



HAL
open science

Tortured phrases: A dubious writing style emerging in science. Evidence of critical issues affecting established journals

Guillaume Cabanac, Cyril Labbé, Alexander Magazinov

► To cite this version:

Guillaume Cabanac, Cyril Labbé, Alexander Magazinov. Tortured phrases: A dubious writing style emerging in science. Evidence of critical issues affecting established journals. 2021. hal-03596867

HAL Id: hal-03596867

<https://hal.science/hal-03596867>

Preprint 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.



Distributed under a Creative Commons Attribution 4.0 International License

Tortured phrases: A dubious writing style emerging in science Evidence of critical issues affecting established journals

Guillaume Cabanac · Cyril Labbé · Alexander Magazinov

Version: July 12, 2021

Abstract 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 concentration of abstracts flagged as ‘synthetic’ in the journal. We also highlight irregularities in its operation, such as abrupt changes in editorial timelines. We substantiate our call for investigation by analysing several individual dubious articles, stressing questionable features: tortured writing style, citation of non-existent literature, and unacknowledged image reuse. Surprisingly, some websites offer to rewrite texts for free, generating gobbledegook full of tortured phrases. We believe some authors used rewritten texts to pad their manuscripts. We wish to raise the awareness on publications containing such questionable AI-generated or rewritten texts that passed (poor) peer review. Deception with synthetic texts threatens the integrity of the scientific literature.

Keywords AI-generated texts · GPT · Misconduct · Research integrity · Tortured phrases

All parts of this work are contributed jointly by all authors. Authors are listed in alphabetical order.

G. Cabanac

University of Toulouse, Computer Science Department, IRIT UMR 5505 CNRS, 31062 Toulouse, France

E-mail: guillaume.cabanac@univ-tlse3.fr

ORCID: [0000-0003-3060-6241](https://orcid.org/0000-0003-3060-6241)

C. Labbé

Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France

E-mail: cyril.labbe@univ-grenoble-alpes.fr

ORCID: [0000-0003-4855-7038](https://orcid.org/0000-0003-4855-7038)

A. Magazinov

Yandex, 82 Sadovnicheskaya str., Moscow 115035, Russia

E-mail: magazinov-al@yandex.ru

ORCID: [0000-0002-9406-013X](https://orcid.org/0000-0002-9406-013X)

1 Introduction

In science there is a history of scholarly publishing stings (Faulkes, 2021). Scholars and journalists have submitted nonsensical papers to various venues to expose dysfunctional peer review. These nonsensical papers submitted can be written by humans (e.g., the *Sokal Affair* and Bohannon, 2013) or computer generated (e.g., SCIGen, Mathgen).

Computer programs designed to generate fake papers and sting publishers are also reused by academic tricksters who easily produce the (fake) publications or (fake) citations they desperately need. As a result, meaningless randomly generated scientific papers end up being served and sometimes sold by various publishers with a prevalence estimated to 4.29 papers every one million papers (Cabanac & Labbé, in press; Van Noorden, 2021). Such papers can be easily spotted by both human and machine; natural language generation tools thus appear to be a cheap and dirty alternative to buying publications from paper mills, which also seems on the rise (Else & Van Noorden, 2021; Mallapaty, 2020).

The major recent advances in language models based on neural networks may sooner or later lead to a new kind of scientific writing. Incurable optimists would consider that automatic translation, writing enhancement, and summarising tools help authors to produce better scientific papers. Whole books are now generated from thousands of articles used as input (Beta Writer, 2019; Day, 2019; Visconti, 2021). But the generative power of modern language models can also be considered a threat to the integrity of the scientific literature. For example, the dangerous nature of the GPT-3 language model (Brown et al., 2020) was discussed extensively (Hutson, 2021).

With this in mind, we report observations about a reputable journal along several lines: occurrences of tortured phrases in publications (e.g., ‘flag to clamor’ in lieu of the established ‘signal to noise’), indication – if not evidence – of AI-generated abstracts, as well as questionable texts and images (including reuse from other sources without proper acknowledgement), as well as recent changes in editorial management (including shortened time between reception and acceptance of manuscripts). Without any definitive proof, we thus provide hints of the rise of a new kind of probably synthetic, nonsensical scientific texts.

The outline of this open call for investigation is as follows. Section 2 reports a set of ‘tortured phrases’ spotted in the literature. We then focus our study on *Microprocessors and Microsystems*, an Elsevier journal in which they concentrate (Sect. 3). We report intriguing irregularities in the editorial timelines of this journal (Sect. 4). The presence of synthetic text generated by advanced language model is hypothesised and Sect. 5 reports the screening of recent publications using an off-the-shelf software detecting synthetic text. Section 6 provides factual evidence of inappropriate and/or poor quality publications. We discuss possible sources of synthetic papers in Sect. 7 before concluding with a call to the scientific community for further investigation on this matter (Sect. 8).

2 Tortured phrases found in published academic articles

While reviewing recent publications, we encountered an unusual and disappointing phenomenon: well-known and well-established scientific terms were replaced by unconventional phrases. In a typical case, a word-by-word synonymical substitution is applied to a multi-word term. We call *tortured phrases* these phrases that are incorrectly used in lieu of well-established ones. Table 1 shows some tortured phrases that we were able to find in the literature (at first by chance and then by snowballing with already identified terms) and retro-engineer to infer the correct wording that readers would expect.

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)

On May 25, 2021 we queried the Dimensions academic search engine (Herzog, Hook, & Konkiel, 2020) to retrieve the set of papers containing tortured phrases known at that date (see Fig. 1). Note that some tortured phrases may be used in a legitimate way (e.g., ‘enormous information’ in certain contexts) and that the full-text indexing performed by Dimensions ignores punctuation. This may lead to retrieve few articles not using a tortured phrase. Dimensions was chosen for its coverage of the literature that is larger than the Web of Science and Scopus (Singh, Singh, Karmakar, Leta, & Mayr, 2021) and because it is free for scientometric research.¹

The *Microprocessors and Microsystems* journal was ranked first among the venues listed by Dimensions in decreasing number of matching articles (Fig. 1). We selected this journal for further investigation in the remainder of this study.

¹ <https://www.dimensions.ai/scientometric-research/>

The screenshot shows the Dimensions search engine interface. The search query is "tortured neural organization". The results are displayed in a list format, with the first result being "Evaluating the ecological state of Chinese Lake Baiyangdian (BYD) based on Ecological Network Analysis" by Yong Zeng, Yanwei Zhao, Zefeng Qi, and 2021, Ecological Indicators - Article. The interface includes filters on the left, a central list of results, and an analytical views panel on the right showing research categories and an overview chart of citations from 2012 to 2021.

Fig. 1 Published articles retrieved with the Dimensions academic search engine (<https://bit.ly/3vm8tAW>). The query targets the full-text index with 30 tortured phrases that we listed as of May 25, 2021 (earlier version of Tab. 1).

3 The *Microprocessors and Microsystems* journal

Founded in 1976, the *Microprocessors* journal² was quickly renamed *Microprocessors and Microsystems* starting from Volume 3 in 1978. It is now published by Elsevier³ and classified by Scopus⁴ in four subject areas of Computer Science:

- Artificial Intelligence
- Computer Networks and Communications
- Hardware and Architecture
- Software

Based on Scopus data, *Scimago Journal Ranking* ranked *Microprocessors and Microsystems* in Q3, that is the third quartile for the four subject areas.⁵ In the latest *Journal Citation Reports* curated by Clarivate Analytics, this journal appears in the *Science Citation Index Expanded* under three categories:

- Computer Science, Hardware & Architecture
- Computer Science, Theory & Methods
- Engineering, Electrical & Electronic

As of June 2021, the latest *Journal Citation Reports* entry for *Microprocessors and Microsystems* covered 2017–2019. The journal published 378 articles with the top 5 contributing countries and organisations in Tab. 2. Its *Journal Impact Factor* increased from 0.471 to 1.161 over 2015–2019, that is a 146% increase over four years.

² <https://www.sciencedirect.com/journal/microprocessors>

³ <https://www.sciencedirect.com/journal/microprocessors-and-microsystems>

⁴ <https://www.scopus.com/sourceid/15552>

⁵ <https://www.scimagojr.com/journalsearch.php?q=15552&tip=sid>

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

In what follows, we conduct a more in-depth analysis of this venue over the period February 2018 to June 2021 for which we collected data. Figure 2 shows a radical change in the number of articles published per volumes starting in 2020.

