Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Academic Impact vs. Societal Attention: A Dual-Analysis of Top-Cited Artificial Intelligence Articles in Medicine

Furkan Şakiroğlu

fsakiroglu@gmail.com

Çat State Hospital

Cemil Çolak

Inonu University

Research Article

Keywords: Artificial Intelligence, Bibliometrics, Altmetrics, ChatGPT, Medical Research, Societal Impact, Large Language Models, Citation Analysis

Posted Date: March 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8920991/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Objective

This study aims to examine the relationship between the academic citation success and social visibility of Artificial Intelligence (AI)-based medical research using bibliometric and altmetric methodologies.

Methods

The top 100 most-cited articles indexed in the Web of Science Core Collection from 1 January 2023 to 27 January 2026 were analyzed; citation counts and Altmetric Attention Scores (AAS) were retrieved on 27 January 2026 (See Methods for the full search query). Academic impact was measured by Web of Science citation counts, while social impact was evaluated using the Altmetric Attention Score (AAS). Data were assessed through Spearman’s correlation analysis and the Mann-Whitney U test.

Results

A statistically significant but weak positive correlation was identified between citation counts and AAS ($r = 0.299$, $p = 0.0025$). Open access status characterized 92% of the articles. The highest academic impact was achieved by the ChatGPT-USMLE study by Kung et al. (2023) with 2,193 citations, whereas the highest social impact was held by the "Physician vs. Chatbot" study by Ayers et al. (2023) (AAS: 6,388). A notable finding was that publications originating from China exhibited remarkably low altmetric scores (Median AAS: 12) despite high academic citation rates, suggesting a 'digital isolation' effect that may stem from Western-centric altmetric data coverage.

Conclusion

Academic success and societal popularity are governed by distinct dynamics, indicating the need for researchers to adopt science communication strategies and for funding agencies to use multidimensional impact metrics. While academia prioritizes conceptual depth—such as ethics and methodology—the general public shows greater interest in sensational competition (e.g., physician vs. AI). It is recommended that researchers enhance their science communication competencies and that funding agencies adopt multidimensional evaluation approaches.

1. Introduction

Artificial Intelligence (AI), particularly with the introduction of ChatGPT (November 2022) and Large Language Models (LLMs), has triggered an unprecedented paradigm shift in medical research (1). AI-based applications are rapidly proliferating across nearly every domain of healthcare, ranging from diagnostic imaging to personalized treatment, and from medical education to clinical decision support

systems. This technological revolution has fundamentally altered not only clinical practices but also the pace of scientific knowledge production and the dissemination of that knowledge (2).

Traditional bibliometric analyses and citation counts have been recognized as the 'gold standard' for measuring scientific productivity and impact for many years. However, because citation-based metrics are inherently cumulative and often require years post-publication to yield meaningful data, they may prove inadequate for measuring the real-time impact in a field evolving at such a breakneck pace as Artificial Intelligence (3). Furthermore, citation counts solely reflect the interest of the 'academic community' while neglecting the reaction—and thus the societal impact—of patients, policymakers, the media, and the general public. To overcome this limitation, 'Altmetrics' have emerged as a complementary tool to traditional metrics, offering the opportunity to monitor the real-time reflections of scientific outputs across social media, news outlets, blogs, and policy documents (4). Since AI research in the field of medicine concerns not only scientists but also public health directly (e.g., physician vs. AI debates), these studies possess a high potential for going 'viral' on social media. Consequently, a linear relationship may not always exist between a paper's scientific quality and its societal popularity. There is a broad consensus in the literature that altmetrics measure societal curiosity and media visibility rather than scientific quality. In their meta-analysis examining the correlation between altmetric scores and citation counts in health sciences, Kolahi et al. (2021) identified a pooled correlation coefficient of 0.19, demonstrating that these two metrics reflect distinct dimensions of impact (5). However, the methodological limitations of altmetrics should not be overlooked. Thelwall (2021) emphasized that while altmetric indicators reflect attention, they fail to distinguish the nature of that attention (e.g., endorsement, criticism, or the dissemination of misinformation). Furthermore, it is well-known that social media mentions can be artificially inflated through bot activity or coordinated campaigns (6).

Although numerous bibliometric studies exist on AI in medicine, they have predominantly focused on either academic citations or altmetric scores in isolation. To our knowledge, no study has yet jointly analyzed the top-cited articles from the post-ChatGPT era (2023–2026) using both metrics to uncover the divergence between scientific and societal attention. This study fills that gap by providing a comprehensive, dual-perspective analysis of the most influential AI research in medicine. Distinguishing which topics gain academic depth (e.g., machine learning ethics) versus those that generate societal excitement (e.g., chatbots) is critical for directing future research funding and health policies.

The aim of this study is to analyze the most-cited articles in the fields of medicine and informatics within the Web of Science (WoS) database for the recent period (2023–2026) using bibliometric and altmetric methodologies. The study examines the correlation between academic citations and social media engagement, the performance of the most influential journals and countries, and the thematic differences (keyword analysis) between priorities of academia and society. The findings intend to provide a holistic perspective on the scientific and societal reflections of AI research.

This study seeks to answer the following research questions:

1. What is the relationship between Web of Science citation counts and the Altmetric Attention Score?

2. How does open access status influence the academic and societal impact of articles?
3. Which thematic areas receive more attention from the academic community versus the general public?
4. How do academic-societal impact profiles differ across countries and journals?

Based on previous meta-analyses (e.g., Kolahi et al., 2021), we hypothesized that there would be a weak positive correlation between citation counts and AAS. We also anticipated that open access status would be associated with higher AAS, and that certain thematic areas (e.g., chatbots) would garner more social attention than technical topics (e.g., machine learning).

2. Materials and Methods

2.1. Research Design and Data Sources

This study is a cross-sectional, descriptive bibliometric and altmetric analysis evaluating the academic and societal impact of Artificial Intelligence (AI)-based medical research. Two primary data sources were utilized. The Web of Science Core Collection (Clarivate Analytics, Philadelphia, PA, USA) was employed to obtain scientific impact (citation counts) and bibliometric metadata (7). To measure the online visibility of the articles across social media, news outlets, and policy documents (Altmetric Attention Score-AAS), Altmetric.com (Digital Science, London, UK) was used (8).

2.2. Data Collection and Search Strategy

Data retrieval was conducted between January 2023 and January 2026 to encompass the rise of generative AI and the most recent literature. WoS citation counts and Altmetric Attention Scores were retrieved on 27 January 2026 at UTC 12:00. The search strategy utilized the following term combinations within the Title (TI) and Topic (TS) fields:

Keywords: ("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Large Language Models" OR "ChatGPT" OR "Generative AI")

Context: ("Medicine" OR "Health*" OR "Clinical" OR "Diagnosis" OR "Prognosis")

Exact query string and the date filter:

TS=("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Large Language Models" OR "ChatGPT" OR "Generative AI") AND TS = ("Medicine" OR "Health*" OR "Clinical" OR "Diagnosis" OR "Prognosis") AND Document Types=(Article OR Review) AND Timespan = 2023–2026 (search conducted on January 27, 2026, covering publications from January 1, 2023, to January 27, 2026))

Inclusion Criteria:

- Studies published in the English language.

- Document types restricted to "Original Research Article" or "Review".
- Publications indexed under Web of Science categories related to clinical medical sciences, healthcare services, or medical informatics.

Editorial materials, letters, meeting abstracts, and corrections were excluded from the analysis. The results were ranked in descending order based on their Web of Science Core Collection citation counts, and the top 100 most-cited articles constituted the study sample. We selected the Web of Science Core Collection due to its comprehensive coverage of high-impact medical journals and its reliable citation data; Scopus and PubMed were not included to maintain consistency with previous top-cited bibliometric studies.

2.3. Data Extraction and Altmetric Mapping

The data extraction process was finalized on January 27, 2026. Bibliometric metadata (Title, Author, Journal, Publication Year, Author Keywords, Affiliations, Citation Count, and DOI) for the selected 100 articles were exported from the WoS database in .xlsx format.

