Generative Artificial Intelligence Will Transform Post-Print Academic Research Methods

Authors: Xue Gang

Date: 2025-01-08T00:00:00+00:00

Abstract

, when GAI reaches the singularity and becomes AGI—when AI no longer indirectly understands the world through human thought (the human-created and natural) but directly interacts with it (constituting a subject beyond human conceptual standards that humans cannot define)—and when silicon-based and carbon-based life coexist on Earth, how will we judge whether "its" research remains meaningful to humans? The conclusion is that all patterns of human thought will necessarily transform.

Full Text

Preamble

How Generative AI Will Transform Academic Research Methods
Xue Gang A
(Chinese Social Sciences Press, Beijing 100026)

DOI: 10.31193/SSAP.J.ISSN.2096-6695.2024.04.06

Journal of Literature and Data Science, 2024, 6(4): 015-017.

Following several iterations since 2023, Generative Artificial Intelligence (GAI) has demonstrated capabilities surpassing individuals in certain domains. This discussion focuses on the application of Large Language Models (LLM) in philosophy and social sciences research.

1. GAI Is Transforming Academic Writing

Academia has begun paying attention to GAI, particularly LLMs, because they can generate papers through prompts to some extent—a development that presents both opportunities and challenges for scholars and journals alike. In the

absence of universal standards for AI usage, determining whether a paper or certain sections and arguments were AI-generated remains difficult. Although some domestic institutions have launched AI writing detection services, questions persist about their effectiveness at scale and whether their iteration speed can keep pace with AI development itself. Consequently, many traditional academic journals have expressed varying degrees of concern about publishing LLM-generated papers that go undetected.

On this issue, we argue that such concerns are unwarranted and that LLM participation in paper writing should be accepted to a certain degree. Of course, questions about how and when LLMs can be widely implemented belong to the realm of policy and practical operations, which fall outside our present discussion.

Maintaining an objective and calm attitude is essential for treating GAI scientifically. While LLMs are indeed remarkable in language generation, examining their underlying technical logic reveals both strengths and weaknesses. Built upon deep neural network learning, LLMs construct their corpora through crawling, tagging, pre-training, and fine-tuning publicly available web data. Their advantage lies in the breadth of this data, enabling high accuracy in commonsense knowledge responses. Conversely, specialized knowledge often remains submerged in massive datasets, preventing it from being specially tagged or incorporated into professional corpora. When such knowledge is absent from the training data, LLMs perform at roughly the level of search engines when addressing specialized questions.

Therefore, given current LLM technology, data quality is paramount. The quality of training data directly determines LLM quality. The vast expanse of public web data suffers from uneven quality, filling LLM corpora with 杂乱信息. Even with biased filtering during data collection, only a tiny fraction of public data is actually captured. In contrast, professional non-public data remains beyond LLM reach for two reasons. First, well-funded companies capable of training LLMs are mostly technology-focused or general-purpose entities that pay little attention to specific domains, and thus do not separately crawl or train on domain-specific data. Second, departments that need domain-specific LLMs possess high-quality internal data but lack the funds, technology, and personnel to develop them. This is especially true for traditional academic journal editorial offices, where a technological gap prevents them from engaging with LLMs and fuels their concerns about publishing GAI-generated papers.

Current research methods are literature-based, typically involving topic selection, literature review, identifying gaps in previous research, posing questions, and offering solutions—making data quality a critical factor. GAI has three main parameters: algorithm, computing power, and data. While the first two are products of computer development, the third encompasses anything that can be digitized, particularly text. In the public domain, the internet is flooded with low-quality, repetitive, and even spam information, which significantly increases GAI's computational costs and prevents resulting LLMs from achieving

elite human-level performance.

While hardware and software advances may enable more effective data filtering, the current energy crisis driven by computing demands suggests such methods remain elusive. By contrast, professional domain data has been naturally filtered, features limited volume, and allows for human intervention, yielding far higher quality than public web data. This explains why technology companies have recently shifted toward developing specialized LLMs for various social domains. In philosophy and social sciences, the data volume is even smaller than on the public internet. Consequently, the path forward for academia and academic journals lies in developing targeted "small models" through secondary development based on open-source large model frameworks.