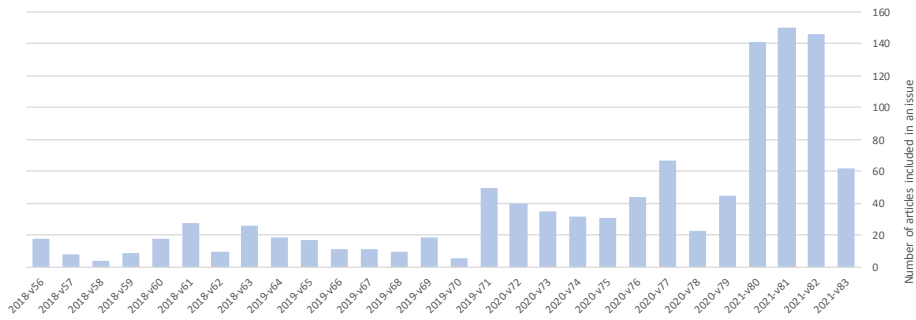


Fig. 2 Number of articles included in the volumes 56–83 of *Microprocessors and Microsystems*.

Microprocessors and Microsystems publishes articles with DOIs minted by Crossref (Hendricks, Tkaczyk, Lin, & Feeny, 2020). We queried the Crossref REST API⁶ to collect the DOIs of papers published in volumes 56–83 (February 2018 to June 2021) of this journal.⁷ We used the Elsevier subscription of the University of Toulouse (GC’s affiliation) to download each article in fulltext XML via the Elsevier API⁸ and extract the following metadata:

- Identifiers: Publisher Item Identifier (PII) and Digital Object Identifier (DOI)
- Timeline: dates of submission, revision, and acceptance
- Publication type (e.g., full-length article, review article, editorial, erratum)
- Title
- Abstract
- Authors’ countries

We filtered out publication types other than ‘full-length articles’ and removed two articles with a missing acceptance date. The revision date was missing for 41 articles; we

⁶ <https://github.com/CrossRef/rest-api-doc>

⁷ <https://www.sciencedirect.com/journal/microprocessors-and-microsystems/issues>

⁸ <https://dev.elsevier.com>

assumed acceptance without revision for these. We noted that no countries were present in the XML format for 12 articles. The final dataset contains 1,078 articles (See [Appendix](#)).

4 Irregularities of the editorial assessment in *Microprocessors and Microsystems*

We use the term ‘editorial assessment’ to denote the time from submission of a manuscript to its acceptance, including: preliminary screening, invitation of reviewers, rounds of peer review, and final decision. The published metadata for each paper characterises its editorial assessment with three dates: submission, revision, and acceptance.

The analysis of the dates of submission vs dates of acceptance reveals a sudden shortening of editorial assessment for volumes published in 2021. Most articles were published after an editorial assessment surprisingly short. Affiliations from China and India were over-represented. Several blocks of articles shared the same dates of submission and acceptance. These observations depart from the typical publication output of *Microprocessors and Microsystems* before 2021.

Our call for investigation (Sect. 8) invites readers to perform a deeper analysis along the same lines and compare with other reputable journals.

4.1 Shortening duration of editorial assessment

We noted that shorter processing times (below 40 days) became prevalent, starting from volume 80 of February 2021 (Fig. 3). Statistics on the editorial assessment duration (Tab. 3) show a 5-fold decrease in average processing time and a 6-fold decrease in median time when comparing the volumes of 2018–2020 and the volumes of early 2021.

Table 3 Statistics on the editorial assessment duration (in days) for 3 periods of *Microprocessors and Microsystems*.

Period	Volumes	<i>N</i>	Min	Avg	StdDev	Med	Max
2018–2020	56–79	579	19	202	157	161	1024
Early 2020	74–77	174	33	148	130	108	919
Early 2021	80–83	499	15	42	82	25	1206

4.2 Quicker editorial assessment and over-representation of some author countries

Out of 404 papers accepted in less than 30 days after submission, 394 papers (97.5%) have authors with affiliations in (mainland) China. Out of 615 papers of which editorial processing time exceeded 40 days, 58 papers (9.5%) only have authors with affiliations in (mainland) China. This tenfold imbalance suggests a differentiated processing of papers affiliated to China characterised by shorter peer-review duration.

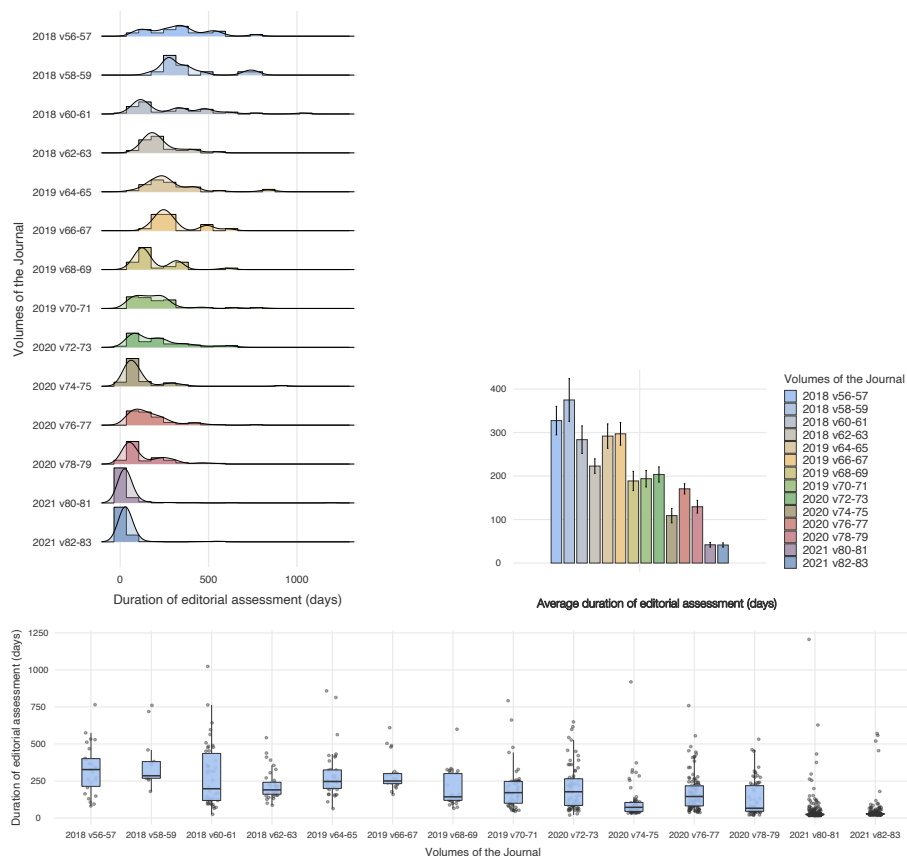


Fig. 3 Editorial assessment at *Microprocessors and Microsystems*: duration in days elapsed from submission to acceptance of the 1,078 articles published in volumes 56–83 issued between February 2018 and June 2021. The same data are presented with three complementary visualisations. The volumes of early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

4.3 Blocks of similar editorial timelines

Skimming through the table of contents, we observed that some papers share identical submission/revision/acceptance dates, which is unusual. This might suggest editorial overload. We thus aimed to identify these blocks of articles and the magnitude of this phenomenon.

Given a triple of dates (x, y, z) we define a *block* of papers characterised by this triple as follows: a paper belongs to the block if its submission date is either x or $x + 1$, its revision date is either y or $y + 1$ and its acceptance date is either z or $z + 1$.

We identified 111 (overlapping) blocks consisting of 10 or more papers, and 40 blocks consisting of 20 or more papers (See [Appendix](#)). Let us discuss two blocks whose publications appeared in special issues of the journal:

- The block generated by dates (November 22, 2020; December 9, 2020; December 14, 2020) consists of 30 papers:
 - 23 belong to the *Special Issue on Embedded Processors*,
 - 2 belong to the *Special Issue on Signal Processing*,
 - 1 belongs to the *Special Issue on Internet of People*,
 - 4 are regular papers.
- The block generated by dates (November 10, 2020; November 24, 2020; November 30, 2020) consists of 24 papers:
 - 16 of which belong to the *Special Issue on Embedded Processors*,
 - 8 are regular papers.

These examples show that a single block may contain papers from different special issues as well as regular papers. The special issues of *Microprocessors and Microsystems* are listed online⁹ with the mention ‘Edited by’ followed by the names of the persons in charge. A few special issues were introduced with a preface¹⁰ where editors present the topics and review process. The four special issues featuring papers from the two aforementioned blocks are not mentioned in the list of special issues. We also failed to find any preface to these special issues.

We were not able to propose a satisfactory explanation to this phenomenon within the bounds of a normal editorial process.

4.4 Discussion

The observed shortening time between submission and acceptance may reflect poor or deficient editorial assessment. Meanwhile, we noted at least two retractions¹¹ in *Microprocessors and Microsystems* for text duplication, indicating that the journal responds to integrity concerns at least in some cases. The closing of these two retraction notices are similar (only difference: the wording ‘severe abuse’ or ‘misuse’) and read as:

“As such this article represents a (misuse | severe abuse) of the scientific publishing system. The scientific community takes a very strong view on this matter and apologies are offered to readers of the journal that this was not detected during the submission process.”

While the suspected low editorial standards may explain how texts with tortured phrases got published, the process by which those tortured phrases were coined is quite mysterious. It seems improbable, for any skilled scientist, to use a non-standard terminology to refer to well-known concepts in one’s field. In addition, when authors are able to cite the literature that uses the standard terminology, it is unexpected that they switch to a tortured version of the terminology in their own manuscripts.