To obtain altmetric data, queries were performed via the Altmetric.com API using the Digital Object Identifier (DOI) of each article. The Altmetric Attention Score (AAS) and its source distribution were recorded. For each publication, the following metrics were documented: Total AAS, number of news mentions, Twitter/X posts, Facebook mentions, blog posts, Wikipedia citations, and Mendeley readership counts.

2.4. Analytical Framework

In the analytical framework of the study:

- Scientific impact was assessed via total WoS citations.
- Societal impact was evaluated through the AAS.
- Conceptual and geographical structures were analyzed using author keywords and the institutional affiliations of the corresponding authors.

Furthermore, articles were categorized by access type ("Open Access" vs. "Subscription-based") and content focus ("Methodological/Technical" vs. "Clinical Application"). The complete list of the 100 articles included in the study is provided in Supplementary File 1.

2.5. Statistical Analysis, Bibliometric Mapping, and Visualization

Statistical analyses were performed using Python 3.12 with pandas (v2.0.0), scipy (v1.12.0), and numpy (v1.26.3) libraries. Given the non-normal distribution of bibliometric data (Shapiro-Wilk test, $p < 0.05$), non-parametric tests were employed. The relationship between scientific impact (citations) and social impact (Altmetric scores) was assessed using Spearman’s rank correlation coefficient. To ensure the robustness of the correlation estimate, the 95% Confidence Interval (CI) was calculated using the

bootstrap method with 2,000 resamples (random seed = 42). Group comparisons (e.g., Open Access vs. Closed Access) were conducted using the Mann-Whitney U test; we report U statistic, p-value, and effect size (Cliff's delta and r). Statistical significance was defined as a two-sided p-value < 0.05. No multiple comparison correction was applied as the analyses were exploratory and hypothesis-generating rather than confirmatory.(9–11).

Scientific mapping techniques were employed to reveal the conceptual structure, research trends, and inter-topic relationships. For the construction and visualization of bibliometric networks, VOSviewer (Version 1.6.19, Leiden University, The Netherlands)(12) and Python-based network analysis libraries (NetworkX)(13) were used.

The analysis process involved the following steps:

- **Data Standardization**: To ensure data consistency, synonymous or variably spelled terms in the "Author Keywords" column (e.g., "Artificial Intelligence" and "AI", "Large Language Models" and "LLM", "Chat-GPT" and "ChatGPT") were merged under a single standardized term.
- **Co-occurrence Analysis**: Keyword co-occurrence analysis was applied to determine the intellectual structure of the AI literature. To reduce noise in the network map and identify dominant trends, a minimum occurrence threshold was set, and only the most frequently used concepts were included.
- **Visualization Parameters**: In the generated network map, node size represents the frequency of the keyword, the lines between nodes indicate the frequency of co-occurrence in the same article, and the colors represent thematic clusters determined by the clustering algorithm.

The normality of continuous variables (citation counts and AAS) was examined using the Shapiro-Wilk test, which indicated that the data were not normally distributed ($p < 0.05$). Consequently, the Spearman rank correlation coefficient was used to determine the direction and strength of the relationship between academic citations and altmetric scores. Differences in citation and altmetric scores based on open access status and article type (review vs. research) were analyzed using the Mann-Whitney U test.

For all statistical tests, the significance level was set at $p < 0.05$.

2.6. Ethical Considerations

As this study is a bibliometric analysis based on publicly available data obtained from the Web of Science and Altmetric databases and does not involve any human or animal subjects, it is exempt from ethics committee approval.

We used ChatGPT (OpenAI; version GPT-4o; accessed Jan 2026) and QuillBot Premium (accessed Jan 2026) for English editing and language polishing only. The authors retain responsibility for the content and interpretation of the data. Use of these tools complied with journal guidelines

3. Results

3.1. General Characteristics of the Dataset

The top 100 most-cited articles retrieved from the Web of Science Core Collection were included in the analysis. All included publications were released between 2023 and 2025, reflecting the early-stage impact of artificial intelligence and generative language models in medicine.

The median citation count of the articles was 226 (IQR: 170–307.5), while the median Altmetric Attention Score (AAS) was determined to be 82 (IQR: 16–396.5). Regarding accessibility, 92% (n = 92) of the articles were published under Open Access status, with the remaining 8% appearing in subscription-based journals.

In terms of document type, 65% (n = 65) were categorized as original research articles, and 35% (n = 35) were reviews. The distribution by year revealed a high concentration in 2023 (71%), followed by 2024 (27%) and 2025 (2%). The average number of authors per publication was calculated at 10.2.

3.2. Relationship Between Academic Impact and Societal Interest

The relationship between WoS citation counts, representing academic impact, and altmetric scores, representing societal interest, was examined using Spearman's rank correlation test. The analysis revealed a statistically significant, positive, yet weak correlation between the two variables ($r = 0.299$, $p = 0.0025$; 95% CI bootstrapped with 2,000 resamples [0.107, 0.457], $n = 100$). According to Cohen's (1988) effect size criteria, this correlation coefficient indicates a small-to-medium relationship (14).

This finding demonstrates that studies receiving high citations within the scientific community do not always garner a corresponding level of engagement on social media; conversely, studies that go "viral" on social platforms do not necessarily achieve the highest levels of academic citation success (Fig. 1).

3.3. Comparison of the Most Influential Articles

An examination of the most successful articles in terms of scientific and societal impact reveals two distinct patterns (Table 1 and Table 2). The study titled 'Performance of ChatGPT on USMLE' by Kung et al. stands at the pinnacle of academic impact with 2,193 citations, while simultaneously achieving high visibility on social media with an Altmetric score of 1,766.

Table 1
Top 5 Most Cited Articles in Academic Literature (Scientific Impact)

Rank	Article Title (Shortened)	Category	WoS Citations (Scientific Impact)	Altmetric Score (Social Impact)
1	Performance of ChatGPT on USMLE: Potential for AI-assisted medical education...(15)	Education / Exam	2193	1766
2	Large language models encode clinical knowledge(16)	Basic Science / AI	1719	1248
3	ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review...(17)	Review	1592	94
4	Comparing Physician and AI Chatbot Responses to Patient Questions...(18)	Clinical Comparison	1385	6388
5	Segment anything in medical images (19)	Technical / Imaging	1338	283

The study by Ayers et al. titled 'Comparing Physician and Artificial Intelligence Chatbot Responses' emerged as the most debated publication across social media and news outlets, reaching a peak Altmetric score of 6,388 (18). While its citation count (1,385) is also substantial, its societal impact is proportionally much more dominant than its academic influence. Similarly, the study titled 'Distinguishing features of long COVID' generated a massive public resonance with an Altmetric score of 4,181, yet remained further behind in academic citation rankings with 452 citations.

