

# Sliced, diced and digested: science ready in minutes

AI technology can decide which papers are worth reading, and condenses them to make the literature more accessible. **By Chris Woolston and Jeffrey M. Perkel**

**N**o matter how many months or years authors take to produce a scientific paper, Sabine Louët needs only a few seconds to generate a coherent 300-word summary of it. But she leaves the thinking to an artificial intelligence (AI) algorithm that statistically analyses the text, identifies meaningful words and phrases, and pieces it all together into a crisp, readable chunk. “We’re trying to tell a story, and we want to make it as digestible as possible,” says Louët, chief executive of SciencePOD, a Dublin-based science communication company.

As the volume of research continues to grow, natural-language processing programs that can rapidly sort and summarize scientific papers have become an increasingly important tool for scientific publishers and researchers alike, says Markus Kaindl, senior manager for data development at Springer Nature, which publishes Nature Index. (Nature Index is editorially independent of its publisher.)

The company has engaged SciencePOD and others to explore the use of AI to enhance content appeal and accessibility. “AI can really help us as science publishers, by summarizing information, translating it for wider audiences and increasing the impact,” says Kaindl.

**“We’re trying to tell a story, and we want to make it as digestible as possible.”**

He points to the roughly 2,000 papers published on COVID-19 each week, enough to overwhelm anyone trying to stay on top of the field. “It’s like an ocean of content, and it feels like our users are close to drowning,” he says. “We need to help them surf that wave instead.”

AI can help identify the papers most suited to a particular user’s needs. For example, Semantic Scholar, developed by the Allen Institute for Artificial Intelligence in Seattle, Washington, goes beyond keywords to rank the most relevant papers for any query. “It’s a brilliant

platform because it really tries to understand what the publications are about,” Kaindl says. Springer Nature expects to go further by offering personalized summaries and search results. “If you are a senior career researcher, a postdoc or a principal investigator, your needs from a paper or a chapter may be very different from someone at an earlier career stage,” he says.

Finding papers is an important first step, but AI technology can also help researchers to decide if a selected paper is actually worth reading, says David Konopnicki, manager of the Language and Retrieval group at IBM Research AI in Haifa, Israel. Researchers used to rely on abstracts to make those decisions, but a computer-generated summary that synthesizes the main points of the paper can be more helpful, he says. “The role of the abstract is to convince you to read the article, but that’s maybe not what you want to do,” he adds. “I want to be able to understand if a piece of research is impactful very quickly. It’s a very difficult task.”

In 2019, Konopnicki and his team launched the IBM Science Summarizer, a service that, as he puts it, “slices and dices the literature” to help users track the latest papers in the field of AI that have been uploaded to sites such as preprint service arXiv and ACL Anthology, a reference repository of conference and journal papers in natural-language processing and computational linguistics. The project was partly self-serving, in response to a boom in research on natural-language processing, natural-language generation and information retrieval. “Every week I have to search for new articles that impact our work,” he says. “We used to do that just a couple of times a year.” The search function identifies papers based on key words, data sets or author name. The Summarizer then condenses the main sections of each paper into short, readable chunks.

As Konopnicki explains, the Summarizer uses extractive text, which means the words and sentence structures are pulled directly from the paper. “We’re not rewriting sentences,” he says. The challenge is identifying the sentences and sections that truly matter. During the development phase, Konopnicki and his



**Markus Kaindl.**

team used transcripts from conference presentations to help train the algorithm to spot the most important parts of papers. Authors were given an opportunity to grade the results and offer feedback.

The IBM team is now exploring the possibility of extending the Summarizer to other scientific fields, but it’s not a trivial transition. “Each domain has its peculiarities,” Konopnicki says. “You need to apply different techniques to a physics paper or a mathematical paper.”

Wiley, another leading scientific publisher, plans to start generating AI summaries of papers in the next several months, says David Flanagan, the company’s director of data science, based in Frankfurt, Germany. The summaries will not only help researchers who are trying to stay abreast of their fields, but will also help the general public and funding agencies to make sense of technical subjects. “Researchers



SPRINGER NATURE: SHUTTERSTOCK

### One vision for the future use of AI in scientific publishing is to synthesize information from multiple journal articles.

are at the centre of what we do, so we're always looking for ways to make dissemination of their results easier," he says.

Wiley is also using AI to suggest possible alternative destinations for rejected papers. "That saves the authors, the editors and the reviewers work and shortens the overall time to publication," Flanagan says. "We anticipate deploying additional AI-powered tools like this to streamline the publication process and improve the quality of the literature."

Wiley, like other publishers, is developing its own AI tools to complement technologies from other industries. "A typical AI product at Wiley might combine proprietary data, publicly available data, open source tools and our own in-house developed tools," he says. The company is currently partnering with UNSILO, an AI company based in Aarhus, Denmark, to develop a deep-learning tool that would help

authors automatically generate data availability statements, a requirement of many funding agencies. He hopes that AI could be used to search for bad data or manipulated images, preferably before they are ever published.

Louët says that SciencePOD's algorithm, developed in conjunction with a team at the University of Avignon in France, is continually evolving to make the results more useful for readers and researchers. "There's a lot of fine-tuning," she says. "A lot of people are trying to make summaries with off-the-shelf algorithms. We're telling the algorithms how to find the most important point."

As AI becomes more sophisticated, the generated text will become more indistinguishable from text written by humans, Louët says. She adds that SciencePOD algorithms can already produce journalistic overviews of scientific papers. Journalists aren't out of business, she

says. For most journalistic purposes, it still takes a human to polish text, interview experts and put science into context.

Although there are kinks to be ironed out for AI to reach its full potential, Kaindl says, the near future looks promising. One ambitious vision is scientific searches that can synthesize information from many different journal articles. As Kaindl explains, a researcher could be looking for the best reagent for a particular experiment or for some other complex piece of information. "The system will be smart enough to understand what you want to know and return a summary by searching Springer Nature's corpus of 13 million publications," he says. "That's maybe a couple of years down the line, but it's what we should strive for."

**Chris Woolston** is a writer in Billings, Montana. **Jeffrey M. Perkel** is Nature's technology editor.