⁹ <https://www.sciencedirect.com/journal/microprocessors-and-microsystems/special-issues> and <https://www.journals.elsevier.com/microprocessors-and-microsystems/special-issues>

¹⁰ e.g., doi:10.1016/j.micpro.2020.103236 and doi:10.1016/j.micpro.2020.103187

¹¹ doi:10.1016/j.micpro.2020.103229 and doi:10.1016/j.micpro.2017.11.007

Our hypothesis is that the observed tortured phrases were coined by *misused* natural language processing (NLP) tools: automatic translation, automatic re-writing or even automatic generation of text. Today, the vast majority of these tools relies on advanced language models. In the next section we investigate a way to detect the use of such models.

5 Abstracts with high Generative Pre-Training (GPT) detector score

Advanced NLP models are now core building blocks for any natural language-related task: translation, information retrieval, classification, named-entity recognition, text generation, and so on. Regarding text generation, detectors of synthetic texts have been released. Automatic detection of computer generated text has already drawn attention in the past, see for example (Amancio, 2015; Cabanac & Labbé, in press; Dalkilic, Clark, Costello, & Radivojac, 2006; Labbé, Labbé, & Portet, 2016). This section focuses on one of the most recent detectors.

5.1 GPT and the GPT-2 Output Detector

The OpenAI company has released several advanced language models: Generative Pre-training (GPT, Radford, Narasimhan, Salimans, & Sutskever, 2018), Generative Pre-trained Transformer 2 (GPT-2, Solaiman, Clark, & Brundage, 2019), and GPT-3 (Brown et al., 2020). The generative power of these models has been extensively discussed:

- “Humans find GPT-2 outputs convincing. Our partners at Cornell University surveyed people to assign GPT-2 text a credibility score across model sizes.” (Solaiman, Clark, & Brundage, 2019)
- “We’ve seen no strong evidence of misuse so far. While we’ve seen some discussion around GPT-2’s potential to augment high-volume/low-yield operations like spam and phishing, we haven’t seen evidence of writing code, documentation, or instances of misuse. We think synthetic text generators have a higher chance of being misused if their outputs become more reliable and coherent. We acknowledge that we cannot be aware of all threats, and that motivated actors can replicate language models without model release.” (Solaiman, Clark, & Brundage, 2019)
- “With its apparent ability to artificially read and write, GPT-3 is perhaps different from other forms of AI, in that writing seems more fluid, open-ended, and creative than examples of AI that can beat people in a game or classify an image” (“New chapter in intelligence writing [Editorial]”, 2020)
- There is anecdotal evidence¹² that GPT-2 was re-trained on Pubmed abstracts to generate scientific texts (Lang, 2019).

A report from OpenAI discusses the ability of humans to differentiate between genuine texts and texts generated with GPT-2. It also presents and evaluates different versions of classifiers aiming to detect synthetic text and claims “Our classifier is able to detect 1.5 billion parameter GPT-2-generated text with approximately 95% accuracy” (Solaiman, Brundage, et al., 2019, p. 10). Unfortunately, the report does not provide any clue about precision and

¹² See writemeanabstract.com and <https://twitter.com/DrJHoward/status/1188130869183156231>

recall (i.e., false positive and false negative rates). Nevertheless, several versions of GPT-2 detectors are provided along with the generators so to flag synthetic texts. One is based on RoBERTa (Liu et al., 2019) and available as a website called ‘GPT-2 Output Detector Demo.’ This detector estimates a ‘fake’ score for a text given as input, reflecting the probability the text was generated. This prediction comes with a caveat: ‘The results start to get reliable after around 50 tokens.’

How GPT-2 relates to tortured phrases? Some non-native English authors write in their mother tongue and then translate into English using a translation service, such as DeepL or Google Translate.¹³ We hypothesised that observed tortured phrases would result from advanced language models: either through uncorrected translations or through text generation. Using a sample of texts in French, we checked the GPT-2 detector score before and after translation into English. The automatically translated results were clearly marked as ‘fake.’ This suggests that the GPT-2 detector flags text generated using GPT-2 and synthetic texts from other sources. If true, the GPT-2 detector may prove useful to flag questionable papers. This we investigate in the next section.

5.2 Datasets for evaluation

First, we retrieved abstracts for all full-length articles from volumes 80–83 of *Microprocessors and Microsystems* that were processed in less than 30 days. We thus obtained a set of 389 articles, which we call the *experimental set*. Given our earlier observations about editorial timelines, it is natural to expect articles from this set to be ‘probably questionable.’ Table 4 shows a breakdown of the 389 articles by special issue; regular papers are accounted for separately.

Having run the RoBERTa base GPT detector¹⁴ against all abstracts of the articles in the experimental set, we observed a prevalence of high GPT detector scores, see Tab. 5. Then we proceeded to assemble control sets to pursue the following goals:

- Answering the question, “do abstracts from *Microprocessors and Microsystems* exhibit a higher prevalence of articles with greater GPT detector scores compared to other sets of articles?”
- Finding possible explanations for prevalence of high GPT detector scores, other than use of GPT. The GPT detector may be sensitive to output of other advanced language models, at the basis of automatic translation or cross-language (self-)plagiarism.

Five control sets were created, each consisting of 50 samples except for the last one:

- (A) The abstracts of 50 most recent (by acceptance date) articles published in volumes 57–79 and processed in 41 days or more. We expected this set to represent “least concerning articles” from *Microprocessors and Microsystems*.
- (B) The abstracts of the 50 most recent articles in a selected set of SIAM journals, conditioned that the full text of the article contain terms among: IoT, wireless, sensor, sensors, deep learning, neural network, and neural networks.
- (C) The same set of abstracts as (B), but translated to Chinese and then back to English using Google Translate.

¹³ See <https://www.deepl.com/translator> and <https://translate.google.com>

¹⁴ <https://github.com/openai/gpt-2-output-dataset/tree/master/detector>

- (D) 50 Chinese-language abstracts from *Wireless Internet Technology*¹⁵ translated to English using Google Translate. The abstracts were selected at random from volumes 1/2021 and 2/2021. Before sampling, we excluded 3 abstracts from Volume 1/2021 that appeared to be advertisements of other journals.
- (E) 139,236 abstracts from randomly-selected articles published in 2021 by Elsevier. We retrieved this sample from the Web of Science on May 21, 2021 with query PY=2021 AND PUBL="Elsevier" AND DT="Article" AND LA="English" run on the *Science*, *Social Science*, *Art and Humanities*, and *Emerging Sources* citation indexes.

The control set (A) was chosen to represent the content of *Microprocessors and Microsystems* before the apparent change in the journal's operation mode.

The control set (B) was expected to represent high-quality, well-written, and thoroughly proofread articles. We infer these traits from the reputation of the society that publishes the journals. In addition, by selecting articles with certain terms we aim to ensure similarity with the experimental set by topic.

The control sets (C) and (D) were created in order to emulate a situation in which an English-language paper is prepared using automated translation from some other language.

The control set (E) was created to reflect a large proportion of the articles Elsevier published in early 2021 irrespective of the journals and scientific fields.

Table 4 Articles of the Experimental set: breakdown by Special Issue.

Headings	Articles vol. 80–83		Share (%)
	editorial assessment <30d (Exp)	all articles	
Special issue on Signal Processing	155	176	88.1
Special issue on Internet of People	98	102	96.1
Special issue on Embedded Processors	74	84	88.1
Regular Papers	49	83	59.0
Special issue on AI-SIGNAL PROCESSING	12	18	66.7
Special issue on CyberSECHARD2019	1	4	25.0
Grand Total	389	467	83.3

5.3 Results and analysis

The evaluation of abstracts from experimental and control sets against the RoBERTa base GPT detector is given in Tab. 5.

For the further analysis it is important to stress its main limitation: we assume that GPT scores in each considered case are sampled independently from a distribution related to the case. We consider this assumption reasonable within our setup.

Each of the six samples yields an empirical distribution function which we denote by F_{exp} , F_A , F_B , F_C , F_D and F_E for the experimental set and control sets A, B, C, D and E, respectively. A confidence band is then constructed around each empirical distribution function using the Dvoretzky–Kiefer–Wolfowitz inequality, see (Dvoretzky, Kiefer, & Wolfowitz, 1956) for the original work and (Massart, 1990) for the inequality with sharp constants; an exposition is also available in Section II of (Learned-Miller & DeStefano, 2008).

¹⁵ <https://wap.cnki.net/touch/web/Journal/Index/WXHK.html>

Table 5 Distribution (in %) of GPT detection scores by RoBERTa base. Rounded values may add up to greater than 100.0.