Table 2
Top 5 Most Popular Articles on Social Media and News (Social Impact)

Rank	Article Title (Shortened)	Focus Area	Altmetric Score (Social Impact)	WoS Citations (Scientific Impact)
1	Comparing Physician and AI Chatbot Responses to Patient Questions...(18)	Doctor vs. AI	6388	1385
2	Distinguishing features of long COVID identified through immune profiling(20)	Public Health / COVID	4181	452
3	A high-performance neuroprosthesis for speech decoding and avatar control(21)	Futurism / BCI	3683	229
4	Persistent complement dysregulation ... in active Long Covid(22)	Public Health / COVID	2945	259
5	Organ aging signatures in the plasma proteome track health and disease (23)	Aging / Prevention	2036	291

3.4. Analysis of Societal Engagement Sources

Table 3 compares the distribution of societal dissemination sources for the top three studies with the highest altmetric scores. The results indicate that societal interest is not monolithic; rather, three distinct "Engagement Profiles" emerge based on the study's subject matter:

- The "Hybrid/Viral" Profile (Physician vs. Chatbot): By addressing a popular and controversial question of universal concern—"Will AI replace doctors?"—this study garnered record-breaking attention across both news outlets (413 stories) and Twitter (5,158 posts). This profile represents rare instances where a scientific topic simultaneously penetrates the mainstream media agenda and triggers intensive public discourse. This tripartite classification—'Hybrid/Viral', 'Community-Driven', and 'Futuristic/Media-Centric'—demonstrates that societal engagement patterns vary systematically with research topic.
- The "Community-Driven" Profile (Long COVID): Despite receiving less than half the news coverage of the top-ranked article (191 stories), this study achieved the highest Twitter engagement with 6,108 tweets. This suggests that topics involving patient advocacy and diagnostic uncertainty, such as Long COVID, are primarily adopted and disseminated by social media communities rather than news organizations. This represents popularity amplified by the public rather than the media.
- The "Futuristic/Media-Centric" Profile (Neuroprosthesis / Avatar): This study, which enables a paralyzed patient to communicate via brain signals, remained relatively quiet on Twitter (only 668 tweets) but was featured in 409 news outlets, nearly matching the first-ranked article. This profile proves that futuristic technologies with high visual appeal (e.g., neural interfaces) are highly favored

as "Technology News" by traditional media (TV, newspapers), even if they do not spark extensive daily discussion among the general public on social platforms.

Table 3
Breakdown of Social Attention Sources for the Top 3 High-Scoring Papers

Rank	Paper Topic & Context (Descriptive Name)	News Mentions	Twitter (X) Posts	Altmetric Score	Dominant Source
1	Physician vs. Chatbot (Clinical Comparison)(18)	413	5,158	6,388	Mixed (Viral)
2	Long COVID (Immune Profiling)(20)	191	6,108	4,181	Social Media
3	Neuroprosthesis (Speech Decoding / Avatar)(21)	409	668	3,683	News Media

3.5. Journal and Country Performance

Analysis of the distribution of publications by journal reveals that Nature (n = 11) and NPJ Digital Medicine (n = 11) emerged as the most productive outlets. Nature stood at the pinnacle of academic impact with a median citation count of 445, while also maintaining its leadership in societal impact (Median AAS: 908).

A notable finding is that Science magazine, despite having the highest social media influence with a median Altmetric score of 913, trailed behind Nature in terms of academic citations (Median: 259). The clinically-oriented JAMA Network Open demonstrated a balanced and robust performance in both academic (Median Citations: 200.5) and social (Median AAS: 226.5) metrics. Conversely, the median Altmetric score for Scientific Reports was notably low at 8; this indicates that the high overall average for this journal was driven by a few viral articles rather than a widespread distribution of engagement across its publications.

When country impact is evaluated based on median values, the United States (USA) maintains its dominance in the field with 58 articles, exhibiting a balanced and high profile in both academic impact (Median Citation: 247.5) and social influence (Median AAS: 227). In contrast, studies originating from China have been unable to translate their success in the academic world (Median Citation: 204) into social media visibility; the median altmetric score for Chinese publications was found to be only 12.

This striking discrepancy suggests that while research from China possesses high scientific value, it does not achieve sufficient circulation within Western-centric social media networks and news outlets—a phenomenon that may be characterized as 'digital isolation.' Furthermore, Canada (Median Citation: 271, Median AAS: 369), despite a smaller number of publications (n = 7), emerged as one of the countries reaching the highest median values in both domains.

3.6. Thematic Analysis

An impact analysis conducted via author keywords reveals that academia and the general public prioritize distinct subjects.

Academic Focus: Studies featuring the keywords "Ethics" (Median Citations: 309.5) and "Machine Learning" (Median Citations: 307) stand at the forefront of scientific impact. This indicates that the academic community prioritizes not only the technical capabilities of AI but also its ethical boundaries and methodological foundations.

Societal Focus: Contrary to mean values, an examination of median values shows that the standard social media impact of popular topics such as "Chatbot" (Median AAS: 30.5) and "Large Language Models" (Median AAS: 28.5) does not differ significantly from more technical subjects (e.g., Ethics: 34). This critical finding suggests that social media interest is not necessarily driven by a specific topic itself, but rather focuses on individual articles with sensational qualities within those topics.

"ChatGPT" remains the most balanced and central term in the literature, characterized by both high academic citation rates (Median: 248) and consistent social visibility (Median AAS: 30).

The impact analysis conducted according to Web of Science categories (Fig. 5) reveals a sharp divergence in character across different disciplines.

The 'Medicine, General & Internal' category emerged as the most influential domain in the literature, with a median citation count of 255 and a median Altmetric score of 200. This trend indicates that AI discussions have evolved beyond being purely technical subjects and have shifted toward the axes of clinical practice and public health.

In stark contrast, studies categorized under 'Computer Science' were found to have a median Altmetric score of 0 (zero). While publications in the computer science field continue to receive academic citations (Median: 163), they failed to achieve any significant visibility across social media platforms or news outlets.

3.7. The Impact of Open Access

While the prevailing expectation in the literature is that open access articles achieve higher visibility, the data obtained in this study present a different picture (Fig. 6).

A Mann-Whitney U test was conducted to determine if Open Access status influenced social impact. The analysis revealed no statistically significant difference in Altmetric scores between Open Access (n = 92, Mdn = 30.5) and Closed Access (n = 8, Mdn = 14.5) articles (U = 234.0, p = 0.090). The effect size was small to medium (Cliff’s δ = -0.364), indicating that accessibility alone was not the primary driver of social attention in this high-impact dataset. The difference approached but did not reach statistical significance (p = 0.090).

3.8. Conceptual Structure and Trend Analysis

The keyword map generated using the VOSviewer algorithm (Fig. 7) demonstrates that the literature is concentrated along two primary axes: the technical infrastructure (blue cluster) and social/clinical applications (red cluster).

4. Discussion

The present study stands as one of the first comprehensive bibliometric investigations to systematically compare the academic (citations) and societal (Altmetric) impact of the top-cited AI research in medicine from the generative AI era (2023–2026). Our findings reveal a statistically significant but weak correlation between academic impact and societal interest. This result corroborates the thesis that "scientific quality and popularity do not always coincide," indicating the existence of "two distinct audiences" (scholars and the general public) in AI research. Similarly, Karabay et al. (2024) found no significant correlation between citations and altmetric scores in their analysis of dental literature (24). This suggests that, independent of the discipline, scholarly accumulation (citations) and societal awareness (Altmetrics) constitute two separate universes governed by distinct dynamics. Khan et al. (2022) previously warned the literature that even publications lacking scientific validity could garner record-level attention on social media, emphasizing that Altmetric scores may not mirror scientific rigor (25). Our findings validate this divergence on a different scale: although all articles in our dataset were high-cited (high-quality) works, this academic success did not consistently translate into high social impact. As noted by Thelwall (2021), altmetric indicators measure "societal curiosity" and "media visibility" rather than intrinsic scientific merit (26).

The distribution of publications revealed a high concentration in 2023 (71%), followed by 2024 (27%) and 2025 (2%). This trend indicates that academic interest in generative artificial intelligence technologies reached its pinnacle specifically in 2023. Furthermore, the average number of authors per article was calculated at 10.2, reflecting the field’s multidisciplinary nature (combining engineering and medicine) and its reliance on multi-center collaboration.

In a comprehensive bibliometric analysis conducted by Wu et al. (2024), it was reported that approximately 41% of the ChatGPT-related literature consisted of 'Letters to the Editor' or opinion pieces. This finding serves as robust evidence that the field was initially in a phase of speculation and 'early hype' (27).

In the present study, to eliminate the 'noise' identified by Wu et al. (2024) and to analyze the scientific backbone of the literature, the dataset was methodologically filtered to include only the top 100 most-cited works categorized as 'Original Articles' or 'Reviews.' Consequently, this research represents the 'evidence-based' facet of the AI literature with the potential to shape clinical practice, rather than its purely 'popular' aspect. Our results indicate that this 'elite' cluster resonates within society not only through academic citations but also through high social media engagement (Median AAS: 82) (27).

