# Editor guide: Papermills, paraphrasing software, nonsensical papers, AI-generated papers, citation rings.

This document aims to help EPL Editors to detect some **emerging forms of publication malpractice**. It is not an exhaustive list of types of publication malpractice and tools to detect these. Please also refer to EPL's ethical policy at: https://www.epletters.net/ethical-policy/

## Papermills

Papermills are the process by which manufactured manuscripts are submitted to a journal for a fee on behalf of researchers with the purpose of providing an easy publication for them, or to offer authorship for sale (see this report https://publicationethics.org/sites/default/files/paper-mills-cope-stm-research-report.pdf). **Hallmarks** listed below do not appear in every case, and do not definitively determine if a paper has been generated by a papermill. In combination with one another, and depending on the context, we can assess the likelihood that a paper is papermill derived.

## Paraphrasing software

Software can be used by academics, and papermills, to generate nonsense papers by rewording often genuine research (https://spinbot.com/, https://quillbot.com/ etc). The original source could be an entire paper, or portions of different papers stitched together to create an article, which is then run through the software. The end product is a paper that evades plagiarism detection software, and contains peculiar phrases that have been termed 'tortured phrases' e.g. 'attractive reverberation' instead of the established term 'magnetic resonance'

## Automatic paper generators

A group of academics created a software in 2005 called Scigen for their own amusement, that would generate nonsense content (https://pdos.csail.mit.edu/archive/scigen/). Academics began using the software to generate content to submit to conferences and journals, sometimes to entrap these and expose flaws in the peer review process. The creation and use of similar technology has increased over the years.

## AI-generated papers

With the development of generative AI technology such as ChatGPT, the risk of receiving submissions of AI-generated papers is increased. AI-generated papers are generally free of grammatical and spelling errors, but lack personality, unique perspective and opinions.  If an article feels repetitive or formulaic, it may have been generated by a machine. AI detection

tools can be used to determine the likelihood that a paper was generated by an AI, e.g.
https://platform.openai.com/ai-text-classifier

(In table below: EO: Editorial Office, PO: Production Office)

| Papermill hallmarks | Resources available |
| --- | --- |
| **Title, abstract** | |
| Inappropriate use of "buzzwords" in article title: big data, internet of things, machine learning, artificial intelligence… | |
| A significant point made in the title and/or abstract is not discussed at any point in the remainder of the paper. | |
| **Materials and methods** | |
| Methods & materials section is overly vague, providing no real way for other researchers to reproduce. | |
| The method does not match the result at all. | |
| **Language** | |
| Repeat miss-spellings, e.g. logical instead of logarithmic. | |
| Tortured phrases, e.g. counterfeit consciousness instead of artificial intelligence. | |
| **Data, Figures, Images** | |
| Author unable to provide full raw data. | |
| Duplicate images, within the image, within the paper, or with other published papers. | |
| Fuzzy/pixelated images – suggests images have been taken from another source, especially if mismatch in the quality of different figures. | PO checks that figures are sufficiently good for online publication, if not requests alternative figures. |
| Author not able to send images in other/original formats/resolutions. | |

| | |
|---|---|
| The same or similar image/article structure is used across multiple active submissions by different authors. | |
| Unnatural background noise on image. | |
| Identical bar charts representing different experiments. | |
| Results 'too good to be true', very small error bars with low sample in a sample expected to be heterogenous. | |
| **Declarations** | |
| Ethics statement shows ethics approval coming from different institution compared to affiliations of authors. | |
| Ethics statement inappropriate for the study. | |
| No funding has been declared in studies where funding would be expected (large experimental collaboration…) | |
| **References** | |
| Irrelevant/fictional references. | |
| Repeated phrases in references (often same as phrase in title). | |
| Majority of references are old (e.g. the most recent was published 4 years before the submission of the work). | |
| Stretched references: e.g. J. K o, K. L ee, S. H ong. | |
| Significant citation padding (e.g. 9-22 for only one point.) | |
| **Authors' and reviewers' names, emails, affiliations** | |
| Authors institutions/departments don't match subject matter of paper. | |
| Author suggested reviewers – contain names of well-known people but not institutional email address (fake reviewer). | Google the email address |
| Email address contains name and numbers and numbers correspond to date of submission e.g. firstnamelastname201908@ | Google the email address |

| | |
|---|---|
| Similar/same email address used by different authors with different names. | |
| Different papers submitted by different authors where the corresponding authors all have the same e-mail address. | |
| When searching the corresponding author's email address in Google, other articles by the author are not in legitimate journals and/or focus on different unrelated topics. | Google the email address |
| The IP of emails sent by the authors (@institution.org) does not match their location and there is no sufficient explanation (if different, ask the authors to confirm if they are residing near their institution). | |
| Empty ORCID profiles | Check in ORCID |
| Author list contains non-existing email addresses. | Google the email address |
| **Process** | |
| Requested changes to the authorship post-submission, especially if the request is suspicious (addition of more than 1 or 2 authors) and if the request is received post-acceptance. Frequent changes to author list is very suspicious. | Pre-acceptance: EO sends requests to the Editor for approval Post-acceptance: PO sends change requests to the Editor for approval |
| Significant requested changes to the author proofs. Especially if they request to add new references, to their own work or a particular author(s) work. | PO does not accept requests affecting the scientific content of the paper. The paper is returned to the EO and to CE. €100 charge for large volume of changes may apply. |
| Similarly worded (and timed) email responses to request for info/data from apparently different groups/people. | |
| **Plagiarism** | |
| Crossref Similarity Check overall percentage is **too high**. | Crossref Similarity Check integrated in ScholarOne. Editor should check the overall |

| | similarity, and in doubt look at the full similarity report. |
|---|---|
| Crossref Similarity Check overall percentage is **too low**. Too low percentage may indicate paraphrasing. | Crossref Similarity Check integrated in ScholarOne. |
| Translated plagiarism | Check title/abstract/conclusion in some mainstream languages – Russian, Spanish |
| **Submission files** | |
| The author or title of the original file in the properties does not match the author or title of the article. | EO checks the details of the paper vs information entered by the author in Scholar One |
| Revision number of original submission file is very high, >50. The total editing time is very high, especially compared to the number of revisions. | |