Score	Experiment Set ($N = 389$)	Control A ($N = 50$)	Control B ($N = 50$)	Control C ($N = 50$)	Control D ($N = 50$)	Control E ($N = 139, 236$)
[0.0, 0.1[8.5	78.0	90.0	56.0	28.0	89.9
[0.1, 0.2[1.5	6.0	6.0	4.0	12.0	2.0
[0.2, 0.3[0.8	0.0	2.0	2.0	8.0	1.1
[0.3, 0.4[0.5	0.0	0.0	10.0	4.0	0.8
[0.4, 0.5[0.5	0.0	0.0	0.0	0.0	0.7
[0.5, 0.6[1.5	0.0	0.0	2.0	2.0	0.6
[0.6, 0.7[1.5	0.0	2.0	2.0	0.0	0.6
[0.7, 0.8[2.1	0.0	0.0	2.0	6.0	0.6
[0.8, 0.9[2.1	0.0	0.0	2.0	4.0	0.8
[0.9, 1.0]	81.0	16.0	0.0	20.0	36.0	3.0
Sum	100.0	100.0	100.0	100.0	100.0	100.0

$$\Pr\left(\sup_x |F_{emp}(x) - F(x)| > \varepsilon\right) \leq 2\exp(-2N\varepsilon^2), \quad (1)$$

where N is the sample size, F_{emp} is the empirical distribution function, F is the cumulative function of the underlying distribution. Adjusting the value of ε will allow to control the right-hand side of (1) as follows:

$$\alpha = 2\exp(-2N\varepsilon^2) \Leftrightarrow \varepsilon = \sqrt{\frac{\ln(2/\alpha)}{2N}}.$$

We chose $\alpha = \frac{1}{120}$ so that with 95% ($0.95 = 1 - 6 \cdot \frac{1}{120}$) confidence all cumulative distribution functions for the 6 considered cases lie within their respective confidence bands. This approach is designed to account for multiple comparisons of the experimental case with control cases. For the distribution sizes at hand, i.e. $N = 389$, $N = 50$ and $N = 139, 236$, the half-width of the respective confidence bands are $\varepsilon \approx 0.084$, $\varepsilon \approx 0.234$ and $\varepsilon \approx 0.004$, respectively.

In Fig. 4 the confidence band for the experimental case is plotted against each confidence band for control cases. We hereby see that any of the cut-offs 0.3, 0.4, ..., 0.9 distinguishes the experimental case from all control cases in the sense that the scores in the experimental case occur above the selected cut-off significantly more often than the scores for the control cases do so. With the exception of Control D (abstracts from a Chinese journal translated to English), the same holds for the cut-offs 0.1 and 0.2. Under the design of our comparison, it is undecided whether the cut-offs 0.1 and 0.2 draw distinctions between the experimental set and the control set D; we hypothesise that the small size of set D does not provide sufficient power for comparison.

Table 6 reveals that several journals published articles with abstracts having a 70% or higher GPT detector score. The 70% threshold was selected because it belongs to the flat part of the cumulative distribution function of Control set E. Let us stress that the concentration of articles with high GPT scores is outstanding in *Microprocessors and Microsystems* with 72.1% compared to 13.6% maximum in the other journals tabulated. The column ‘Number of articles’ shows that many journals published papers with abstracts featuring a high GPT score. While a high GPT score for an abstract does not necessarily indicate flaws in an individual paper, high concentrations of such articles in certain venues invites a further assessment of this phenomenon.

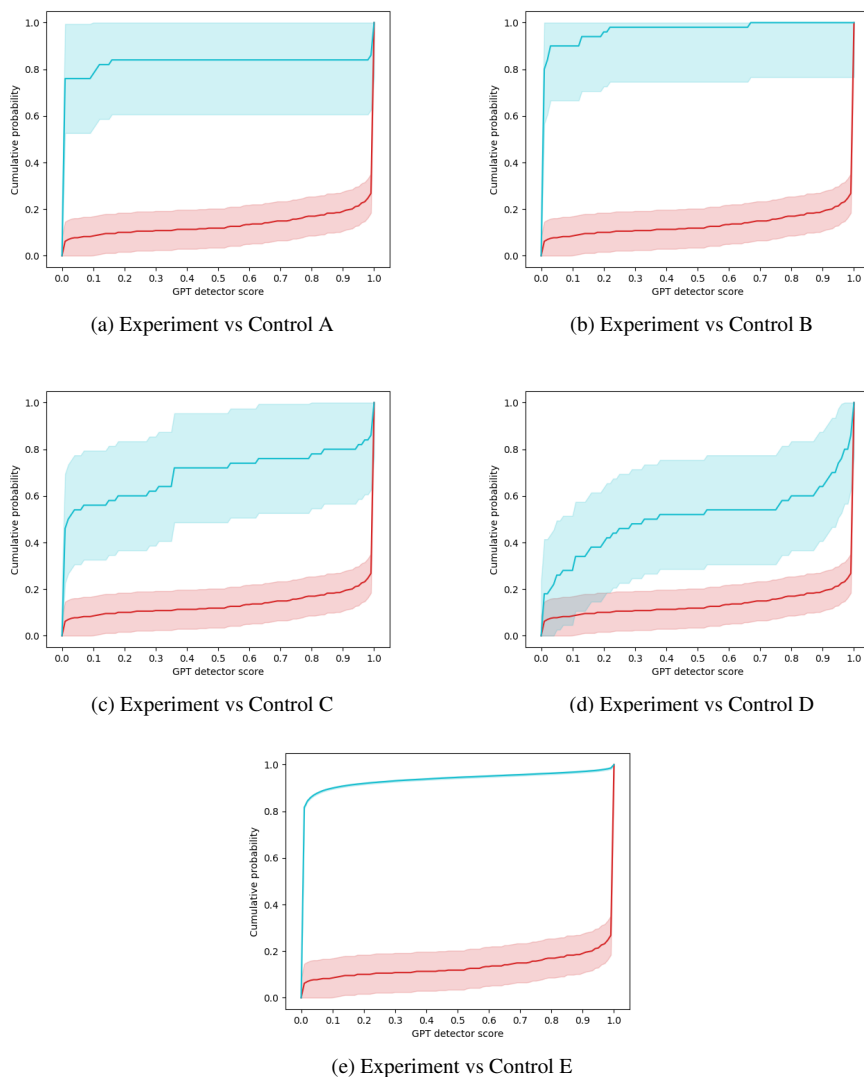


Fig. 4 Cumulative distribution functions for the Experiment set (red line) vs Control sets A–E (blue line) with confidence bands (Tab. 5).

The concentration of abstracts with a high GPT detector score in *Microprocessors and Microsystems* (Experimental Set) is intriguing. Nonetheless, texts flagged as synthetic by the GPT detector might be scientifically sound. We visually examined several publications from this journal to go beyond automatic screening. The next section reports critical flaws we found in several papers, including nonsensical text featuring tortured phrases, plagiarised text, and image theft. We believe these publications should be considered for retraction as they “represent a severe abuse of the scientific publishing system”, as quoted in the retraction notices reproduced in Sect. 4.4.

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
<i>Renew. Energy</i>	93.1	39	5.4
<i>J. Comput. Appl. Math.</i>	92.7	39	13.6
<i>Fuel</i>	91.6	39	3.6
<i>Constr. Build. Mater.</i>	93.5	39	3.5
<i>Spectroc. Acta Pt. A</i>	91.5	38	7.3
<i>Energy</i>	93.9	34	4.1
<i>Bioresour. Technol.</i>	92.0	34	6.5
<i>Food Chem.</i>	90.8	33	2.6
<i>Powder Technol.</i>	94.5	32	8.7
<i>Measurement</i>	94.0	31	6.1
<i>Carbohydr. Polym.</i>	90.1	31	5.3
<i>J. Differ. Equ.</i>	89.1	27	11.2
<i>Electrochim. Acta</i>	89.8	27	5.5
<i>Alex. Eng. J.</i>	92.6	26	9.8
<i>Opt. Laser Technol.</i>	90.7	25	8.7
<i>Neurosci. Lett.</i>	94.6	25	11.3
<i>J. Math. Anal. Appl.</i>	89.3	25	7.5
<i>Ecotox. Environ. Safe.</i>	92.2	25	4.4
<i>Environ. Pollut.</i>	91.2	25	3.4
<i>Carbon</i>	92.8	25	6.6

6 Critical flaws found in questionable and problematic publications: individual cases

All the above quantitative observations suggest that certain editorial processes in several venues were (and might still be) arranged in a non-conventional manner. In order to test this hypothesis, we analysed several individual papers from the journal *Microprocessors and Microsystems*. For each case presented in this section, we expose various flaws that, in our opinion, are unacceptable in published scientific literature. Our observations include:

- reuse of text and / or images without acknowledgement;
- references to non-existing literature;
- references to non-existing internal entities of the paper (e.g., theorems and variables in formulas);
- sentences for which we failed to infer any meaning.

Excerpts from each case are reported with a score computed by the [GPT-2 Output Detector](#) for which “The results start to get reliable after around 50 tokens.” We searched [Google Images](#) — either via the ‘Search by image’ feature or by typing in characteristic keywords — for potential earlier occurrences of selected images that appeared most suspect to us (e.g., irrelevant, of poor visual quality) in the papers we inspected. Note that we did not perform this image screening systematically. As of July 8, 2021 there were no citations for the six cases except for Case 5 and Case 6 with one citation each.