Our analysis reveals a structural, rather than merely thematic, divergence between the priorities of academia and the general public (Fig. 4). When examining median citation values, concepts such as

'Ethics' (Median Citation: 309.5) and 'Machine Learning' (Median Citation: 307) were found to exert a significantly greater scientific impact than the popular 'Chatbot' studies (Median Citation: 239). Conversely, an intriguing paradox was observed on the social media front. Although the mean altmetric scores for 'Chatbot' and 'ChatGPT' terms were exceedingly high, their median scores (30.5 and 30, respectively) were nearly identical to technical subjects like 'Machine Learning' (Median AAS: 31).

This critical finding proves that the public does not engage with every study on chatbots; rather, interest is reserved for studies with a specific narrative. The most concrete example of this selective interest is the study by Ayers et al. (2023). The reason this paper stood out from other chatbot research to garner record-breaking attention (AAS > 6,000) was not just its technical analysis, but its focus on the competitive inquiry: 'Who is better: the Physician or Artificial Intelligence?' (18). Consequently, our data confirm that while academia prioritizes 'conceptual depth,' society grants a premium to 'sensational competition.' Social media tends to respond with sensational fervor to competitive narratives such as 'Physician vs. Artificial Intelligence' (Ayers et al., 2023). In contrast, the fundamental vision of the academic literature is built upon the 'convergence' of human and artificial intelligence—a concept defined by Topol (2019) as 'High-Performance Medicine (28).

The academic community's intense focus on ethical issues can be linked to the striking findings of Obermeyer et al. (2019), which exposed structural biases within healthcare algorithms. This underscores a fundamental scholarly stance: academia prioritizes the safety and reliability of technology over its mere speed of development (29).

When examining the shifts within the literature, bibliometric studies covering the pre-generative AI era (Guo et al., 2020) show that the focus was predominantly concentrated on technical metrics such as 'algorithm performance,' 'disease diagnosis,' and 'image processing.' Guo et al. (2020) identified that while the AI literature in healthcare was growing rapidly, the majority of studies remained limited to theoretical modeling rather than clinical integration (30).

In contrast, our current study, which covers the post-ChatGPT era, observes that the axis has shifted from technical performance toward 'social interaction' (e.g., Chatbot vs. Physician) and 'ethical discourse.' The steady growth identified in the analysis by Guo et al. (2020) has been superseded by an explosive popularity as of 2023. This phenomenon proves that artificial intelligence is no longer confined to the engineers' laboratories but has become central to daily public discourse (30).

A striking finding in our study is that while publications originating from China (n = 16) achieved high academic citation counts, they exhibited statistically significant lower Altmetric Attention Scores (AAS) compared to studies from the USA and the UK. This discrepancy stems from the structural limitations of current altmetric data aggregators (e.g., Altmetric.com) rather than a lack of societal impact by Chinese research.

The Altmetric.com algorithm comprehensively tracks Western-centric platforms such as Twitter (X), Facebook, and Reddit; however, it either fails to index or provides very limited coverage of platforms like

WeChat and Weibo, which serve as the primary communication channels for the Chinese academic ecosystem (31). Consequently, the substantial social impact generated by Chinese researchers within their local digital ecosystems is rendered 'invisible' by Western-centric metrics. This situation creates a significant risk of 'Data Coverage Bias' in bibliometric analyses, leading to the systematic underestimation of the societal influence of publications originating from the Global South or East Asia. This 'digital isolation' is not a reflection of low societal interest in China, but rather a manifestation of data coverage bias inherent in current altmetric aggregators, which inadequately index major Chinese platforms like WeChat and Weibo.

Furthermore, the limited social impact of articles in the Computer Science category, despite their academic citations, is directly related to our methodological choices. During the construction of the dataset, we prioritized Medicine and Health Sciences journals that directly influence clinical practice over pure engineering journals.

Consequently, the studies labeled as 'Computer Science' in our analysis do not represent internal theoretical debates within the engineering community, but rather interdisciplinary works published in medical journals. It is expected that such technically dense articles are shared less by medical journal readers and the general public compared to articles with direct clinical implications, such as 'Chatbot vs. Physician.' This finding proves that while a 'technical infrastructure' is an academic necessity in the field of AI in healthcare, the primary factor determining 'social visibility' is the clinical and human narrative.

The conceptual mapping analysis of our study (Fig. 7) demonstrates that the intellectual structure of the literature is distinctly clustered along two main axes. At one end of the map lies the technical infrastructure—represented by concepts such as 'Machine Learning' and 'Neural Networks'—while the other end hosts the social and clinical application areas represented by 'ChatGPT,' 'Ethics,' and 'Education.' The gradual displacement of technical terms by concepts focused on 'social impact' over the years provides the clearest visual evidence that AI literature has moved beyond the laboratory stage and onto a platform of societal debate.

Contrary to expectations, the finding that Open Access (OA) articles had lower social impact scores than subscription-based articles demonstrates that 'access to content' alone is insufficient. The fact that subscription-based articles published in high-prestige journals like Nature, Science, or NEJM are more frequently reported by global news networks proves that 'Journal Branding' is a more potent visibility factor than mere accessibility. This finding aligns with studies showing that journal prestige can override accessibility in driving media attention.

4.1. Practical Implications and Conclusion

The findings of this study offer significant implications for various stakeholders:

Researchers: To enhance both the academic and societal impact of AI research, authors should focus on clinical and human narratives alongside technical content and improve their science communication

competencies such as engaging with science journalists, using plain language summaries, and actively disseminating findings on multiple social media platforms.

Funding Agencies: Multi-dimensional evaluation approaches should be adopted rather than relying on a single metric when assessing research impact.

Journal Editors: It should be considered that the brand value of high-prestige journals is a stronger driver of visibility than accessibility.

Policy Makers: Given that public debates regarding AI applications in health are often shaped by sensational competition, mechanisms should be established to support the flow of evidence-based information.

5. Limitations of the Study

Despite its comprehensive scope, this study has certain limitations that should be acknowledged. First, the dataset is restricted to the Web of Science (WoS) database; consequently, preprint servers such as arXiv, which are extensively utilized in the field of artificial intelligence, were not included in the analysis. Second, Altmetric data are subject to external factors, including API policy changes by social media platforms (notably Twitter/X) and instantaneous fluctuations in online popularity. Altmetric scores are platform-dependent and subject to temporal fluctuations; they may not fully capture the nature (positive vs. negative) of online attention. Finally, the analysis was limited to English-language publications, which may have resulted in the omission of societal debates occurring in local languages.

6. Conclusion

This study provides quantitative evidence of a distinct divergence between "academic significance" and "societal excitement" within the medical artificial intelligence literature. While the academic community focuses on how AI works (ethics, methodology), the general public and media concentrate on whom AI will replace (e.g., the "Physician vs. Chatbot" narrative).

The findings underscore that researchers and funding agencies should avoid relying on a single metric—whether exclusively citations or social media scores—when evaluating the impact of a study. Especially for AI applications involving public health, it is of critical importance that researchers enhance their science communication competencies to ensure that scientifically rigorous studies are accurately disseminated to the public. Future research should utilize long-term data to monitor whether the current surge in popularity translates into lasting scientific impact. As AI continues to reshape medicine, bridging the gap between academic rigor and public engagement will be essential for translating innovation into real-world health benefits.

The original dataset analyzed in this study was retrieved from the Web of Science Core Collection. The processed dataset used for the final analysis, including citation counts and Altmetric scores (retrieved in

January 2026), is provided as Supplementary File 1. The Python code used for data preprocessing, statistical analysis, and VOSviewer file preparation is available as Supplementary File 2.

All data and code are also openly available in the Zenodo repository at [https://doi.org/10.5281/zenodo.18474927].

