International Journal of Essential Competencies in Education
https://journal-center.litpam.com/index.php/ijece/index

June 2025 Vol. 4, No. 1
e-ISSN: 2962-2131
pp. 1-14

Charting the Future of Prompt Engineering: Critical Reflections on Methodology, Ethics, and Research Directions

Ni Nyoman Sri Putu Verawati [1]*, Lalu Nurul Yaqin [2]

[1] Physics Education Department, University of Mataram, Mataram, Indonesia.
[2] Faculty of Arts and Social Sciences, Universiti Brunei Darussalam, Gadong, Brunei Darussalam
*Correspondence: veyra@unram.ac.id

Article Info

Article History
Received: May 2025;
Revised: June 2025;
Published: June 2025

Keywords
Prompt engineering;
Large language model;
Bibliometric methodology;
Ethical and governance frameworks;
Future research directions

Abstract

Prompt engineering has emerged as a transformative strategy for optimizing Large Language Models (LLMs), offering a cost-effective alternative to full model fine-tuning. In a recent bibliometric review, Fatawi et al. (2024) analyzed 437 Scopus-indexed publications from January 2022 to February 2024, using VOSviewer to identify key thematic clusters—including transformer architectures, deep learning innovations, and few-shot learning—and documenting a fivefold increase in related publications over the review period. Building on their macro-level mapping, this commentary extends the discussion by articulating the strategic and democratizing potential of prompt engineering while addressing critical gaps in methodology and ethical oversight. We critique the review's reliance on a single English-language database, its exclusion of preprints and non-English sources, and its omission of qualitative insights into user practices and system impacts. In response, we offer concrete recommendations to guide future research: diversify data sources for bibliometric analysis, implement rigorous prompt audit frameworks, conduct longitudinal A/B testing in real-world environments, and adopt mixed-methods approaches to capture human-centered dynamics. We also explore emerging synergies—such as quantum-enhanced NLP and neuro-linguistic prompt design—as promising frontiers for advancing prompt optimization. By addressing these gaps, this commentary aims to ensure that prompt engineering evolves not only as a technical solution but as a responsible and inclusive foundation for next-generation AI development.

https://doi.org/10.36312/ijece.v4i1.2908

How to Cite
Verawati, N. N. S. P., & Yaqin, L. N. (2025). Charting the Future of Prompt Engineering: Critical Reflections on Methodology, Ethics, and Research Directions. *International Journal of Essential Competencies in Education*, *4*(1), 1–14. https://doi.org/10.36312/ijece.v4i1.2908

INTRODUCTION

The past few years have witnessed an unprecedented surge in the capabilities of Large Language Models (LLMs), catalyzed by advancements in transformer architectures and large-scale pre-training. As organizations increasingly deploy LLMs in domains ranging from automated customer service to biomedical literature synthesis, a central challenge has emerged: how to reliably generate high-quality outputs without incurring the substantial computational costs, time, and domain-specific data required for full model fine-tuning. Prompt engineering—the practice of crafting input queries to elicit desired model behaviors—

has thus emerged as a promising, low-resource alternative to fine-tuning. Its appeal lies in its accessibility and adaptability, allowing non-specialist users to leverage powerful language models across diverse applications.

In their article "Empowering Language Models Through Advanced Prompt Engineering: A Comprehensive Bibliometric Review," Fatawi et al. (2024) provide a valuable macro-level mapping of the field. Analyzing 437 Scopus-indexed publications from January 2022 to February 2024 using VOSviewer, they identify major thematic clusters—such as deep learning innovations, transformer architectures, and few-shot learning—and trace the exponential growth of research activity in prompt engineering. Their work successfully highlights institutional collaborations and topical trends, offering a panoramic view of the domain's rapid evolution.

However, Fatawi et al.'s analysis does not interrogate the methodological blind spots or socio-ethical ramifications associated with the proliferation of prompt engineering. Specifically, the review does not address key concerns such as the variability of prompt efficacy, issues of reproducibility, and the contextual fragility of prompt-based outputs in real-world applications. For example, recent studies in healthcare contexts reveal considerable uncertainty in the reliability of prompts. Gimeno et al. (2024) and Chen et al. (2024) both underscore how prompt phrasing can significantly alter LLM outputs, introducing instability in clinical decision-making simulations. Similarly, Skryd and Lawrence (2024) caution that limited user expertise in prompt design can exacerbate biases or reduce output validity, particularly in sensitive domains. These limitations call into question the assumption that prompt engineering is universally reliable or sufficient for high-stakes tasks.

