
"discourse acknowledgement" OR "discourse acknow	ledgment"	voice recognition		1,004	
"observational examination"	observational study	/	988		
"supply route" AND artery		artery	Nb of	906	
"choice tree"	Boolean	decision tree	retrieved	877	
	conjunctive		Papers		

operator





# **Building a community**

Media Coverage

Fingerprints

✓ Detectors:

1 Tortured

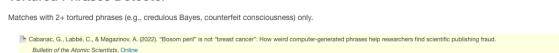
2 SCIgen 3 Mathgen 4 SBIR 5 Suspect 6 Seek&Blastn 7 Problematic C

? FAQ ≥ Admin

All Problematic Papers

PPS: snowballing effect with PubPeer





🕒 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.

Qv			Go	1. Primary Report	~	Rows	20	~	Action	ons V					
														1	1 - 20 of 4259
Year	Doctype	Publisher	Venu	ue	Title				Cita	tations	Altmetric	Tips ↓=	Tortured Phrases	Assessor	PubPeer User

Year	Doctype	Publisher	Venue	Title	Citations	Altmetric	Tips ↓=	Tortured Phrases	Assessor	PubPeer User
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	Al 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	4
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	



Last Update: 15/02/2022 22:32

II View results by publisher, venue, or year.

### Tortured Phrases: PubPeer sleuths

#### 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:

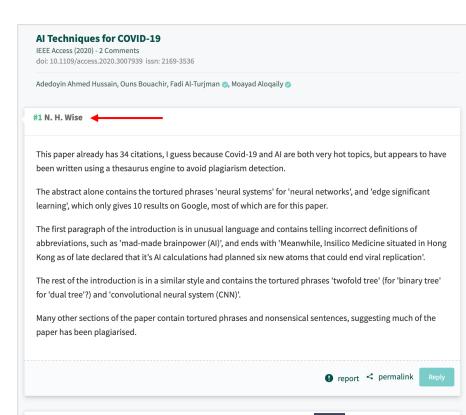
https://commons.wikimedia.org/wiki/File:Microarray\_exp\_horizontal.svg

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, DEAL 2006, LNCS 4224, pp. 1087 – 1094, 2006, DOI: 10.1007/11875581\_129. Here is an example:

tortured

Q

1266 results



This paper contains a prime example of thesaurus gone wild: "extreme intense respiratory disorder (SARS)".

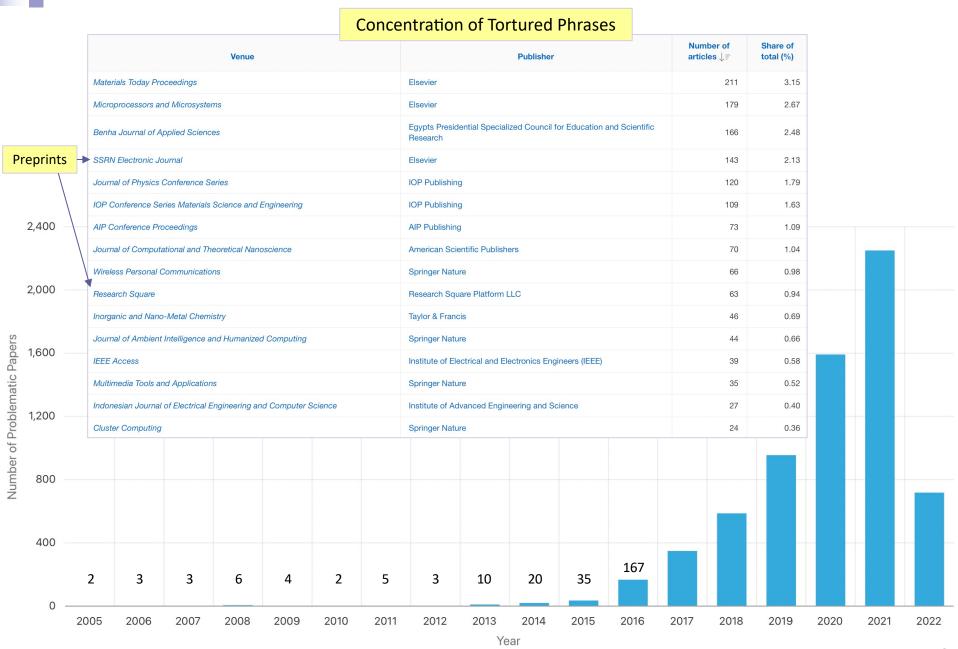
turns up in a search or on a web page, it seems both fair and useful to ask the authors: What contribution to knowledge do you think you have made with this publication? The same question could fairly and usefully be

Given that those who use the PubPeer brower plugin are going to be directed to this page whenever this paper

#2 Salsola Zygophylla

asked of those who cited it.