Declarations

Author Contribution

F.Ş. and C.Ç. contributed equally to all stages of this study. Both authors were involved in the study's conceptualization, methodology design, and data collection. F.Ş. and C.Ç. performed the bibliometric and altmetric data analysis, prepared the figures and tables, and wrote the main manuscript text. All authors reviewed and approved the final version of the manuscript.

Data Availability

The original dataset analyzed in this study was retrieved from the Web of Science Core Collection. The processed dataset used for the final analysis, including citation counts and Altmetric scores (retrieved in January 2026), is provided as Supplementary File 1. The Python code used for data preprocessing, statistical analysis, and VOSviewer file preparation is available as Supplementary File 2.All data and code are also openly available in the Zenodo repository at [https://doi.org/10.5281/zenodo.18474927].

References

1. Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Aleman FL, et al. Gpt-4 technical report. arXiv preprint arXiv:230308774. 2023.
2. Cai Y, Deng Q, Lv T, Zhang W, Zhou Y. Impact of GPT on the Academic Ecosystem. Science & Education. 2025;34(2):913 – 31.
3. Priem J, Taraborelli D, Groth P, Neylon C. Altmetrics: A manifesto. 2011.
4. Williams AE. Altmetrics: an overview and evaluation. Online information review. 2017;41(3):311-7.
5. Kolahi J, Khazaei S, Iranmanesh P, Kim J, Bang H, Khademi A. Meta-Analysis of Correlations between Altmetric Attention Score and Citations in Health Sciences. BioMed research international. 2021;2021:6680764.
6. Byram JN, Lazarus MD, Wilson AB, Brown KM. Could the altmetrics wave bring a flood of confusion for anatomists? Anatomical sciences education. 2023;16(4):600-9.
7. Web of Science Core Collection: Clarivate; 2024 [cited 2026 01]. Available from: https://www.webofscience.com.

8. Adie E, Roe W. Altmetric: enriching scholarly content with article-level discussion and metrics. Learned Publishing. 2013;26(1):11 – 7.
9. McKinney W. Data structures for statistical computing in Python. scipy. 2010;445(1):51 – 6.
10. Waskom ML. Seaborn: statistical data visualization. Journal of open source software. 2021;6(60):3021.
11. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature methods. 2020;17(3):261 – 72.
12. Van Eck N, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. scientometrics. 2010;84(2):523 – 38.
13. Hagberg A, Swart PJ, Schult DA. Exploring network structure, dynamics, and function using NetworkX. Los Alamos National Laboratory (LANL); 2007.
14. Cohen J. Statistical power analysis for the behavioral sciences: routledge; 2013.
15. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLoS digital health. 2023;2(2):e0000198.
16. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature. 2023;620(7972):172 – 80.
17. Sallam M, editor ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare; 2023: MDPI.
18. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA internal medicine. 2023;183(6):589 – 96.
19. Khandakar S, Al Mamun M, Islam M, Hossain K, Melon M, Javed M. Unveiling early detection and prevention of cancer: Machine learning and deep learning approaches. Educational Administration: Theory and Practice. 2024;30(5):14614-28.
20. Klein J, Wood J, Jaycox JR, Dhodapkar RM, Lu P, Gehlhausen JR, et al. Distinguishing features of long COVID identified through immune profiling. Nature. 2023;623(7985):139 – 48.
21. Metzger SL, Littlejohn KT, Silva AB, Moses DA, Seaton MP, Wang R, et al. A high-performance neuroprosthesis for speech decoding and avatar control. Nature. 2023;620(7976):1037-46.
22. Cervia-Hasler C, Brüningk SC, Hoch T, Fan B, Muzio G, Thompson RC, et al. Persistent complement dysregulation with signs of thromboinflammation in active Long Covid. Science. 2024;383(6680):eadg7942.
23. Oh HS-H, Rutledge J, Nachun D, Pálovics R, Abiose O, Moran-Losada P, et al. Organ aging signatures in the plasma proteome track health and disease. Nature. 2023;624(7990):164 – 72.
24. Karabay F, Demirci M, Tuncer S, Tekçe N, Berkman M. A bibliometric and Altmetric analysis of the 100 top most cited articles on dentin adhesives. Clinical oral investigations. 2024;28(1):92.

25. Khan H, Gupta P, Zimba O, Gupta L. Bibliometric and altmetric analysis of retracted articles on COVID-19. Journal of Korean Medical Science. 2022;37(6).
26. Thelwall M. Measuring societal impacts of research with altmetrics? Common problems and mistakes. Journal of economic surveys. 2021;35(5):1302-14.
27. Wu J, Ma Y, Wang J, Xiao M. The application of ChatGPT in medicine: a scoping review and bibliometric analysis. Journal of Multidisciplinary Healthcare. 2024:1681-92.
28. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nature medicine. 2019;25(1):44–56.
29. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447 – 53.
30. Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial intelligence in health care: bibliometric analysis. Journal of medical Internet research. 2020;22(7):e18228.
31. Griffiths J. The great firewall of China. 2021.

Figures

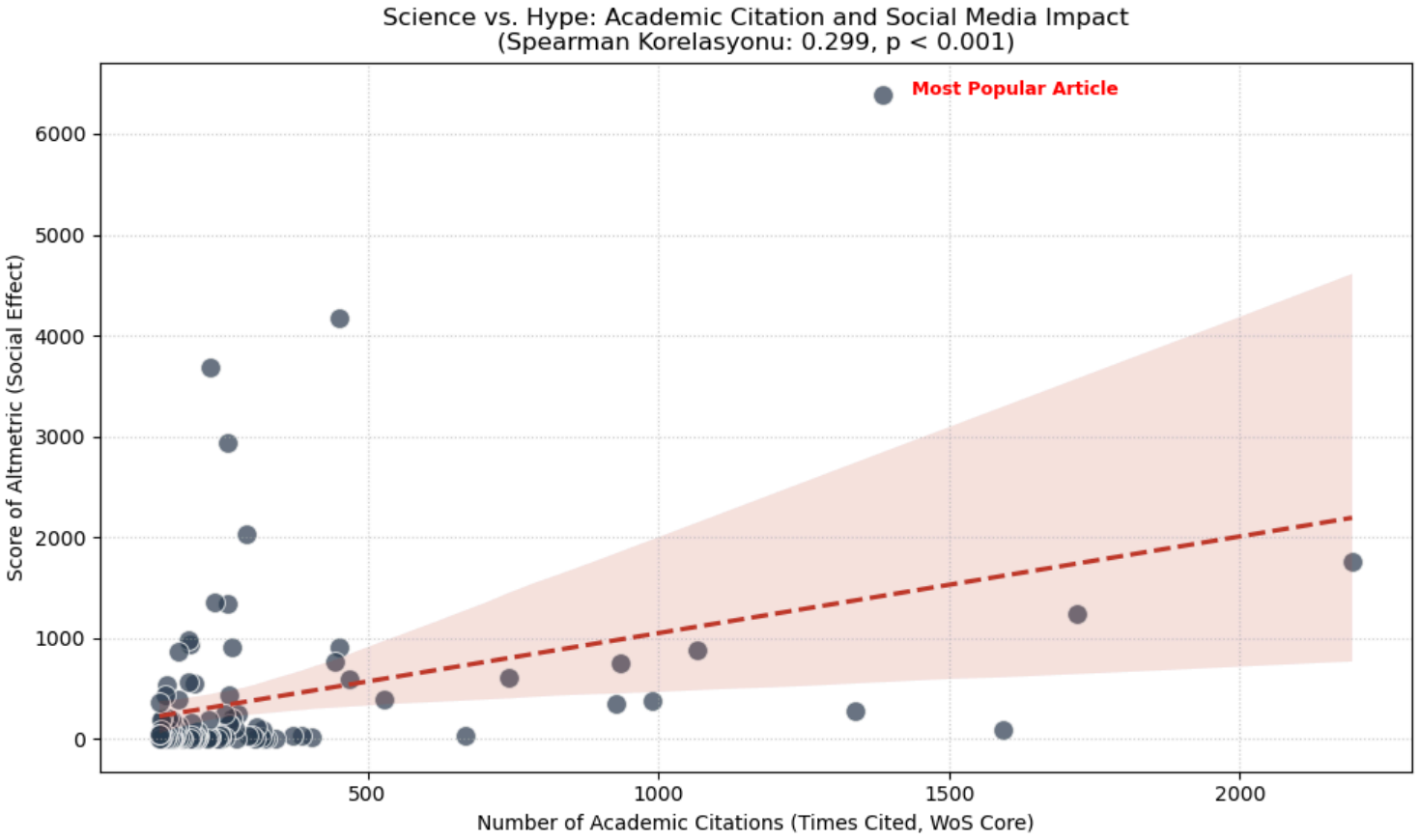


Figure 1

Comparing Academic Citation and Social Media Impact (Scatter plot illustrating the relationship between citation counts and Altmetric Attention Scores. Each data point represents an individual article (n=100).