Moreover, comparative studies reveal nuanced trade-offs between prompt engineering and traditional fine-tuning approaches. While prompt-based techniques have been shown to provide competitive results in low-resource settings or for rapid prototyping (Maharjan et al., 2024; Hauna et al., 2025), other investigations demonstrate that fine-tuning continues to outperform prompting in more complex tasks such as medical imaging analysis and multi-class classification (Botunac et al., 2024; Kim et al., 2025). These mixed results highlight the contextual nature of effectiveness and the necessity of method selection tailored to domain constraints and operational goals (Banda et al., 2025).

This commentary fills a critical void by extending the findings of Fatawi et al. (2024) through a conceptual and normative lens. We interrogate the limitations of bibliometric methodology, which, despite offering a bird's-eye view, often fails to capture the qualitative dimensions of prompt engineering practices, including human-computer interaction, reproducibility barriers, and the lived experiences of practitioners. As highlighted by Lai et al. (2020), Gurung et al. (2022), and Sacca et al. (2024), bibliometric analyses and AI research more broadly continue to grapple with constraints such as database bias, methodological heterogeneity, and the lack of standardized evaluation metrics—factors that undermine generalizability and hinder evidence-based policy design.

Accordingly, this commentary advances the discourse on prompt engineering by (1) elucidating its strategic and democratizing potential; (2) critically evaluating the methodological foundations and epistemological limits of current bibliometric analyses; (3) engaging with the ethical and governance challenges surrounding prompt deployment; and (4) offering targeted, actionable recommendations to shape a more responsible and inclusive research agenda. By situating prompt engineering within a broader socio-technical context, we aim to provide a complementary perspective that bridges empirical insight with conceptual depth.

STRATEGIC VALUE OF PROMPT ENGINEERING

Prompt engineering has emerged as a strategically significant and cost-effective approach for leveraging the capabilities of Large Language Models (LLMs), especially in contexts where computational resources and training data are limited. By carefully crafting inputs, practitioners can steer model outputs toward specific objectives, often achieving comparable results to fine-tuned models in few-shot or zero-shot scenarios. Studies such as those by Kim et al. (2023) and Saeed et al. (2023) demonstrate that prompt tuning can match or even exceed the performance of traditional fine-tuning in tasks such as political perspective detection, with minimal performance degradation and significantly lower resource demands. Wu (2024) further reinforces this perspective, highlighting the parameter efficiency and adaptability of prompt-based frameworks as compelling advantages over conventional model retraining methods.

This democratization of advanced AI capabilities has particular significance for smaller organizations—such as educational institutions, non-profits, and startups—that often lack access to the infrastructure required for full-scale fine-tuning. For instance, Nyaaba and Zhai (2024) emphasize that effective prompt design is crucial for educators aiming to leverage generative AI tools, while Heston and Khun (2023) point to the role of prompt engineering in mitigating biases in resource-constrained learning environments. Similarly, Liu et al. (2024) demonstrate how mobile-edge platforms, when paired with optimized prompts, can deliver scalable AI services without high overhead. These findings collectively underscore the role of prompt engineering as a critical enabler of inclusive AI adoption, equipping underserved sectors to integrate cutting-edge technologies with limited financial or technical investment (Wang et al., 2024).