While mentioning individual papers, we explicitly refrain from including them in our list of references not to distort the scholarly record. None of the cases we discuss in the following sections had been reported to PubPeer ([Barbour & Stell, 2020](#)). We posted a PubPeer comment for each to trigger discussions.

Our purpose is not to blame individual authors but to trigger the necessary investigation to be conducted by editors and publishers.

6.1 Case 1: Unacknowledged (mis)use of a water leak detector description from elsewhere

▷ *Real time monitoring of medical images and nursing intervention after heart valve replacement*, Published in volume 82 of April 2021, [doi:10.1016/j.micpro.2020.103766](https://doi.org/10.1016/j.micpro.2020.103766).

The English is hard to understand and the meaning of some sentences is quite difficult to infer. For example, the section *literature survey* starting on the first page reads as follows:

Case 1, *literature survey* – GPT detector score: 99.98%

A **pamphlet of sickness** or harmed heart valves, ailment, or passing is one of the world’s significant reasons. **Accessible medicines** for patients with a heart valve **are abused**; however, to fix the valve because the fix is incredible, **it have** to supplant a heart valve in the most genuine cases.

Figure 5 shows a figure and its caption that are mostly irrelevant to each other. In fact, it appears that both the figure and the corresponding paragraph come from <https://www.edn.com/water-leak-detector-uses-9v-batteries/>, where a logical diagram for a water leak detector is presented. The original text has been heavily modified making it hard to understand.

Case 1 – GPT detector score: 88.22%

Fig. 4 shows this design, Image Detection Sensor built-in 1.2V reference Maxim, circuit MAX934, use integrated the four comparators of ultra-low power consumption. **Power consumption of the chip is about 6 μ A.** IC1A, R1, and R2 provides a water leak detection. R1 is water detector may be a two bare copper wire wound around the sponge. **R1 is, because the sponge is dry, when the left output of IC1A is high, you have a high impedance.** Circuit detects a water leakage, the value of R1 is less than one hundred, reduced to several kilo-ohms, and it is the low output of the force IC1A. Through D1, it will make the output of high IC1B.

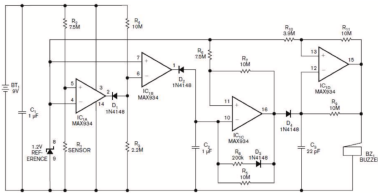


Fig. 4. Image Detection Sensor Using Intervention after Heart Valve Replacement

(a) Figure in Case 1 and its caption.

Case 1's source – GPT detector score: 3.53%

The design uses Maxim Integrated Circuits' MAX934, an ultra-low-power quad comparator with a built-in 1.2V reference. **The chip uses about 6 μ A.** IC1A, R1, and R2 provide water-leakage detection. R1 is the water probe, which can be two bare copper wires wrapped in a sponge. **R1 has high impedance when the sponge is dry, so IC1A's output stays high.** Once the circuit detects the water leak, R1's value decreases to less than a few hundred kilohms, which forces IC1A's output low. Through D1, it makes the output of IC1B high.

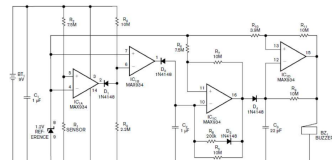


Figure 1 IC1A and IC1B determine the conditions for sounding the buzzer, and IC1C oscillates to provide the trigger for IC1D.

(b) Original figure and its caption.

Fig. 5 Case 1 (a) reusing without acknowledgement an original image (b) taken from <https://www.edn.com/water-leak-detector-uses-9v-batteries/>.

Moreover, Fig. 3 of Case 1 (not shown here) is identical to Fig. 7 (not shown here, caption: 'Magnetic resonance imaging in prosthetic heart valves...') published in a 2015 article [doi:10.1161/circimaging.115.003703](https://doi.org/10.1161/circimaging.115.003703) with no visible acknowledgement to the original source of the image.

6.2 Case 2: Image reuse

▷ *Case 2.1: Computer aided medical system design and clinical nursing intervention for infantile pancreatitis*, Published in volume 81 of March 2021, [doi:10.1016/j.micpro.2020.103761](https://doi.org/10.1016/j.micpro.2020.103761).

▷ *Case 2.2: Big Data Prediction of Sports Injury Based on Random Forest Algorithm and Computer Simulation*, Not included in a volume, online since January 2021, [doi:10.1016/j.micpro.2021.104002](https://doi.org/10.1016/j.micpro.2021.104002).

Case 2.2 contains the following tortured phrases (Tab. 1): *irregular timberland* (in lieu of *random forest*) and *innocent Bayes* (in lieu of *naïve Bayes*).

These two papers share a common image (without any mention to each other). This figure features an unexpected Spanish-language annotation on one of the blocks. It has been obtained by cropping of the figure 5.5 on page 136) of:

A. Ramírez Agundis, *Diseño y experimentación de un cuantizador vectorial hardware basado en redes neuronales para un sistema de codificación de video*, Doctoral thesis, Politecnica de Valencia, 2008. doi:10.4995/Thesis/10251/3444

The description of the image from each of the two questionable papers, as well as the description of the original image from the thesis by Ramírez Agundis are below. We also reproduce the images themselves in Fig. 6. A notable feature is that the description of the image in Case 2.2 has a relatively low GPT detector score.

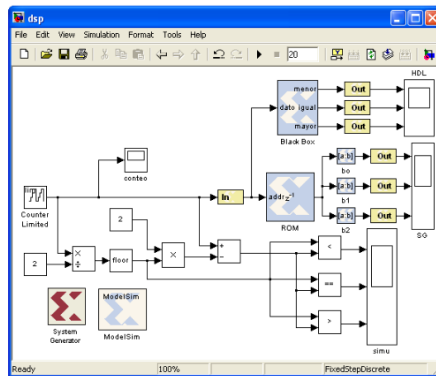
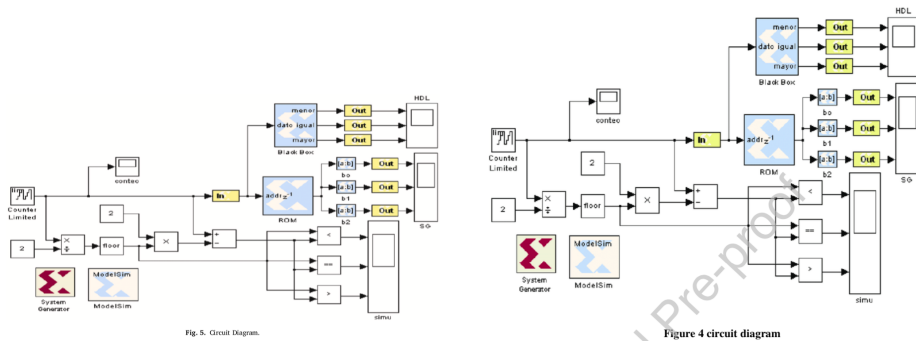


Fig. 5.5: Diagrama a bloques para codiseño Simulink-SG

(a) Original image and caption in Spanish



(b) Case 2.1 image and caption

(c) Case 2.2 image and caption

Fig. 6 Original image from a doctoral thesis (a) and its cropped versions in Case 2.1 (b) and Case 2.2 (c).

Description of the original image – GPT detector score 0.02%

Como es bien sabido, Simulink es un entorno gráfico que permite el diseño y simulación de modelos usando una metodología basada en diagramas de bloques. Proporciona además bibliotecas de bloques para las tareas comunes de procesamiento en muy diversas áreas y, lo que es importante para el codiseño, crear nuevas bibliotecas e incorporar otras proporcionadas por terceras partes, en este caso las que suministra el fabricante de los dispositivos FPGA (por lo que se refiere a Xilinx, a través de System Generator) y las del fabricante de la placa donde está alojada la FPGA. La Fig. 5.5 muestra un ejemplo que incluye bloques de los tres tipos.

Description of the image in Case 2.1 – GPT detector score 99.98%

Fig. 5 shown as if all is considered to be completed, after the hunter and viral contamination, T-gloomy period based on the average cycle-specific clinical course staged the febrile period, step-down stage, hypourinary stage diuretic phase. It can be improved—an essential part of all stages of kidney contribution.

Description of the image in Case 2.2 – GPT detector score 36.75%

Figure 4 the following stage expects us to change from likelihood to chances. The pre-test changes can be determined utilizing the recently sketched out condition (see area "Relative dangers and chances proportions") or determined utilizing the pre-test likelihood.

Additionally, we show other samples of questionable text in Cases 2.1 and 2.2, with some irregularities highlighted.