The dashed line indicates the Spearman correlation trendline (r=0.299, p=0.0025). Points highlighted in red represent outliers characterized by 'high citation-low AAS' and 'low citation-high AAS' profiles. Notable outliers include Ayers et al. (high AAS, moderate citations) and Sallam et al. (high citations, low AAS))

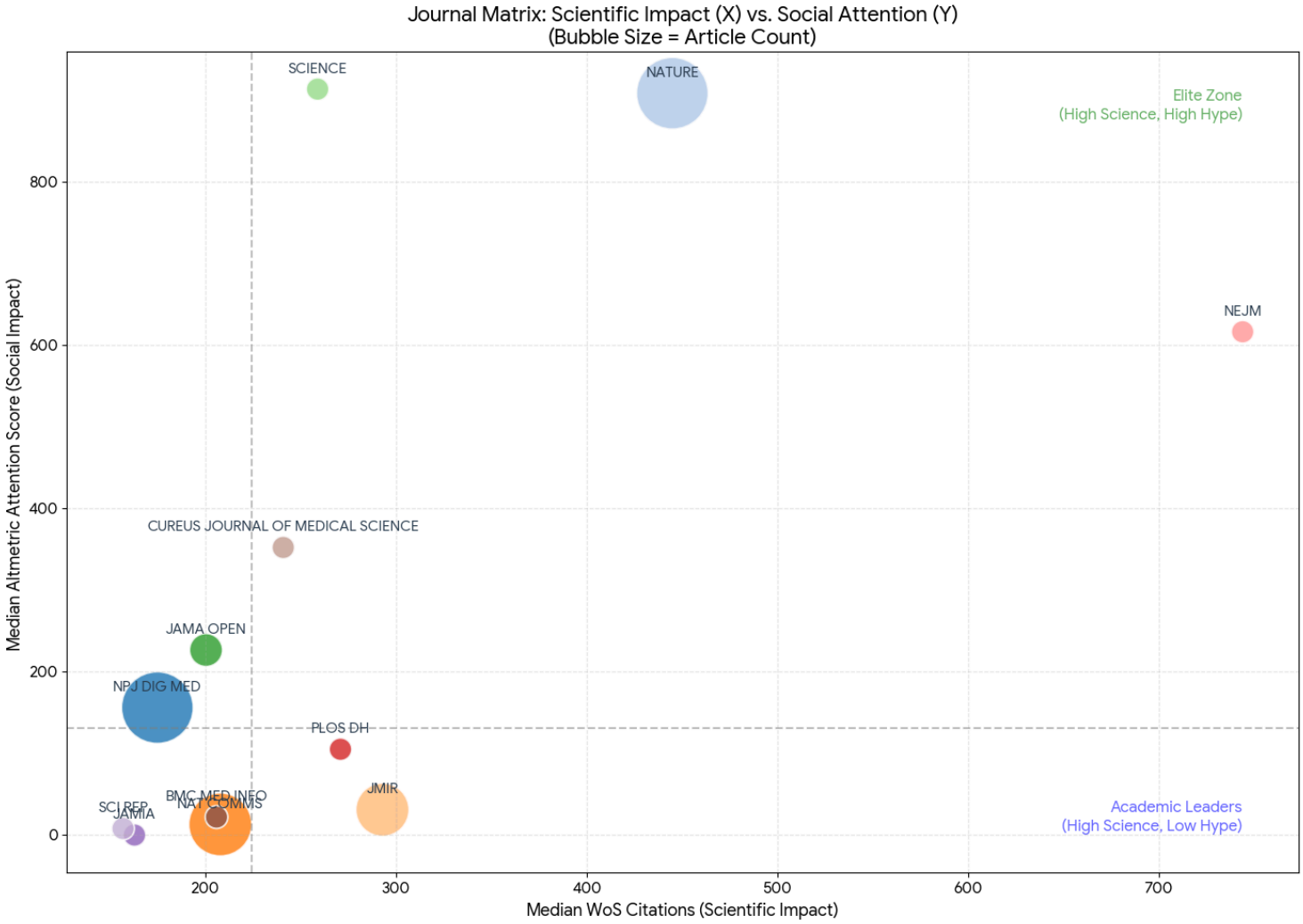


Figure 2

Scientific and Social Impact Distribution by Top Journal (The plots show median citation counts and median AAS for journals with at least 5 articles)

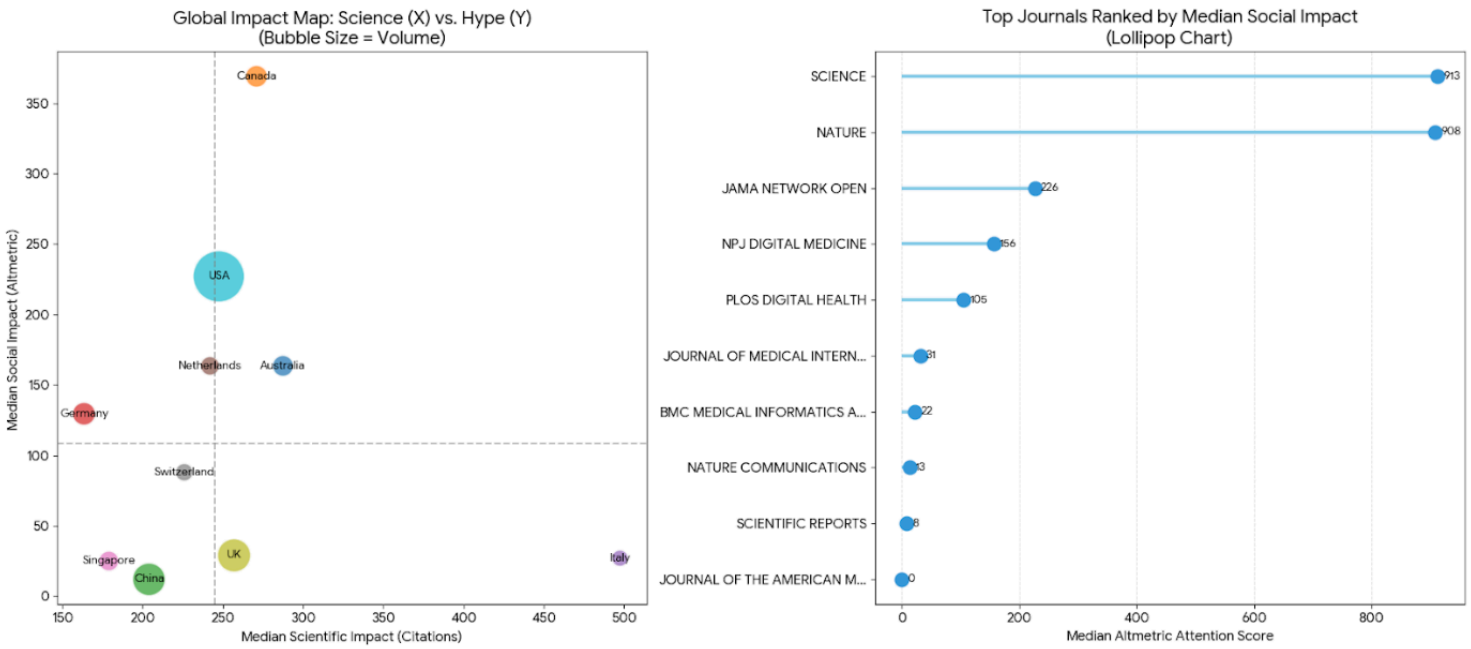


Figure 3

Social and Scientific Impact Distribution by Top Countries (The plots show median citation counts and median AAS for countries with at least 3 articles)