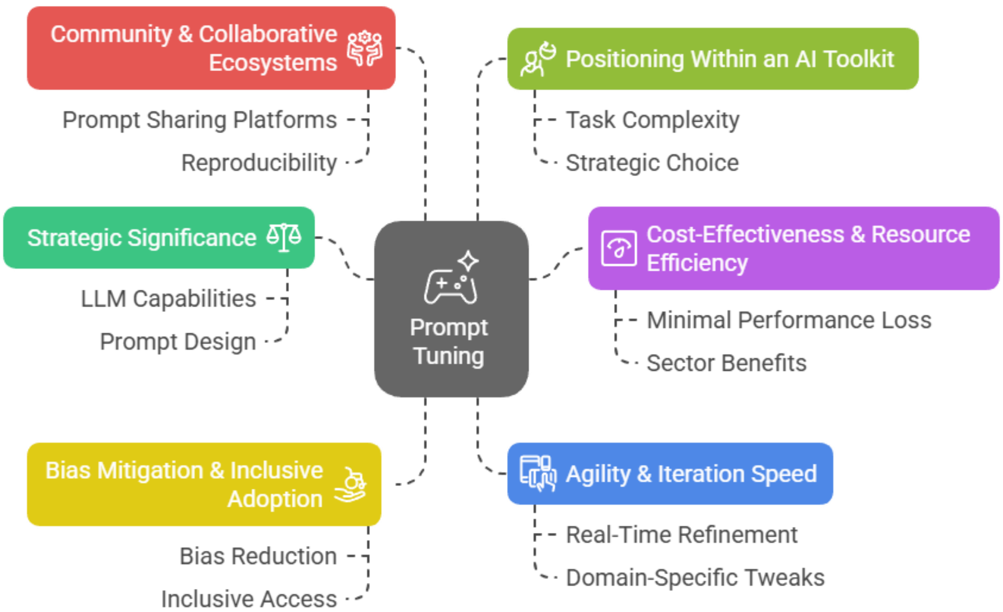

Figure 1. Strategic value of prompt engineering

In addition to expanding accessibility, prompt engineering fosters rapid experimentation and agile iteration. Unlike full fine-tuning cycles, which require substantial retraining, practitioners can iteratively refine prompts in real-time to adapt to dynamic requirements or new data formats. This agility accelerates innovation and supports domain-specific customization, especially in fields such as healthcare, education, and public administration. The growing emergence of collaborative ecosystems—such as PromptCraft (Vemprala et al., 2024) and PromptSource—has further amplified this agility by enabling practitioners to share, evaluate, and reuse effective prompts. These platforms encourage communal knowledge exchange and serve as living repositories of best practices, aligning with broader open science principles. While related work by Huang et al. (2024), Leung (2024), and Johnson-Eilola et al. (2024) emphasizes the value of structured instructional design and collaborative learning environments, PromptCraft represents a concrete example of how prompt sharing can enhance reproducibility and accelerate cross-disciplinary knowledge transfer.

These developments reinforce the strategic value of prompt engineering as both a technical methodology and a social practice when taken together. It empowers users with limited resources to deploy AI systems effectively while enabling rapid innovation cycles supported by community-driven prompt development. However, this potential must be contextualized within specific domains, as performance outcomes may still vary based on task complexity and user expertise. Therefore, prompt engineering should be seen not as a panacea, but as a critical component within a broader toolkit of adaptive, efficient, and inclusive AI deployment strategies.

METHODOLOGICAL STRENGTHS AND LIMITATIONS

The bibliometric methodology employed by Fatawi et al. (2024)—analyzing 437 Scopus-indexed documents using VOSviewer—offers a valuable quantitative snapshot of scholarly output in the rapidly expanding field of prompt engineering. The use of keyword co-occurrence networks and collaboration mapping enables the identification of thematic clusters, such as deep learning, few-shot learning, and transformer architectures, while also visualizing institutional and geographic research hubs. These visualizations provide a helpful macro-level overview of who is contributing to the field and what subtopics are gaining traction.

However, the exclusive reliance on Scopus introduces several methodological limitations that compromise both the comprehensiveness and generalizability of the findings. A growing body of research points to biases inherent in using Scopus as a single-source database. For instance, Kumar et al. (2023) and Tan et al. (2023) demonstrate that Scopus and Web of Science (WoS) differ significantly in their citation coverage, indexing practices, and journal selections—differences that can distort the perceived prominence of research trends. Additionally, Adeosun (2024) notes that Scopus has a slower indexing pace compared to some preprint repositories, resulting in the exclusion of emerging or rapidly disseminated research. These concerns echo earlier critiques by Wilder and Walters (2021) and Sutar et al. (2024), who advocate for multi-database strategies in bibliometric studies to avoid regional and disciplinary omissions. In the context of prompt engineering, this bias may disproportionately underrepresent contributions from non-English, open-access, or regionally focused research outlets, thereby skewing interpretations of the field's development.

