



THE PROJECT TWINS

ARTIFICIAL-INTELLIGENCE SEARCH WRANGLES LITERATURE

Developers want to free scientists to focus on discovery and innovation by helping them to draw connections from a massive body of publications. **By Amanda Heidt**

For a researcher so focused on the past, Mushtaq Bilal spends a lot of time immersed in the technology of tomorrow.

A postdoctoral researcher at the University of Southern Denmark in Odense, Bilal studies the evolution of the novel in nineteenth-century literature. Yet he's perhaps best known for his online tutorials, in which he serves as an informal ambassador between academics and the rapidly expanding universe of search tools that make use of artificial intelligence (AI).

Pulling from his background as a literary scholar, Bilal has been deconstructing the

process of academic writing for years, but his work has now taken a new tack. "When ChatGPT came on the scene back in November, I realized that one could automate many of the steps using different AI applications," he says.

This new generation of search engines, powered by machine learning and large language models, is moving beyond keyword searches to pull connections from the tangled web of the scientific literature. Some programs, such as Consensus, give research-backed answers to yes-or-no questions; others, such as Semantic Scholar, Elicit and Iris, act as digital assistants – tidying up bibliographies, suggesting new papers and generating research summaries.

Collectively, the platforms facilitate many of the early steps in the writing process. Critics note, however, that the programs remain relatively untested and run the risk of perpetuating existing biases in the academic publishing process.

The teams behind these tools say they built them to combat 'information overload' and to free scientists up to be more creative. According to Daniel Weld at the Allen Institute for Artificial Intelligence in Seattle, Washington, and Semantic Scholar's chief scientist, scientific knowledge is growing so rapidly that it's nearly impossible to stay on top of the latest research. "Most search engines help you find

the papers, but then you're left on your own trying to ingest them," he says. By distilling papers into their key points, AI tools help to make that information accessible, Weld says. "We were all loyal fans of Google Scholar, which I still find helpful, but the thought was, we could do better."

The next great idea

The key to doing better lies in a different type of search. Google Scholar, PubMed and other standard search tools use keywords to locate similar papers. AI algorithms, by contrast, use vector comparisons. Papers are translated from words into a set of numbers, called vectors, whose proximity in 'vector space' corresponds to their similarity. "We can parse more of what you mean, the spirit of your search query, because more information about the context is embedded into that vector than is embedded into the text itself," explains Megan Van Welie, lead software engineer at Consensus, who is based in San Francisco, California.

Bilal uses AI tools to follow connections between papers down interesting rabbit holes. While researching descriptions of Muslims in Pakistani novels, AI-generated recommendations based on his searches led Bilal to Bengali literature, and he ultimately included a section about it in his dissertation. For his postdoc, Bilal is studying how Danish author Hans Christian Andersen's stories were interpreted in colonial India. "All that time spent on the history of Bengali literature came rushing back," he says. Bilal uses Elicit to iterate and refine his questions, Research Rabbit to identify sources and Scite – which tells a user not only how often papers are cited, but in what context – to track academic discourse.

Mohammed Yisa, a research clinician at the Medical Research Council Unit The Gambia of the London School of Hygiene & Tropical Medicine, follows Bilal on Twitter (now known as X), and sometimes spends evenings testing the platforms that Bilal tweets about.

Yisa particularly enjoys using Iris, a search engine that creates map-like visualizations that connect papers around themes. Feeding a 'seed paper' into Iris generates a nested map of related publications, which resembles a map of the world. Clicking deeper into the map is like zooming in from a country-wide view down to, say, states (sub-themes) and cities (individual papers).

"I consider myself a visual learner, and the map visualization is not something I've seen before," Yisa says. He's currently using the tools to identify papers for a review on vaccine equity, "to see who is talking about it at the moment and what is being said, but also what has not been said".

Other tools, such as Research Rabbit and LitMaps, tie papers together through a network map of nodes. A search engine targeted at medical professionals, called System Pro,

creates a similar visualization, but links topics by their statistical relatedness.