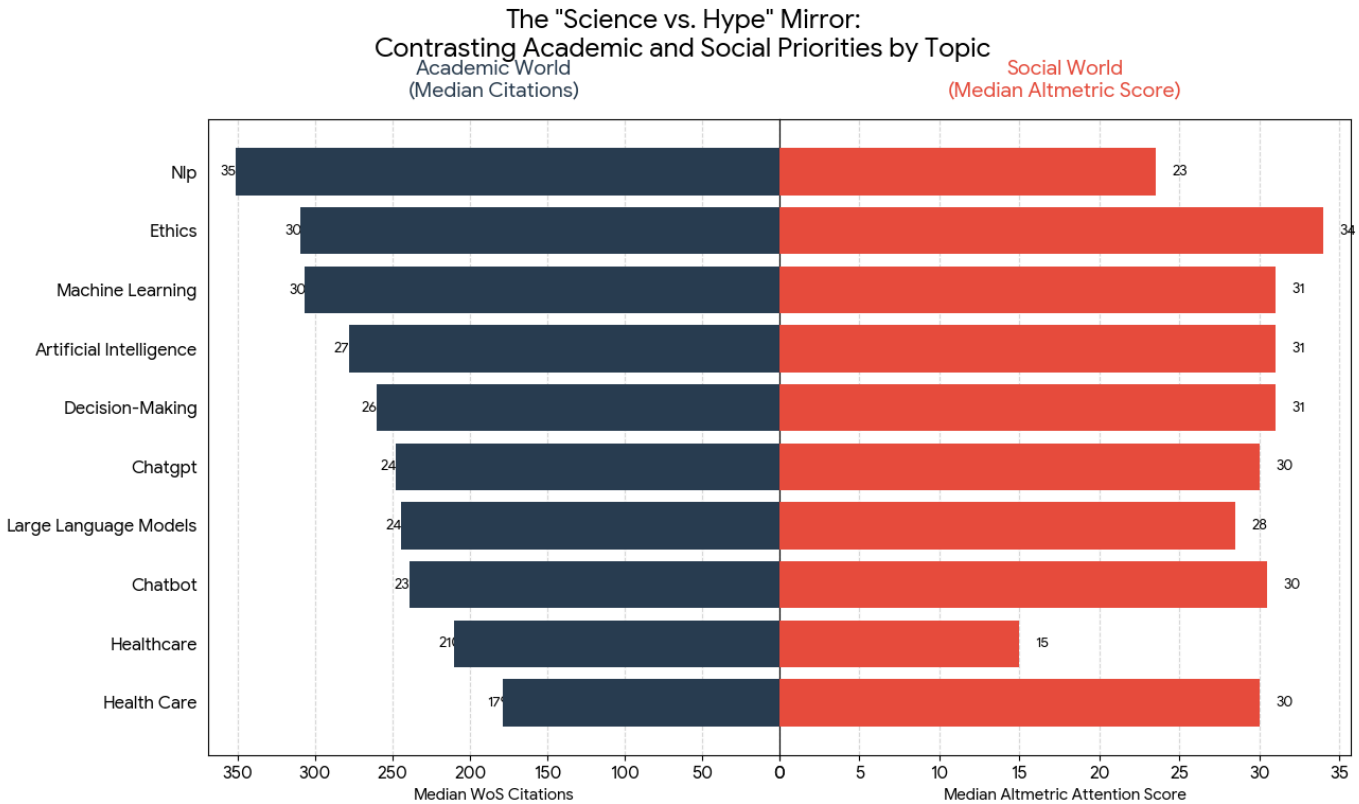


Figure 4

Comparing Keywords (Median citation counts and median AAS for selected keywords)

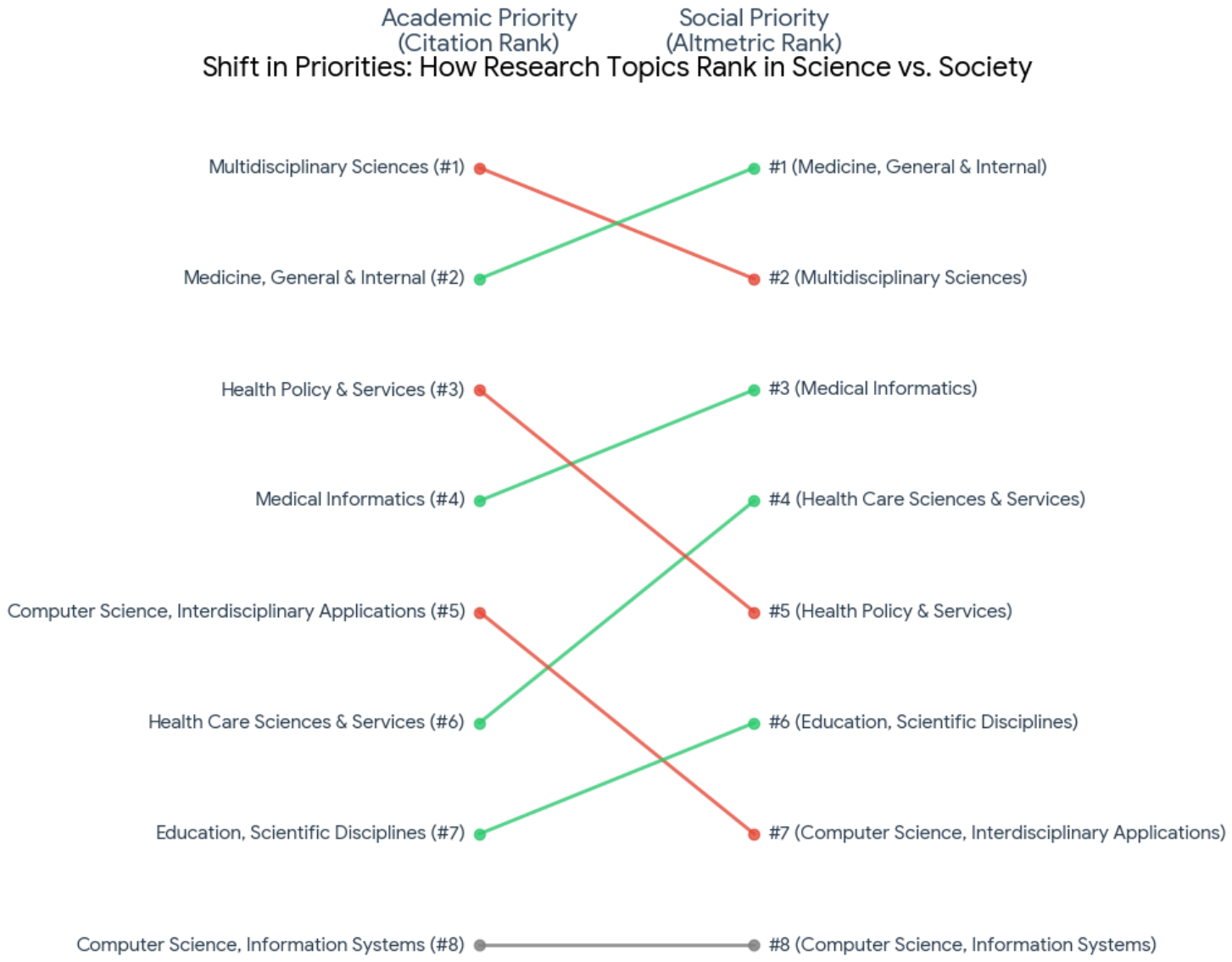


Figure 5

Shift in Priorities (Median citation counts and median AAS by Web of Science category)