Equally important is the limited epistemological scope of traditional bibliometric analyses, which tend to excel at identifying "what" and "where" but often neglect the "how" and "why." While Fatawi et al. (2024) successfully map topical clusters, their review does not unpack the methodological rigor, reproducibility, or practical deployment contexts of the studies they identify. This gap is particularly consequential in areas such as prompt engineering, where empirical performance can vary significantly based on task type, prompt design, and user expertise. Integrating qualitative inquiry—such as content analysis, expert interviews, or case studies—would help surface insights about prompt formulation strategies, evaluation standards, and domain-specific implementation challenges.

Indeed, the value of mixed-method bibliometric studies has been demonstrated across diverse fields. For example, Su and Rungruang (2023) successfully combined bibliometric mapping with qualitative thematic analysis to examine research on workplace conflict outcomes, yielding richer interdisciplinary insights. Tardin et al. (2024) similarly blended bibliometric data with qualitative evaluations to explore sustainability orientations in organizational research, while Dana et al. (2023) applied a dual-method approach to map research on women's entrepreneurship in developing economies. These examples underscore the potential of integrating qualitative dimensions to contextualize and critically assess bibliometric findings—an approach notably absent in Fatawi et al.'s study.

In addition to expanding data sources and methods, the selection of bibliometric tools also warrants careful consideration. While VOSviewer remains popular for its intuitive interface and visually engaging network maps (An et al., 2024; Jiang et al., 2024), it lacks the temporal analytical features of CiteSpace, which excels in identifying citation bursts and tracking topic evolution over time (Wu et al., 2023; Que et al., 2025; Wei et al., 2025). Bibliometrix, an R-based package, offers advanced statistical and bibliographic capabilities, including performance metrics and co-word analyses, which are especially useful for large, dynamic fields like AI and NLP (Feng et al., 2023; Peng et al., 2025). A more nuanced methodological design in bibliometric research—one that combines multiple tools and data sources—would yield a deeper, more representative understanding of prompt engineering's landscape.

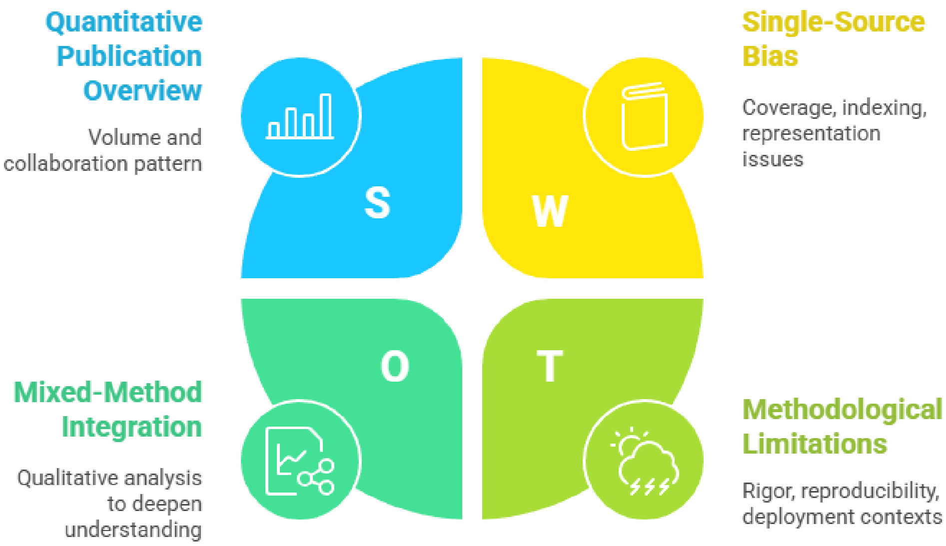

Figure 2. Methodological strengths and limitations of a scopus-based bibliometric study

The methodological limitations of Fatawi et al.'s (2024) approach constrain the interpretability and scope of their conclusions, even though they provide a foundational overview of research activity in prompt engineering. Future reviews should adopt a more diversified, mixed-methods strategy—incorporating multi-source data aggregation, complementary analytical tools, and qualitative inquiry (as recommended in Figure 2), to produce more accurate and actionable insights into the development, diffusion, and ethical implications of prompt-based optimization in LLMs.