Although these searches rely on 'extractive algorithms' to pull out useful snippets, several platforms are rolling out generative functions, which use AI to create original text. The Allen Institute's Semantic Reader, for instance, "brings AI into the reading experience" for PDFs of manuscripts, Weld says. If users encounter a symbol in an equation or an in-text citation, a card pops up with the symbol's definition or an AI-generated summary of the cited paper.

Elicit is beta-testing a brainstorming feature to help generate better queries as well as a way to provide a multi-paper summary of the top four search results. It uses Open AI's ChatGPT but is trained only on scientific papers, so is less prone to 'hallucinations' – mistakes in generated text that seem correct but are actually inaccurate – than are searches based on the entire Internet, says James Brady, the head of engineering for Elicit's parent company, Ought, who is based in Oristà, Spain. "If you're making statements that are linked to your reputation, scientists want something a bit more

"We were all loyal fans of Google Scholar, which I still find helpful, but the thought was, we could do better."

reliable that they can trust."

For his part, Miles-Dei Olufeagba, a biomedical research fellow at the University of Ibadan in Nigeria, still considers PubMed to be the gold standard, calling it "the refuge of the medical scientist". Olufeagba has tried Consensus, Elicit and Semantic Scholar. Results from PubMed might require more time to sort through, he says, but it ultimately finds higher-quality papers. AI tools "tend to lose some info that may be pivotal to one's literature search", he says.

Early days

AI platforms are also prone to some of the same biases as their human creators. Research has repeatedly documented how academic publishing and search engines disadvantage some groups, including women¹ and people of colour², and these same trends emerge with AI-based tools.

Scientists who have names that contain accented characters have described difficulties in getting Semantic Scholar to create a unified author profile, for instance. And because several engines, including Semantic Scholar and Consensus, use metrics such as citation counts and impact factors to determine ranking, work that is published in prestigious journals or sensationalized inevitably gets bumped to the top over research that might be more relevant, creating what Weld

calls a "rich-get-richer effect". (Consensus co-founder and chief executive Eric Olson, who is based in Boston, Massachusetts, says that a paper's relevance to the query will always be the top metric in determining its ranking.)

None of these engines explicitly mark preprints as worthy of greater scrutiny, and they display them alongside published papers that have undergone formal peer review. And with controversial questions, such as whether childhood vaccines cause autism or humans are contributing to global warming, Consensus sometimes returns answers that perpetuate misinformation or unverified claims. For these charged questions, Olson says that the team sometimes reviews the results manually and flags disputed papers.

Ultimately, however, it's the user's responsibility to verify any claims, developers say. The platforms generally mark when a feature is in beta testing, and some have flags that indicate a paper's quality. In addition to a 'disputed' tag, Consensus is currently developing ways to note the type of study, the number of participants and the funding source, something Elicit also does.

But Sasha Luccioni, a research scientist in Montreal, Canada, at the AI firm Hugging Face, warns that some companies are releasing products too early because they rely on users to improve them – a common practice in the tech-start-up world that doesn't gel well with science. Groups have also become more secretive about their models, making it harder to address ethical lapses. Luccioni, for instance, studies the carbon footprint of AI models, but says she struggles to access even fundamental data such as the size of the model or its training period – "basic stuff that doesn't give you any kind of secret sauce". Whereas early arrivals such as Semantic Scholar share their underlying software so that others can build on it (Consensus, Elicit, Perplexity, Connected Papers and Iris all use the Semantic Scholar corpus), "nowadays, companies don't provide any information, and so it's become less about science and more about a product".

For Weld, this creates an extra imperative to ensure that Semantic Scholar is transparent. "I do think that AI is moving awfully quickly, and the 'let's stay ahead of everyone else' incentive can push us in dangerous directions," he says. "But I also think there's a huge amount of benefit that can come from AI technology. Some of the main challenges facing the world are best confronted with really vibrant research programmes, and that's what gets me up in the morning – to help improve scientists' productivity."

Amanda Heidt is a science writer in Moab, Utah.

1. Ross, M. B. et al. *Nature* **608**, 135–145 (2022).
2. Salazar, J. W. et al. *JAMA Int. Med.* **181**, 1248–1251 (2021).

Correction

This article gave the incorrect title for Mohammed Yisa.