



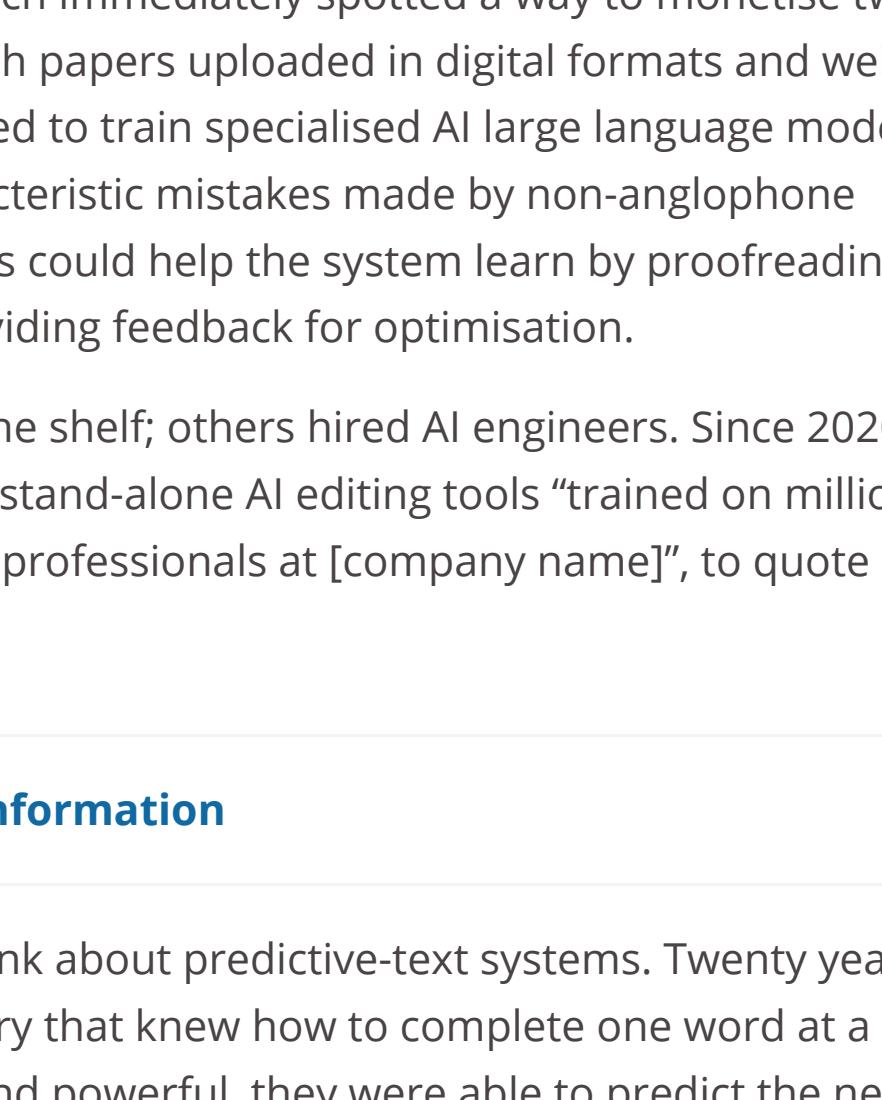
## Editing companies are stealing unpublished research to train their AI

Both publishers and the editing firms they outsource to must seek informed consent to use academics' IP, say Alan Blackwell and Zoe Swenson-Wright

January 12, 2024

Alan Blackwell Zoe Swenson-Wright

Natalia Kucirkova, a professor in Norway, recently [wrote movingly in Times Higher Education](#) about the language discrimination experienced by scholars who use English as a second language. She described the stress caused by insensitive referee comments and the time and money spent preparing articles for journal submission. In the right context, she argued, AI "bots" could level the publication playing field.



Source: iStock

They could. Sadly, in 2024, AI systems are actually being used to exploit non-anglophone scholars by stealing their intellectual property.

Many academic publishers collaborate with large, private editing firms to provide "author services", which include English language editing. The arrival of AI has triggered a frantic race to the bottom among such firms, which immediately spotted a way to monetise two resources they had in abundance: research papers uploaded in digital formats and well-trained editors. Client papers could be used to train specialised AI large language models (LLMs) to recognise and correct the characteristic mistakes made by non-anglophone authors from all parts of the world. Editors could help the system learn by proofreading the automatically generated text and providing feedback for optimisation.

