

Touchy information and irregular esteem – on the problem of tortured phrases and possibly fake science

Petrus H. Potgieter^{a, b}  potgiph@unisa.ac.za

^a University of South Africa

^b Institute for Technology and Network Economics

An unscientific sample of reviewers for and readers of scientific journals have anecdotally reported to the author their frustration at the profusion of papers that:

- seem (at best) formulaic, and
- are written in odd English.

The content of these papers also suggests the extensive use – in an honest or dishonest way – of paraphrasing and automated writing or translation tools. One of the amusing attendant phenomena is tortured phrases (Else, 2021), the nonsensical thesaurus calques of well-known and established terminology. The following table gives a small sample.

Normal terminology	Tortured phrase
sensitive information	touchy information
random value	irregular esteem
deep neural network	profound neural organisation
more and more	extra and extra
big data	colossal information
back propagation algorithm	back spread calculation
number of features	quantity of highlights
data loss	data misfortune

An interested reader can easily find examples of all of the above by searching for the tortured phrases on Google Scholar. The paper by Deepa et al. (2021) is perhaps one of the most emblematic examples, as the following quotation shows.

Potgieter, P.H. (2022). Touchy information and irregular esteem – on the problem of tortured phrases and possibly fake science [Viewpoint]. *South African Computer Journal* 34(2), 149–151. <https://doi.org/10.18489/sacj.v34i2.1186>

Copyright © the author(s); published under a [Creative Commons NonCommercial 4.0 License](https://creativecommons.org/licenses/by-nc/4.0/) .

SACJ is a publication of the [South African Institute of Computer Scientists and Information Technologists](https://www.unisa.ac.za/).

ISSN 1015-7999 (print) ISSN 2313-7835 (online).

The encoder some portion of the organization diminishes the quantity of highlights starting with one concealed layer then onto the next, ideally bringing about a bunch of (pseudo-) symmetrical highlights with insignificant data misfortune, so the decoder part can recreate the info. The back spread calculation changes the auto encoder neuron loads toward the misfortune work negative inclination, limiting the distinction between the information and remade yield picture.

One cannot claim dishonesty in this case with any certainty since it is perfectly feasible that this tortured paragraph was produced by writing in a language other than English and then machine translating. Even writers that are relatively fluent in English (like the author) occasionally use paraphrasing tools to tidy up text but it seems unlikely:

- that the text quoted above was produced in this manner, and
- subsequently reviewed by an author who is familiar with the literature on the subject in the English language.

Some papers contain only the occasional tortured phrase such as “heterosexual structure of carbon” in Wu et al. (2020), reported in a letter to the journal by Teixeira da Silva (2021). Other papers however contain loads of nonsensical prose. The following example from the paper by Wang (2021), in the *Springer* journal *Microprocessors and Microsystems* that has been flagged by Else (2021) as containing many suspicious papers, is a good example.

Fig. 6 describes the system service by the different categories, ecosystem service based on regulatory service, cultural service, provisioning services, and disservices. ... The graph is based on the ISI network to use the long-term ecological and ecosystem services (S) to search for science. It has been included mention of ecosystem services, basically does not solve these problems.

Both Wang (2021) and Deepa et al. (2021) fit a template which can be observed in many published articles. It consists of:

- some anodyne and/or nonsensical prose with (almost) no references,
- followed by a relatively reasonable literature survey (which artificial intelligence engines are known to be skilled at producing),
- concluding anodyne and/or nonsensical methods and results with (almost) no references.

The incentives for publishing research papers of whatever quality are clear: professional advancement as well as direct financial incentives (in many cases). In South Africa, for example, the government pays a subsidy of well over \$10 000 for each publication in its list of “accredited” journals. That list includes *Microprocessors and Microsystems*. The sheer volume of publications produced by some authors (not working in laboratories) suggests that a degree

of mechanised production has come into play although this need be neither dishonest nor unethical. Governments and other institutions are promoting publications at almost any cost and those who react to these incentives should not be blamed for it.

The first research question is to formulate a relatively simple criterion for identifying papers with questionable content. Subsequently, we would like to gauge the extent of the problem through random sampling of publications in a subset (to be determined) of the extant published content. Barriers to gathering the data include

- the sheer size of the published scientific corpus,
- works being behind a paywall and having their text not indexed by search engines other than possibly Google Scholar,
- a lack of transparency as to the methods by which search engines select results, and
- citation and other bibliographic databases being likewise proprietary.

We believe that questions about the academic integrity of the scientific publishing environment are sufficiently serious for at least an exploratory investigation into the data to be fully warranted at this stage. This should be relevant for funders but especially for journal editors and the publishers as well as the readership of such journals.

DISCLAIMER

Views expressed are entirely personal ones of the author.

References

- Deepa, S., Thanammal, K. & Sujatha, S. (2021). Phishing website detection using novel features and machine learning approach. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(7), 2648–2653. <https://www.turcomat.org/index.php/turkbilmat/article/view/3638>
- Else, H. (2021). ‘Tortured phrases’ give away fabricated research papers. *Nature*, 596, 328–329. <https://doi.org/10.1038/D41586-021-02134-0>
- Teixeira da Silva, J. A. (2021). A tortured phrase claims heterosexuality of the carbon structure. *Results in Physics*, 30, 104842. <https://doi.org/10.1016/j.rinp.2021.104842>
- Wang, X. (2021). Research on inversion of ecosystem dynamics model parameters based on improved Neural Network algorithm. *Microprocessors and Microsystems*, 80, 103605. <https://doi.org/10.1016/j.micpro.2020.103605>
- Wu, T., Yao, M., Li, J., Li, M. & Long, M. (2020). First-principles prediction of the electronic property, carrier mobility and optical absorption in edge-modified pristine sawtooth penta-graphene nanoribbons (SSPGNRs). *Results in Physics*, 17, 103103. <https://doi.org/10.1016/J.RINP.2020.103103>