Case 2.1, *related work* – GPT detector score 99.98%

Gadget embedded clinical tools are typically coupled to locate, visualize and test the disease [1]. With such a device, including busy period and **most manifested people: meters above sea level, cardiovascular screen, screen blood glucose**, the ECG (Electro Cardio Gram), X-ray imaging, the MRI (Magnetic Resonance Imaging). CT (Computed Tomography) and PET (Positron Emission Tomography). (After the elimination of human formalism) the installation size and cost associated with clinical applications increased demand for corrective fully computerized gadgets due to the **dynamic need to squeeze it** is to promote the development control framework.

Case 2.2, *conclusion* – GPT detector score 99.95%

Competitor's exhibition to anticipate, recommend the sequential request of elective forecast techniques. The proposed innovation is an altered form of the group of the move learning model. **Utilizing many arranged history neural organizations**, searching the principle focus with the proper boundaries, and spotting the best model with at least wellness work in the initial set. **To assess the proposed expectation's greatness**, competitors' informational execution index around the globe has been utilized.

6.3 Case 3: Circuits Today heart rate monitor presented as something else

▷ *Computer aided intelligent medical system and nursing of breast surgery infection*, Published in volume 81 of March 2021, doi:10.1016/j.micpro.2020.103769.

Figure 2 in Case 3 contains a clear indication of its source (www.circuitstoday.com) and is most probably reused from <https://todayscircuits.wordpress.com/2014/06/26/tc-heart-rate>

[-monitor-using-8051/](#). It seems to picture a heart rate monitor while the reference cited in the caption (number [13] in its reference section) is about *Channel attention module with multiscale grid average pooling for breast cancer segmentation in an ultrasound image*. The reference to <https://www.circuitstoday.com/heart-rate-monitor-using-8051>, is consistent with older versions of that page (namely, those before June 2, 2016) — see https://web.archive.org/web/*/https://www.circuitstoday.com/heart-rate-monitor-using-8051.

The *breast surgery dataset* section contains the following text:

Case 3, *breast surgery dataset* – GPT detector score 99.87%

It also used a built-in case-control design. From March 1, 2012, to May 31, 2019, **all breast cancer surgeries are routine**. After discharge, the surgeon has also evaluated three times **more than 30 days of the patient week**.

6.4 Case 4: Citations to non-existent literature

▷ *Blockchain financial development based on FPGA and Convolutional Neural Network*, Not included in a volume, online since November 2020, doi:10.1016/j.micpro.2020.103492.

This article contains the following tortured phrases (Tab. 1): *profound neural organization* (in lieu of *deep neural network*), *fake neural organization* (in lieu of *artificial neural network*), *counterfeit neural organization* (in lieu of *artificial neural network*), *human-made consciousness* (in lieu of *artificial intelligence*).

The reference list contains non-existent or unidentifiable items. The hyperlinks provided in the pdf (and reproduced below) are either broken or leading to unrelated publications:

Case 4 reference section – GPT detector score 97.69%

- [4]: T.D. Chaudhry, Gauche, to decipher the forecast of the instability of the Indian financial exchange, utilizes a **counterfeit neural organization** with various information sources and yields, *J. Estimation* 120 (8) (2016) 7–15. **It applies to.**
- [5]: H. Moncada, M.H. Monhada, M. Esfandiari, Utilizing a **counterfeit neural organization**, *J. Econ. Class Stock File Fig.* 21 (41) (2016) 89–93. **Fund will be logical.**
- [11]: W. Melody, W. Tune, W. Song, R.F. expectation, **ACM bilayer neural organization system transformer**. The executives, *Inf. Framework.* 7 (4) (2017) 1–17.
- [13]: G. Kaur's, J. Dahl, R.K. Ha, Used to foresee the mix with the BSE list in any event adjusted **OWA administrator ANFIS fluffy C-**, *science, Neuropsychol. Rev.* 122 (2016) 69–80.

Additionally, the related work section attributes ref. [4] to ‘Kiyoshi Erwang’ whereas the references section gives ‘T.D. Chaudhry, Gauche.’ We were unable to determine whether the identity of ‘Kiyoshi Erwang’ is real or not.

The *related work* section starts with the following text:

Case 4, *related work* – GPT detector score 98.60%

The strategies for **human-made consciousness** has been invited by an ever-increasing number of residents. The **human-made consciousness** technique that speaks to exploring a **neural organization** has been created at an exceptional rate [1]. **Such a business expectation, Assessment of scores, business misfortune forecast, these fields, for example vision and control framework, has been generally utilized** [2].

6.5 *Case 5: Referring to a theorem that is never introduced*

▷ *Ecological landscape planning and design based on the Internet of Things system and VR technology*, Not included in a volume, online since November 2020, doi:[10.1016/j.micpro.2020.103431](https://doi.org/10.1016/j.micpro.2020.103431).

This article contains the following tortured phrases (Tab. 1): *organization association* (in lieu of *network connection*), *information distribution center* (in lieu of *data warehouse*).

The introduction contains a reference to *Theorem 1.2*, which does not exist within the paper. Most variables in mathematical formulas across the paper are not introduced in any way, their meaning remain unclear from the context.

Figure 2 in the paper is identical to Figure 6 from doi:[10.1016/j.scib.2019.07.004](https://doi.org/10.1016/j.scib.2019.07.004). No acknowledgement for image reuse was provided.

The statement of author's research interests appears odd: "*His research interests include Chinese calligraphy and fine arts, and the comparison of Chinese and Western arts.*" The title of the paper is "*Ecological landscape planning and design based on the Internet of Things system and VR technology*".

An excerpt from the abstract is provided below as an example of language irregularities in the paper.

Case 5, abstract – GPT detector score: 99.96%

Intelligent, real-time, is **essential to low-cost**, the planning and design of a distributed ecosystem to understand, to manage the rapidly changing ecosystem. However, in the era of big data, most new technology, especially in remote areas of fragile ecosystems, **it is not introduced to the planning and design of the business ecosystem**. Innovative by using the Internet of things technology ecosystem in the smart device planning and design and control system in the development and isolated environment, to establish the eco-system of the prototype, **internet of things (IoT) technology it is introduced**.

6.6 *Case 6: Abstracts of other papers rewritten in a tortuous way*

▷ *New technology application in logistics industry based on machine learning and embedded network*, Published in volume 80 of February 2021, doi:[10.1016/j.micpro.2020.103596](https://doi.org/10.1016/j.micpro.2020.103596).

This article contains the following tortured phrases (Tab. 1): *organization association* (in lieu of *network connection*), *huge information* (in lieu of *big data*), *arbitrary timberland* (in lieu of *random forest*).

In the ‘Materials and method’ section, one can read “*FedEx and Uninterruptible Power Supply (UPS), has become two recipients and supporters of improving transport and coordination,*” the abbreviation UPS is obviously incorrect as it should be “United Parcel Service.”

The related work section seems to be a concatenation of automatically re-written abstracts. This text could result from a back and forth automatic translation or the output of an unspecified re-writing tool:

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.

Abstract of reference [12] in case 6 references section – GPT detector score: 0.22%

This study presents a comparison of different **deep learning methods** used for sentiment analysis in Twitter data. In this domain, deep learning (DL) techniques, which contribute at the same time to the solution of a wide range of problems, gained popularity among researchers. Particularly, two categories of neural networks are utilized, **convolutional neural networks (CNN)**, which are especially performant in the area of image processing and recurrent neural networks (RNN) which are applied with success in **natural language processing (NLP)** tasks. In this work we evaluate and compare ensembles and combinations of CNN and a category of RNN the long short-term memory (LSTM) networks.

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.

Case 6 text rewritten from its ref. [12] – GPT detector score: 96.79%

In this investigation, look at the different **profound learning strategies** for information notoriety examination of Twitter. Around there, simultaneously profound Deep learning (DL) innovation to add to the arrangement of a wide scope of issues has been invited by analysts. In particular, two classifications of neural organization uses typically picture preparing and **Neuro-Linguistic Programming (NLP)**, a neurstrategiesizatithe on **Cable News Network (CNN)** convolution is applied to the region of the repetitive neural organization Recurrent Neural Network (RNN) with the undertaking.

6.7 Case 7: Citing items missing from the reference list

▷ *Simulation of football based on PID controller and BP neural network*, Published in volume 81 of March 2021, doi:10.1016/j.micpro.2020.103695

While the reference list contains only 15 items, labelled [1] through [15], the section *Related works* contains citations to items [16] and [17], as below:

Case 7, citations to items [16] and [17] missing from the reference list – GPT detector score: 99.98%

Calculation surveys that anticipate football results Competitions dependent on **neural organizations**. Study the capacity to make forecasts dependent on broad variables chose to utilize the neural organization [16]. Because of test research on neural organization properties, the expectation precision accomplished. A productive and speedy approach **to circulatory strain's neural organization is utilized by football refs to help refs their actual quality and quality**. Partitioned into two classifications: Or can't be against the mediator. This neural organization created under Boy Lab Windows, and its cordial U.I. created under Visual Basic. The Neural Network Method utilizes the past stage official point at the World Cup to foresee the two groups' speed that won the football competition [17].