One company bought a small AI firm off the shelf; others hired AI engineers. Since 2020, most have built LLMs and are now selling stand-alone AI editing tools "trained on millions of research manuscripts [...] enhanced by professionals at [company name]", to quote from one promotional blurb.

### Want to write for THE? Click for more information

The best way to understand LLMs is to think about predictive-text systems. Twenty years ago, a language model was just a dictionary that knew how to complete one word at a time. As models became more complex and powerful, they were able to predict the next word or next several words. The latest generation of large language models, like the ones that drive ChatGPT and Copilot for Microsoft 365, can "predict" hundreds of words.

Like all LLMs, editing-company systems encode *everything*, not just editorial corrections. As soon as a researcher uploads a manuscript, their intellectual property – original ideas, innovative variations on established theories, newly coined terms – is appropriated by the company and will be used, likely in perpetuity, to "predict" and generate text in similar papers edited by the service (or anyone using company-provided editing tools).

ADVERTISEMENT

Ad served by Google

Ad options

Send feedback

Why this ad? ▾

Yet few scholars have noticed this fundamental transformation of academic editing. Publishers avoid mentioning the firms they outsource work to. Editing companies boast about AI advances when marketing new tools, but not when advertising editing services. Researchers are encouraged to believe that their papers will be edited entirely by humans. Instead, they are edited by human editors working with (and increasingly marginalised by) AI systems.

Every journal, publisher and editing company guarantees research confidentiality. Their data protection and privacy policies never mention AI. This is misleading but not illegal; current legislation protecting the confidentiality of *personal* data does not regulate or prohibit the use of anonymised academic work.

### Find out more about how to get full unlimited article access to THE for staff and students.

To stave off future lawsuits, most editing firms provide for AI training in their small-print terms of service, where authors unwittingly give them permission to keep their work in perpetuity, share it with affiliates, and use it to improve, develop and deliver current and future products, services and algorithms.

But other prominent victims of AI exploitation are starting to push back. In December, [The New York Times filed suit against ChatGPT](#) for using "millions of articles published by The Times [...] to train automated chatbots that now compete with the news outlet as a source of reliable information". In June, the National Institutes of Health [prohibited scientific peer reviewers from using AI tools](#) to analyse or critique grant applications or R&D contract proposals because there was no "guarantee of where data are being sent, viewed, or used in the future".

As the [Society of Authors](#) points out, the "ethical and moral" issues around the largely profit-driven AI development race "are complex, and the legal ramifications are not limited to the infringement of copyright's economic rights, but may include infringement of an author's moral rights of attribution and integrity and right to object to false attribution; infringement of data protection laws; invasions of privacy; and acts of passing off".

We call on publishers and editing companies to embrace transparency and the fundamental academic principle of informed consent. Editing-service providers should disclose the AI-based systems and tools they use on client work. They should explain clearly how LLMs work and offer scholars a choice, for example by compensating authors for loss of rights by pricing hybrid human/AI editing as a cheaper alternative to fully confidential human editing.

To protect themselves from lawsuits and their authors from exploitation, publishers who offer branded author services should – at a minimum – name the editing companies they outsource work to so that researchers can make an informed choice.

New laws and regulations around AI training are surely on their way. For now, scholars must protect their own intellectual property by learning the basics of AI, reading the small print and interrogating editing services – even those provided by trusted firms and publishers.

ADVERTISEMENT

Ad served by Google

Ad options

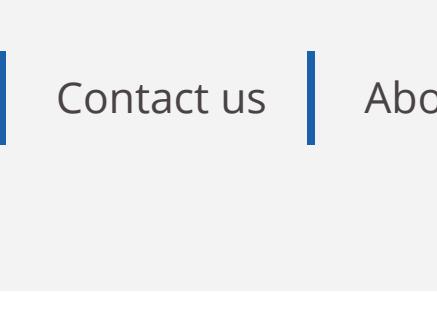
Send feedback

Why this ad? ▾

**Alan Blackwell** is professor of interdisciplinary design in the department of computer science and technology, [University of Cambridge](#), and co-director of Cambridge Global Challenges; his new book, *Moral Codes: Designing Alternatives to AI*, will be published by MIT Press in 2024. **Zoe Swenson-Wright** is a freelance academic editor.