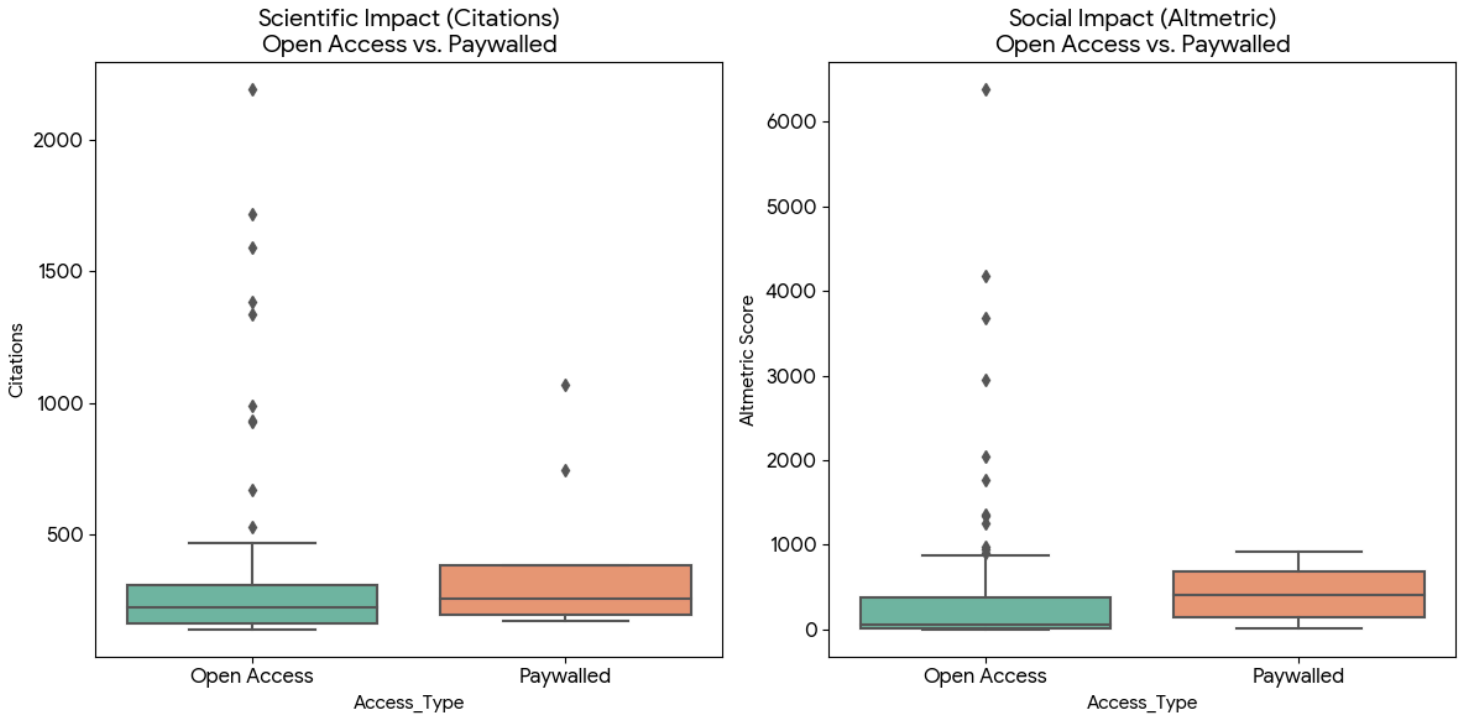


Figure 6

Impact Comparison based on Access Type

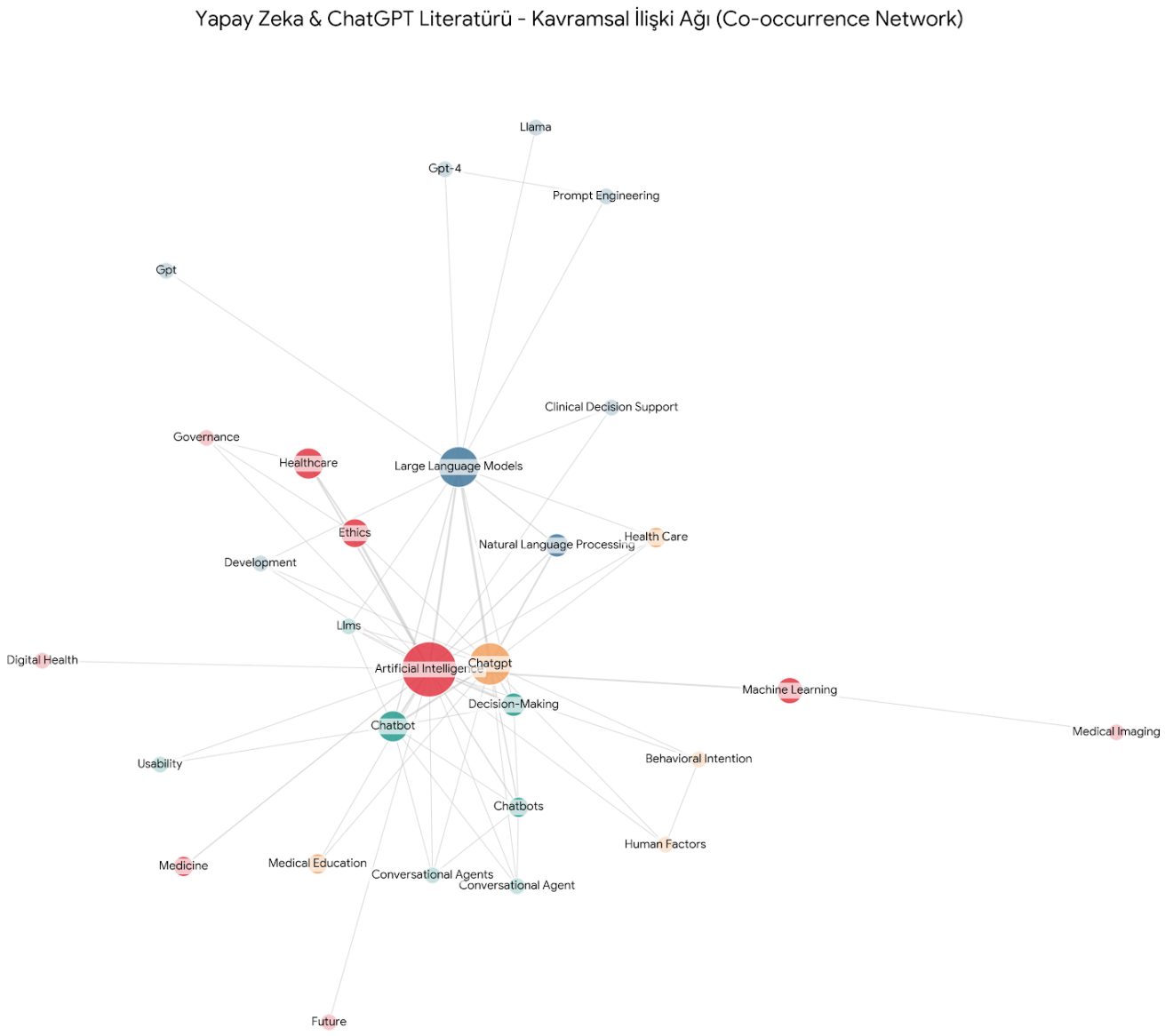


Figure 7

Keyword Co-occurrence Network illustrating the conceptual structure of the artificial intelligence literature. The size of the nodes represents the frequency of keywords within the dataset, while the thickness of the links between nodes indicates the strength of their co-occurrence. Colors signify the algorithmically determined clusters (Blue/Green Cluster: concepts focused on technical and engineering aspects; Red/Yellow Cluster: concepts focused on clinical application and societal impact).