We further provide an excerpt from the *Introduction* to further highlight irregularities in grammar and vocabulary of the paper. It is remarkable that the GPT detector score of this fragment is very low. This kind of examples highlight the actual limitations of deep learning methods by questioning the lack of explanation regarding the computed results.

Case 7, *Introduction* – GPT detector score 1.50%

As the **climate changes**, the ball's situation comparative with the robot's case is continually evolving. Subsequently, it is essential to apply a control framework for robots **to peruse these dynamic ecological conditions**. One of them is the utilization of **dark coherent regulators** in conduct based control and football robot route frameworks.

7 Potential sources of problematic papers

This section discusses the sources we suspect are involved in churning problematic papers out: paper mills and Spinbot-like software.

7.1 Paper mills: Template-based massive production of papers

We suspect papers mills (Else & Van Noorden, 2021) to have produced part of the problematic papers we analysed. Several recurring features shared by most questionable papers from *Microprocessors and Microsystems* may indicate that they come from a single source:

- Similar composition of the papers consisting of five sections named (with slight variations) *Introduction*, *Related work*, *Materials and methods*, and *Results and discussion*, *Conclusion*. This composition is not so common for papers published before volume 80, or even for papers in volumes 80–83 with longer duration of editorial assessment.

- Most questionable papers that we inspected share the same typical set of colours used for diagrams: light blue, orange, grey, yellow, and blue. This suggests that the same software was used to prepare the papers. However, this feature is not common even for all questionable papers.
- We share a subjective impression that there is little variability in the way the images and tables are prepared. For instance, the use of block diagrams is outstandingly common. We expect that presence of non-standard images in these papers will often indicate unacknowledged reuse.

In addition, let us note that the observed changes in the operational mode of the journal, most notably, the increased output, bear some resemblance to how *hijacked journals* operate (Abalkina, in press; Jalalian & Dadkhah, 2015).

7.2 Spinbot: Article Spinning, Text Rewriting, Content Creation Tool

Searching the web for ‘text reformulation’ we stumbled upon spinbot.com, introduced as “a free, automatic article spinner that will rewrite human readable text into additional, intelligent, readable text.” The Wayback Machine of the Internet Archive has records of this website offering this service for a decade now.¹⁶ The [web page](http://spinbot.com) offers to ‘rewrite’ a text of up to 10,000 characters for free. No information is provided regarding the technology Spinbot uses, no computer code is available. Two [paid options](#) are proposed. First, a paid subscription allows customers to use Spinbot without ads or captchas for \$10 a month, \$50 for 6 months, or \$75 a year. Second, a developer can buy ‘Spin Credits’ that correspond to a number of API calls, prices ranging between \$5 for 1,000 credits and \$2,000 for 500,000 credits. Multiple websites such as paraphrasing-tool.com and free-article-spinner.com claim to be ‘powered by the Spinbot API.’ There is no personal information available about the designer of Spinbot yet the [Spinbot Blog](#) advertises the [Mr. Green Marketing, LLC](#) company based in Kansas City, USA.

Studies of academic writing use the term ‘spin’ to describe authors’ attempt to present a more positive description of a technique or drug in health sciences than the data actually support (Boutron et al., 2014; Boutron & Ravaud, 2018). What Spinbot performs, however, is less complex: it replaces words with synonyms. For instance, the text ‘big data’ is transformed into ‘enormous data’ or ‘huge data’ or ‘large data’ when running Spinbot multiple times. The term ‘artificial intelligence’ is spun as ‘counterfeit consciousness’ or ‘man-made brainpower’ or ‘computerized reasoning.’ Feeding the phrases in the ‘Correct wording expected’ column of Tab. 1 to Spinbot we were able to reproduce the associated ‘tortured phrases.’ Different executions of Spinbot with “The road to hell is paved with good intentions.” as input yielded:

- The road to hell is paved with good intentions.
- The way to damnation is cleared with sincere goals.
- The way to hellfire is cleared with honest goals.
- The way to hellfire is cleared with well meaning goals.
- The way to damnation is cleared with well meaning goals.

¹⁶ See the archive of January 28, 2011 at <https://web.archive.org/web/20110128/http://spinbot.com/>. Note that the feature called ‘Spin Any Language to Any Language’ once present is not available anymore.

8 Conclusion and Call for Action

We discovered a number of tortured phrases in the scientific literature, mainly in Computer Science.

We further studied one specific journal, *Microprocessors and Microsystems*, affected by the phenomenon. Our study revealed significant, likely questionable, changes in the journal's operational mode. These changes did not attract much attention, despite being given out by various hints, including:

- abrupt drop of the average / median duration of editorial assessment;
- abrupt surge in the number of articles accepted;
- acceptance of evidently synthetic texts;
- unexpected author affiliations and/or research interests (in authors' biographies) and/or research background outside of the scope of the venue (e.g., 'school of musicology' in a venue on microprocessors).

We specifically discussed the issues with 7 cases covering 8 papers that we are also reporting on PubPeer (Barbour & Stell, 2020). As of 17 June 2021, none of the 1,078 papers had been commented on PubPeer (See Appendix), suggesting that the issues we found went unnoticed.

No systematic screening of the papers containing tortured phrases has been performed to date. Nevertheless, we estimate *Microprocessors and Microsystems* accepted around 500 questionable articles: 389 papers with short duration of editorial assessment in volumes 80–83, plus additional papers not yet included in a volume. As of June 25, 2021 there were 225 such articles queued 'in press' which are 'accepted, peer reviewed articles that are not yet assigned to volumes/issues, but are citable using DOI.'

Our study revealed that multiple other venues also published papers with tortured phrases (Fig. 1) and abstracts with high GPT detector scores (Tab. 6). Tailoring the fingerprint–query approach used in (Cabanac & Labbé, in press) is a promising way to comb the literature for tortured phrases.¹⁷ Preliminary probes show that several thousands of papers with tortured phrases are indexed in major databases. While we managed to identify and retro-engineer several tortured phrases in Computer Science, other tortured phrases related to the concepts of other scientific fields are yet to be exposed.

All in all this work is a call to action and we therefore:

- Encourage other members of the scientific community to extend our findings. We also welcome any effort towards deeper case-by-case analysis of papers in various venues.
- Expect that the relevant parties (Elsevier, COPE, Clarivate Analytics, etc.) initiate an impartial, efficient, transparent and wide investigation, should our concerns be grounded. We believe that the irregularities we raised in this study may fall within the scope of the COPE's guideline/flowchart *Systematic manipulation of the publication process* (COPE Council, 2018).
- Suggest that researchers and, especially, publishers monitor the publishing ecosystem for various hints indicating unusual publication activities (see below). However, we emphasise that hints alone do not mean that misconduct is happening, therefore an analysis of individual papers should be performed in order to support or refute any concerns.

¹⁷ See the screening and assessment results for grammar 'tortured' at <https://www.irit.fr/~Guillaume.Cabanac/problematic-paper-screener>

We wish to broaden the discussion about whether software and natural language models are welcome to generate or modify scientific texts. It is of tremendous importance to provide the detection method that goes along. Such detection methods should be well characterised with regards to both false positives (type I error) and false negatives (type II error). The detection method should also provide a rationale for its decision in line with the current expectations for ‘explainable artificial intelligence.’

Attempts to automatically detect synthetic texts will benefit from open abstracts¹⁸ (Schiermeier, 2020) and their indexing by various academic search engines such as Dimensions (Herzog et al., 2020). Screening pipelines (e.g., Weissgerber et al., 2021) may include such detection software and run it ahead of peer review. While probably useful in short-to-medium term perspective, we fear that any screening initiative is likely to provoke an arms race.

We note, however, that peer review — or rather initial editorial screening — should have detected and filtered out the most blatant examples of synthetic texts; its failure to do so should be analysed.

In our strong opinion, the root of the problems discussed in this work is the notorious *publish or perish* atmosphere (Garfield, 1996) affecting both authors and publishers. This leads to blind counting and fuels production of uninteresting (and even nonsensical) publications.

Update as of July 12, 2021: *Retraction Watch* reports Elsevier issued Expressions of Concern for six special issues of *Microprocessors and Microsystems* (Marcus, 2021). It is not clear whether regular papers (see Tab. 4) will be assessed.

Appendix: List of supplementary materials

We release supplementary materials on Zenodo under [doi:10.5281/zenodo.5031935](https://doi.org/10.5281/zenodo.5031935) for transparency and reproducibility concerns.

Acknowledgements We thank PubPeer for providing a forum to the scientific community. We thank Digital Science for making Dimensions data available for scientometric research. A.M. would like to thank Maxim Panov (Skoltech) for help with initial reconnaissance of the subject. A.M. also acknowledges Mike Downes (independent researcher, Australia) who has already observed the phenomenon of recurring editorial timelines in predatory venues, although his post on the subject at <https://scholarlyoutlaws.com/> appears to be no longer accessible. We are grateful to the colleagues who provided constructive comments and feedback on a previous version of this preprint: Frédérique Bordignon, Ophélie Fraisier-Vannier, Willem Halffman, Vincent Larivière, and François Portet.