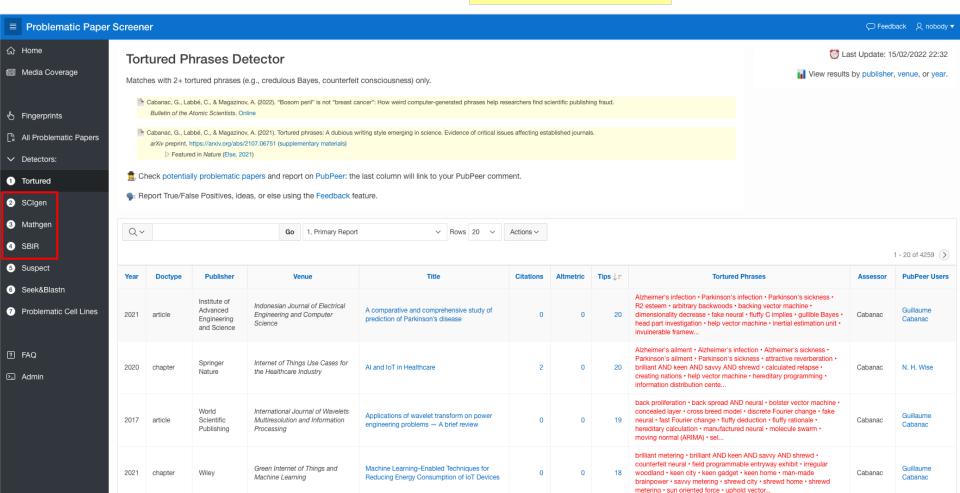
# **Problematic Paper Screener**



Updated May 29, 2022

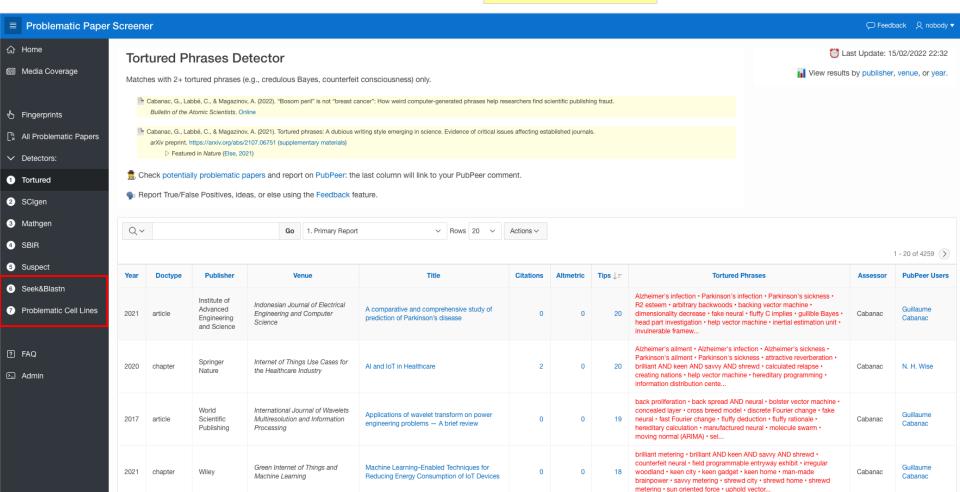
# Building a community

# Detecting SCIgen and its clones



# Building a community

# Potentially erroneous papers



# Problematic Paper Screener

### Guillaume Cabanac @gcabanac · 1 Mar

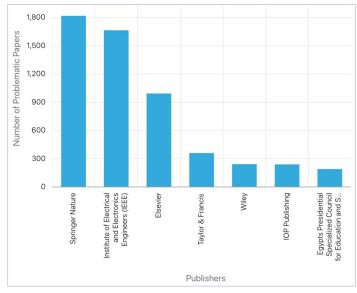
#### Citation count

24,560

3/10,000 problematic papers published in 2021:

- 66% affiliated to India
- 8% affiliated to China

	1 -	200 o 6711
Tortured Phrases	Assessor ↓=	PubPeer Users
area of intrigue · attractive reverberation · blunder rate · calculated relapse · convolutional neural system · designs preparing unit · discourse acknowledgement OR discourse acknowledgment · figured tomography · fluffy C implies · genuine positive (TP) · help vector machine · highlight extraction · prescient capacity · processed tomography (CT) · profound neural system · registered tomography	Magazinov	Alexander Magazinov
arbitrary backwoods • crude information • directing convention • haze figuring • haze registering • keen matrix • malevolent hub • malignant hub • man in the center • middle of the road hub • misfortune rate • specially appointed system • steering convention • brilliant AND keen AND savvy AND shrewd	Magazinov	Alexander Magazinov
choice tree • concealed layer • fluffy rationale • heart assault • heart sickness • hereditary calculation • individuals bite the dust • learning revelation in databases • shrouded examples • shrouded layer • back engendering AND neural • back proliferation AND neural • supply route AND artery • yield layer AND neural	Magazinov	Alexander Magazinov, Hoya Camphorifolia



1 - 2	00 o <sup>‡</sup> 5063 >
Tortured Phrases	Assessor ↓=
arbitrary woodland • backing vector machine • blunder rate • bosom disease • bosom growth • bosom malignant • differentially communicated qualities • electronic wellbeing records • gathering of patients • irregular backwoods • irregular woodland	Invitation for human assessors
crude information • direct blend • hereditary calculation • highlight extraction • man-made consciousness • mind waves • mind-PC interface • sign to commotion • upper appendage	Invitation for human assessors
bolster vector machine • choice tree • concealed Markov • discourse feeling acknowledgment • fake neural • feeling acknowledgment • head part investigation • highlight extraction • speaker autonomous	Invitation for human assessors

Updated May 29, 2022 12