ETHICAL IMPLICATIONS AND POLICY CONSIDERATIONS

As prompt engineering becomes increasingly integral to Large Language Model (LLM) applications, it brings to the forefront a complex landscape of ethical and governance challenges. Central among these is the threat of prompt injection attacks, where adversaries embed malicious instructions within user inputs to subvert intended model behavior—potentially generating toxic, biased, or disallowed content. Recent research emphasizes the importance of adopting black-box defense mechanisms capable of identifying and filtering adversarial prompts without relying on internal model access (Takemoto, 2024). Complementing this, adversarial training strategies have been shown to enhance model robustness against manipulation by exposing models to varied attack vectors during training (Yang et al., 2024). Furthermore, game-theoretic modeling approaches are gaining traction as dynamic tools for anticipating and mitigating evolving attack strategies (Parras et al., 2022).

We propose the design of a "prompt-safe" audit framework that comprises three core components to operationalize these defenses, namely, (1) risk assessment protocols to evaluate prompts for vulnerability to injection; (2) automated red-teaming systems that simulate adversarial prompt behavior during testing phases; and (3) deployment-level safeguards, including blacklists and real-time input filtering mechanisms. Such a framework would serve a similar role to security audits in software engineering, ensuring that prompts meet safety, transparency, and integrity standards before public deployment.

Beyond security, ethical risks also stem from the high sensitivity of LLM outputs to prompt phrasing, which can result in unintentional misinformation dissemination. Research indicates that prompt design significantly shapes user interpretation and acceptance of content. For example, Samayoa and Albarracín (2025) demonstrate that bypass-style prompts can trigger corrective belief formation, while Martel et al. (2024) find that accuracy-framed prompts reduce susceptibility to partisan misinformation. Similarly, Wang and Jacobson (2022) highlight that prompt corrections may reduce misbeliefs, although not eliminate them entirely, and Roozenbeek et al. (2022) report the effectiveness of “accuracy nudges” in encouraging factual content sharing on social platforms. These studies underscore that prompts are not neutral artifacts—they are rhetorical devices capable of shaping beliefs and behaviors. Therefore, automated fact-verification tools, such as those used in systems like TruthfulQA, should be integrated into prompt deployment pipelines, particularly in high-stakes domains such as healthcare, law, and public policy.

Data privacy also emerges as a pressing concern. Prompt engineering workflows frequently involve interactive sessions that collect or process user-generated content, which may inadvertently include personal or sensitive information. As such, clear data governance protocols—including anonymization procedures, informed consent policies, and traceable audit trails—must be instituted. These measures are vital not only for regulatory compliance but also for fostering user trust in AI-driven systems.

To frame these considerations within a broader ethical context, guidance can be drawn from established AI ethics frameworks. The European Union’s Ethics Guidelines for Trustworthy AI (as discussed in Mwogosi, 2025) articulate core principles—such as transparency, accountability, and human agency—that are especially relevant for prompt engineering. Recent work by Park (2025) and Fan (2024) emphasizes the value of human-centric approaches to AI governance, ensuring that prompt strategies align with social values rather than merely optimizing computational output. Moreover, scholars such as Saraiva (2024) and Odero et al. (2024) argue for culturally inclusive frameworks that reflect diverse moral perspectives, recognizing that ethical standards should be sensitive to geographic and philosophical plurality.

Lastly, ensuring inclusive evaluation across languages, dialects, and cultural contexts is essential to avoid reinforcing systemic biases. Without multilingual robustness checks, prompt strategies risk amplifying inequality in access to reliable AI outputs—particularly for communities in the Global South or speakers of low-resource languages. Ensuring linguistic fairness should thus be a design imperative, not an afterthought.