References

- Abalkina, A. (in press). Detecting a network of hijacked journals by its archive. *Scientometrics*. doi: [10.1007/s11192-021-04056-0](https://doi.org/10.1007/s11192-021-04056-0)
- Amancio, D. R. (2015). Comparing the topological properties of real and artificially generated scientific manuscripts. *Scientometrics*, 105(3), 1763–1779. doi: [10.1007/s11192-015-1637-z](https://doi.org/10.1007/s11192-015-1637-z)

¹⁸ <https://i40a.org>

- Barbour, B., & Stell, B. M. (2020). PubPeer: Scientific assessment without metrics. In M. Biagioli & A. Lippman (Eds.), *Gaming the metrics: Misconduct and manipulation in academic research* (chap. 11). MIT Press. doi: [10.7551/mitpress/11087.003.0015](https://doi.org/10.7551/mitpress/11087.003.0015)
- Beta Writer. (2019). *Lithium-Ion batteries: A machine-generated summary of current research*. Springer. (Machine-generated by Beta Writer 0.7 software developed at Goethe University Frankfurt) doi: [10.1007/978-3-030-16800-1](https://doi.org/10.1007/978-3-030-16800-1)
- Bohannon, J. (2013). Who's afraid of peer review? [News]. *Science*, 342(6154), 60–65. doi: [10.1126/science.342.6154.60](https://doi.org/10.1126/science.342.6154.60)
- Boutron, I., Altman, D. G., Hopewell, S., Vera-Badillo, F., Tannock, I., & Ravaud, P. (2014). Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: The SPIIN randomized controlled trial. *Journal of Clinical Oncology*, 32(36), 4120–4126. doi: [10.1200/jco.2014.56.7503](https://doi.org/10.1200/jco.2014.56.7503)
- Boutron, I., & Ravaud, P. (2018). Misrepresentation and distortion of research in biomedical literature. *Proceedings of the National Academy of Sciences*, 115(11), 2613–2619. doi: [10.1073/pnas.1711075115](https://doi.org/10.1073/pnas.1711075115)
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... Amodei, D. (2020). Language models are few-shot learners. *arXiv*. Retrieved from <http://arxiv.org/abs/2005.14165> (preprint)
- Cabanac, G., & Labbé, C. (in press). Prevalence of nonsensical algorithmically generated papers in the scientific literature. *Journal of the Association for Information Science and Technology*. doi: [10.1002/asi.24495](https://doi.org/10.1002/asi.24495)
- COPE Council (Ed.). (2018). *Systematic manipulation of the publication process* (Tech. Rep.). (Version 1) doi: [10.24318/cope.2019.2.23](https://doi.org/10.24318/cope.2019.2.23)
- Dalkilic, M. M., Clark, W. T., Costello, J. C., & Radivojac, P. (2006). Using compression to identify classes of inauthentic texts. In *Proceedings of the 2006 SIAM International Conference on Data Mining*. doi: [10.1137/1.9781611972764.69](https://doi.org/10.1137/1.9781611972764.69)
- Day, C. (2019). Here come the robot authors [Editorial]. *Physics Today*, 72(6), 8. doi: [10.1063/pt.3.4213](https://doi.org/10.1063/pt.3.4213)
- Dvoretzky, A., Kiefer, J., & Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, 27(3), 642–669. doi: [10.1214/aoms/1177728174](https://doi.org/10.1214/aoms/1177728174)
- Else, H., & Van Noorden, R. (2021). The fight against fake-paper factories that churn out sham science. *Nature*, 591(7851), 516–519. doi: [10.1038/d41586-021-00733-5](https://doi.org/10.1038/d41586-021-00733-5)
- Faulkes, Z. (Ed.). (2021). *Stinging the predators: A collection of papers that should never have been published*. (version 18) doi: [10.6084/m9.figshare.5248264](https://doi.org/10.6084/m9.figshare.5248264)
- Garfield, E. (1996). What is the primordial reference for the phrase 'Publish or Perish'? [Commentary]. *The Scientist*, 10(12), 11. Retrieved from [http://www.garfield.library.upenn.edu/commentaries/tsv10\(12\)p11y19960610.pdf](http://www.garfield.library.upenn.edu/commentaries/tsv10(12)p11y19960610.pdf)
- Hendricks, G., Tkaczyk, D., Lin, J., & Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies*, 1(1), 414–427. doi: [10.1162/qss_a_00022](https://doi.org/10.1162/qss_a_00022)
- Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies*, 1(1), 387–395. doi: [10.1162/qss_a_00020](https://doi.org/10.1162/qss_a_00020)
- Hutson, M. (2021). Robo-writers: the rise and risks of language-generating AI [News feature]. *Nature*, 591(7848), 22–25. doi: [10.1038/d41586-021-00530-0](https://doi.org/10.1038/d41586-021-00530-0)
- Jalalian, M., & Dadkhah, M. (2015). The full story of 90 hijacked journals from August 2011 to June 2015. *Geographica Pannonica*, 19(2), 73–87. doi: [10.5937/geopan1502073j](https://doi.org/10.5937/geopan1502073j)

- Labbé, C., Labbé, D., & Portet, F. (2016). Detection of computer-generated papers in scientific literature. In M. D. Esposti, E. G. Altmann, & F. Pachet (Eds.), *Creativity and universality in language* (pp. 123–141). Springer. doi: [10.1007/978-3-319-24403-7_8](https://doi.org/10.1007/978-3-319-24403-7_8)
- Lang, F. (2019, October 28). *OpenAI's GPT2 now writes scientific paper abstracts*. Interesting Engineering. Retrieved from <https://interestingengineering.com/openais-gpt2-now-writes-scientific-paper-abstracts>
- Learned-Miller, E., & DeStefano, J. (2008). A probabilistic upper bound on differential entropy. *IEEE Transactions on Information Theory*, *54*(11), 5223–5230. doi: [10.1109/tit.2008.929937](https://doi.org/10.1109/tit.2008.929937)
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., . . . Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv*. Retrieved from <http://arxiv.org/abs/1907.11692> (preprint)
- Mallapaty, S. (2020). China's research-misconduct rules target "paper mills" that churn out fake studies. *Nature*. doi: [10.1038/d41586-020-02445-8](https://doi.org/10.1038/d41586-020-02445-8)
- Marcus, A. (2021, July 12). *Elsevier says "integrity and rigor" of peer review for 400 papers fell "beneath the high standards expected"*. Retrieved from <https://retractionwatch.com/2021/07/12/elsevier-says-integrity-and-rigor-of-peer-review-for-400-papers-fell-beneath-the-high-standards-expected/> (Retraction Watch)
- Massart, P. (1990). The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Annals of Probability*, *18*(3), 1269–1283. doi: [10.1214/aop/1176990746](https://doi.org/10.1214/aop/1176990746)
- New chapter in intelligence writing [Editorial]. (2020). *Nature Machine Intelligence*, *2*(8), 419. doi: [10.1038/s42256-020-0223-0](https://doi.org/10.1038/s42256-020-0223-0)
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training* (Tech. Rep.). OpenAI. Retrieved from https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf
- Schiermeier, Q. (2020, October 12). Initiative pushes to make journal abstracts free to read in one place [News]. *Nature*. doi: [10.1038/d41586-020-02851-y](https://doi.org/10.1038/d41586-020-02851-y)
- Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*. doi: [10.1007/s11192-021-03948-5](https://doi.org/10.1007/s11192-021-03948-5)
- Solaiman, I., Brundage, M., Clark, J., Askell, A., Herbert-Voss, A., Wu, J., . . . Wang, J. (2019, November). *Release strategies and the social impacts of language models* (Tech. Rep.). OpenAI. Retrieved from <https://arxiv.org/abs/1908.09203> (arXiv preprint)
- Solaiman, I., Clark, J., & Brundage, M. (2019, November). *GPT-2: 1.5B release* (Tech. Rep.). OpenAI. Retrieved from <https://openai.com/blog/gpt-2-1-5b-release/>
- Van Noorden, R. (2021). Hundreds of gibberish papers still lurk in the scientific literature. *Nature*, *594*(7862), 160–161. doi: [10.1038/d41586-021-01436-7](https://doi.org/10.1038/d41586-021-01436-7)
- Visconti, G. (Ed.). (2021). *Climate, planetary and evolutionary sciences: A machine-generated literature overview*. Springer. doi: [10.1007/978-3-030-74713-8](https://doi.org/10.1007/978-3-030-74713-8)
- Weissgerber, T., Riedel, N., Kilicoglu, H., Labbé, C., Eckmann, P., ter Riet, G., . . . Bandrowski, A. (2021). Automated screening of COVID-19 preprints: Can we help authors to improve transparency and reproducibility? *Nature Medicine*, *27*(1), 6–7. doi: [10.1038/s41591-020-01203-7](https://doi.org/10.1038/s41591-020-01203-7)