Title: Scideator: Human-LLM Scientific Idea Generation Grounded in Research-Paper Facet Recombination

Authors: Marissa Radensky (radensky@cs.washington.edu), Simra Shahid, Raymond Fok, Pao Siangliulue, Tom Hope, Dan Weld
Affiliations: University of Washington, Allen Institute for AI, Adobe India, Hebrew University of Jerusalem

Ideation is at the core of the scientific process. To perform interesting research, a scientist must first have a good idea, one that is both novel and useful. Scientists are more likely to come across a good idea when they generate more ideas overall. A common method for generating diverse ideas is to identify analogies between different concepts. In the case of scientific ideation, a major source of inspiration is reading research papers, which provide concepts that can be analogized in order to spark interesting ideas.

However, fixation makes producing many different ideas difficult. For scientists, scholarly filter bubbles often bias them towards reading papers with the same set of concepts. Moreover, when exposed to a larger variety of concepts, scientists often struggle to see connections between familiar and new concepts. Even if a scientist has managed to identify an interesting analogy of concepts to form a research idea, assessing the idea's novelty in comparison to the existing literature is a cumbersome yet critical task. As large language models (LLMs) become more powerful, they have great potential to support scientists in identifying novel analogies. LLMs can help scientists to 1) explore more concepts and analogies between concepts as well as 2) evaluate and improve idea novelty. Still, thoughtful scaffolding is required to support the scientist-LLM interaction necessary for a fruitful ideation session.

In this talk, I will describe an LLM-powered tool called Scideator, developed in collaboration with the Allen Institute for AI, that provides scaffolding for analogy-based ideation as well as evaluation and improvement of idea novelty. Scideator extracts key facets from scientific papers, which includes the purpose (the problem the paper addresses) and mechanism (the paper's proposed solution to the problem). The tool creates analogies that map the purpose and mechanism of one paper's idea to the purpose and mechanism of another paper's idea. These analogies are then used to suggest ideas that combine a purpose from one paper and a mechanism from an analogous paper. The scientist may also add and select facets themselves to generate ideas. For each generated idea, the tool presents its analysis of the idea's novelty in relation to related papers, which can be edited by the scientist. If the analysis indicates that the idea is not sufficiently novel, the tool suggests changes to the idea's facets to make it more novel.

I will also discuss a within-subjects user study through which we compare how computer-science researchers generate research ideas with Scideator versus a baseline of access to an LLM and a scholarly search engine. We find that participants save more ideas with Scideator. We further observe differences in participants' comments on the utility of Scideator versus the baseline.

References Representative of My Work on Computational Discovery:

[1] Radensky, Marissa. "Mixed-Initiative Methods for Co-Creation in Scientific Research." *Proceedings of the 16th Conference on Creativity & Cognition*. 2024. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C48&q=marissa+radensky+mixed-initiative+methods+for+co-creation+in+scientific+research&btnG=
[2] Portenoy, Jason, et al. "Bursting scientific filter bubbles: Boosting innovation via novel author discovery." *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C48&q=Bursting+Scientific+Filter+Bubbles%3A+Boosting+Innovation+via+Novel+Author+Discovery&btnG=