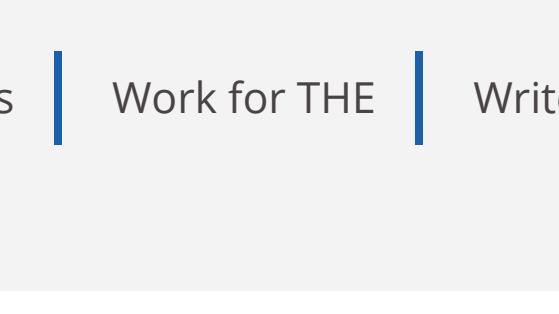
Read more about: Academic publishing Technology and new media

### RELATED ARTICLES



AI writing tools will not fix academia's language discrimination problem

By Natalia Kucirkova 5 September



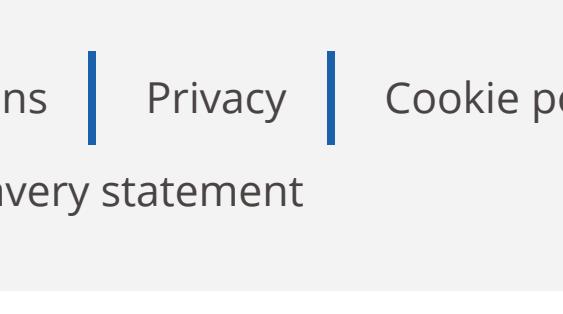
An academic ChatGPT needs a better schooling

By Sorin Adam Matei 28 November



AI writing services proliferating despite essay mill bans

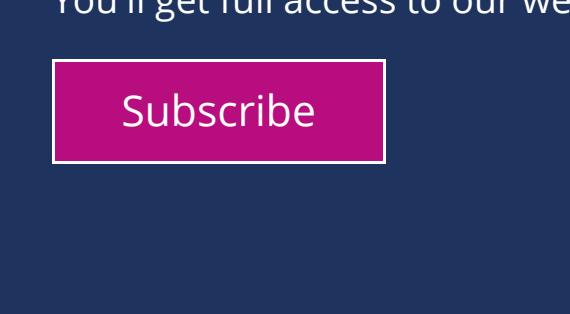
By Tom Williams 20 September



I scramble students' minds so they can unscramble AI

By Jose Marichal 26 October

### YOU MIGHT ALSO LIKE



Open-access to cut 600 jobs

By Emily Dixon 11 January



Universities no one is left behind in the era of AI

By Teruo Fujii 10 January



Online pioneers begin from AI

By Paul Basken 3 January



AI 'lab partners' designs and experiments

By Tom Williams 21 December

### SPONSORED



Jisc Futures: research in the age of open science

Research in the age of open science, UK leads drive towards more open way of sharing science, says Jo Johnson...

Sponsored by Jisc



Student success in a technological world

Extending study spaces, enabling modular learning spaces and using data insights to remove silos were three recommendations from a panel of university and industry thought leaders.

Sponsored by Salesforce



UK Academic Salon 2020: Video highlights and report

Working towards economic stability, working towards economic prosperity and health solutions was explored at the 2020 Academic Salon, in partnership with the UK Academic

Sponsored by Huawei

### RELATED ARTICLES



Open-access to cut 600 jobs

By Emily Dixon 11 January



Universities no one is left behind in the era of AI

By Teruo Fujii 10 January



Online pioneers begin from AI

By Paul Basken 3 January



AI 'lab partners' designs and experiments

By Tom Williams 21 December

### Subscribe

If you like what you're reading online, why not take advantage of our subscription and get unlimited access to all of Times Higher Education's content?

You'll get full access to our website, print and digital editions.

Subscribe

Subscribe

FAQs

Contact us

About us

Work for THE

Write for THE

Terms & conditions

Privacy

Cookie policy

Accessibility statement

THE Connect

Media Centre

Modern slavery

English Simplified Chinese (简体中文)

Copyright © 2024 THE - Times Higher Education