2. Two Philosophical Reflections on GAI

Humanists approach cutting-edge science differently than natural scientists—we lack highly specialized and systematic technical knowledge. Yet technology ultimately concerns humanity and society, which falls within our research domain. Thought experiments are our forte, and as scholars have noted, philosophy's key lies in its future orientation. In this sense, our two philosophical reflections on GAI gain a certain philosophical legitimacy.

The first reflection concerns whether authors of GAI-assisted papers remain human. Two approaches offer possible answers. The technical approach distinguishes components of paper creation, identifying which constitute the core, and judges whether this core represents human creation. In philosophy and social sciences, papers typically involve topic selection, arguments, logical structure, literature review, and specific demonstrations. In our view, the core comprises topic selection, arguments, and logical structure—though current standards also demand originality in literature review and demonstration. Judging whether these core elements are human-created represents the technical route. The philosophical approach, by contrast, determines authorship by defining and confirming the creative subject, invoking philosophy's ancient yet enduring concepts of "self" and "self-consciousness." The self is generally understood as possessing identity—I am not other, not any thing or otherness, but simply myself. Self-consciousness emerges when I can point to myself and say "I." For GAI-assisted paper creation, identifying the subject becomes the criterion. A subject's agency and spontaneity provide the initial impetus for writing—qualities GAI clearly lacks.

Based on these two routes, we have two answers to whether GAI-assisted papers have human authors. The technical answer is more pragmatic: as long as standards are established and widely accepted, the paper's core remains human-created. This raises a bold speculation: if non-core parts are not human-created, must humans still study or practice these aspects? In other words, will future academic writing require only creativity and topic selection? If so, academic research methods will fundamentally change, further supporting the view that

philosophy's object is solely the human spirit. The philosophical answer is more idealistic: GAI lacks self and self-consciousness, and until AGI (Artificial General Intelligence) is created, existing GAI cannot generate a self distinct from human selfhood. GAI's generative capacity depends on human prompts processed through domain-specific models like LLMs—its core lies in computing power and algorithms. It surpasses only certain individual human capabilities, not human creativity, imagination, or self-consciousness. Therefore, authors of GAI-assisted papers remain human. Nevertheless, current AI has already prompted philosophical reflection on concepts like subject, object, subjectivity, object, essence, and otherness, meaning philosophy itself must evolve alongside AI development.

The second reflection concerns whether the mode of expressing ideas will change. This raises the deeper question of whether academic papers will remain necessary in AI's rapid evolution, or whether the paper format will become unsuitable for expressing academic thought—or even whether academia itself will become obsolete.

The paper format for expressing academic thought emerged roughly during German classical philosophy. Earlier forms included dialogue, aphorism, essay, and even geometric-style axiomatic propositions, but papers eventually became the standard. At GAI's current level, it can already organize and review literature, generate content (where contemporary LLMs already surpass most humans in linguistic expression), and even construct logical frameworks. In other words, after a human author provides a topic, subsequent work can be delegated to LLMs. If so, do we still need to read GAI-completed papers to advance our research? Of course, academia is but one mode of expressing thought—others include art and religion. We can predict that future research will be based on questions, logic, and ideas rather than literature and concrete expression. Skipping the demonstration process, our conclusion is that GAI changes how academia expresses thought, not that academia or the need to express thought will disappear.

In summary, when GAI reaches the singularity and becomes AGI—when AI no longer indirectly understands the world through human thought (the human-created and natural) but directly interacts with it (constituting a subject beyond human conceptual standards that humans cannot define)—and when silicon-based and carbon-based life coexist on Earth, how will we judge whether "its" research remains meaningful to humans? The conclusion is that all patterns of human thought will necessarily transform.

Note: Figure translations are in progress. See original paper for figures.