"Where is AI being written?" – A geographic and epistemic meta-analysis of scientific literature

Vlad Morar, Filip Ilievski

Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
v.morar@student.vu.nl, f.ilievski@vu.nl

It could be said that our civilization was built on pillars of writings. From ancient clay tablets and medieval manuscripts to modern digital archives, man's peculiar habit of writing thoughts and discoveries has enabled a continuously expanding foundation of knowledge, leading to many of the things that surround us today, among which, Artificial Intelligence. Although it became prominent only recently, primitive seeds of its conception can be traced to Antiquity and the Middle Ages, in myths such as Talos[1], the bronze automaton, or the golems[2] of Jewish folklore, which are artificial humans made of clay and perceived as the physical form of man before acquiring a soul, denoting unformed substance. Centuries of developments in mathematics, logic, and technology, through writings such as those of Leibniz, Boole, Babbage and Turing, nurtured these seeds and culminated in the digital computer, which sparked the notion of an electronic brain, leading to AI truly blooming in 1956's Dartmouth Summer Research Project[3]. Against this long history of writing that contoured the birth of AI, it is left to wonder where is it being written now? Such question asks for a meta-analysis to answer it. While various meta-analysis works already exist, in regard to AI, they have mostly addressed specific subdomains (e.g., NLP[4]), applied fields (e.g., business[5][6], banking[7]) or niche topics (e.g., misinformation during disasters[8]). This study undertakes a broader meta-analysis to answer "Where is AI being written?", while addressing both its geographical nuance, regarding where contributions to AI literature emerge globally, and its epistemic nuance, concerning where AI concepts intersect with other research domains.

The methodology underpinning this research consisted in the construction of two complementary datasets through the OpenAlex bibliographic index. First, a rigorously extracted corpus of 10,000 AI publications was gathered through specific metadata criteria, from which authorship-based country data was derived and systematically catalogued to assess both the quantity and the weight of country contributions, as well as their percentage of independent versus internationally co-authored works. Second, a corpus of 13,390 "non-AI" publications was established by extending relevant OpenAlex-assigned keywords of the AI set into adjacent OpenAlex-defined research fields, thereby furnishing material for the epistemic dimension of the inquiry. From the dataset which ensued, keywords were extracted from titles and abstracts using a transformer model, and mapped onto a co-occurrence graph, further analyzed through PageRank and Leiden community detection to identify central terms and thematic clusters.

When discussing overall contributions, only 22 countries surpassed 100 works, with the United States standing peerless at the top: its 4,366 publications, nearly

half of the dataset, outnumber the combined 4,271 works of the next four highest-ranking countries. China follows with 1,991, the United Kingdom with 1,040, then Germany with 625 and Canada with 615, while Belgium closes the list at 103. Together, European countries with over 100 publications amassed 3,055 works, more than China, yet still far behind the United States. Turning from quantity to weight, the works' primary authorship paints a more nuanced picture: the United States again dominates with 32.4% of all primary authorship, 80% of which is independent. China, with 15.7%, ranks second yet produces only 53% independently, betraying greater reliance on international collaboration, whereas the United Kingdom holds third with 5.9% and a 60% independence rate. India ascends to fourth with 4.5% and an impressive 71% independence, surpassing Germany's 4.2% and 65%, while Canada follows with 3.8% and 56%. The OpenAlex disciplinary distribution indicates a diverse dispersion across fields: "Computer Science" leads with 33.3%, followed by "Engineering" (14.0%) and "Medicine" (7.0%), while "Biochemistry, Genetics and Molecular Biology" (6.5%), "Social Sciences" (6.3%) and "Psychology" (4.6%) also hold notable shares. The co-occurrence graph derived from the second corpus, comprising 3,450 nodes and 16,088 edges, reveals a moderately sparse structure in which associations remain diffuse yet not incidental, as recurrent pairings surface with modest regularity without consolidating into dense clusters. PageRank uncovers within this structure a semantic hierarchy dominated by "machine learning" and related terms, while "artificial intelligence" itself ranks lower, as if referenced more through its constituent techniques rather than its explicit name. At the same time, PageRank delineates a compact epistemic core of strongly interconnected AI concepts, encircled by a periphery of diverse scientific vocabularies, in which Leiden partitioning further identifies domain-specific communities, from paleopathology and clinical imaging to affective computing, text analysis, and social determinants of health, thus outlining the contours of AI's broader discursive reach.

Regarding the motives shaping these results, the conclusions remain of an unfolding nature that resist definitive contouring, being open to interpretation through a multiplicity of lenses. Still, with regard to the more definable question of "Where is AI being written?", firmer affirmations emerge. Geographically, Western countries, led by the United States, dominate, although China and India stand out as significant exceptions, their publication volumes often rivaling or surpassing European counterparts. This suggests a nuanced global topology of AI production, shaped not only by research capacity, but probably by geopolitical strategies of technological sovereignty as well, China's substantial yet collaborative output possibly exemplifying this dynamic in initiatives of self-reliance such as DeepSeek[9]. Epistemically, AI is mainly written within Computer Science, while extending into Engineering, Medicine, and other statistically intensive fields. Conceptual co-occurrence shows that AI diffuses across disciplines primarily as a methodological toolkit that traverses domain vocabularies from medicine to paleopathology. Yet this diffusion is largely unidirectional, as many disciplines adopt AI techniques, but few contribute to its theoretical development, revealing an asymmetry in epistemic exchange.

References

1. Adrienne Mayor, *Gods and Robots: Myths, Machines and Ancient Dreams of Technology*, 2018
2. Moshe Idel, **Golem: Jewish Magical and Mystical Traditions on the Artificial Anthropoid**, 1990
3. J. McCarthy, L. Minsky, N. Rochester, C.E. Shannon, A proposal for the Dartmouth summer research project, 1955
4. Tim Schopf, Karim Arabi, and Florian Matthes, *Exploring the Landscape of Natural Language Processing Research*, *Proceedings of Recent Advances in Natural Language Processing*, 2023.
5. Sunaina Thakuri, Massimo Bon, Nadire Cavus, Nuriye Sancar, *Artificial Intelligence on Knowledge Management Systems for Businesses: A Systematic Literature Review*, *TEM Journal*, 2024.
6. Elizabeth Real de Oliveira, Pedro Rodrigues, *A Review of Literature on Human Behaviour and Artificial Intelligence: Contributions Towards Knowledge Management*, *Electronic Journal of Knowledge Management*, 2021.
7. Sushil Kalyani, Neha Gupta, *Is artificial intelligence and machine learning changing the ways of banking: a systematic literature review and meta analysis*, 2023.
8. Rosa Vicari, Nadejda Komendatova, *Systematic meta-analysis of research on AI tools to deal with misinformation on social media during natural and anthropogenic hazards and disasters*, 2023.
9. Wenfeng Liang et al., *DeepSeek-V3 Technical Report*, 2025.