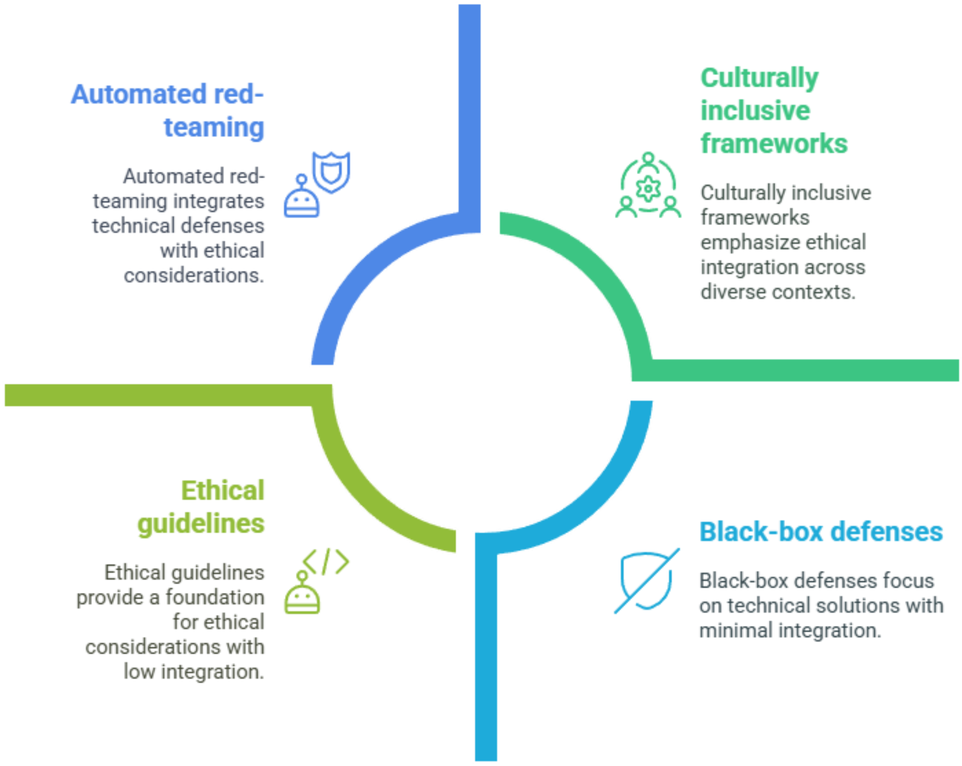

Figure 3. Ethical implications and policy considerations in prompt engineering

The ethical and governance structures that underpin prompt engineering must evolve as the practice of prompt engineering evolves (Figure 3). Integrating technical safeguards, rhetorical awareness, and normative frameworks is not only advisable—it is imperative to ensure that prompt-based systems remain secure, fair, and aligned with human values.

RECOMMENDATIONS FOR FUTURE RESEARCH

To advance the field of prompt engineering beyond its current scope, future research must first address the data source limitations identified in existing bibliometric analyses. As highlighted in this commentary, Fatawi et al. (2024) rely solely on Scopus-indexed publications, which may lead to underrepresentation of emerging or non-English scholarship. Scholars should expand their literature mapping efforts by incorporating additional databases such as Web of Science, arXiv, and region-specific repositories. Doing so would enhance the representativeness and diversity of bibliometric insights and surface trends that are currently overlooked due to linguistic or indexing biases.

Equally critical is the establishment of robust prompt auditing standards. Prompt engineering should adopt practices analogous to software security reviews by systematically assessing prompts for risks related to injection attacks, embedded biases, privacy breaches, and domain misalignment. As discussed in recent work on prompt injection defenses, such audits could draw from black-box testing, adversarial training, and game-theoretic modeling frameworks (Yang et al., 2024; Parras et al., 2022). The creation of standardized "prompt-safe" checklists or validation protocols would improve the transparency and trustworthiness of prompt-based systems before public deployment.

A key gap identified in both the bibliometric review and broader literature is the absence of longitudinal, real-world validation of prompt strategies. Existing studies have demonstrated the importance of contextually adaptive prompts in enhancing user engagement across diverse applications, including dietary adherence and mental health assessments (Lee et al., 2024; Wang et al., 2025). Metrics such as the Overall Prompting Effectiveness (OPE) framework (Wilbers et al., 2023) and iterative refinement techniques (Velásquez et al., 2023) highlight how prompt quality must evolve with system use. Accordingly, long-term A/B testing in operational environments—e.g., conversational agents, clinical decision tools, or education platforms—should be prioritized to assess prompt consistency, user satisfaction, and downstream behavioral impacts (Patil et al., 2024; Stephan et al., 2024).

In parallel, future research should prioritize mixed-methods approaches that integrate bibliometric mapping with ethnographic, behavioral, or interview-based studies. While bibliometrics map conceptual landscapes, qualitative methods offer insight into how practitioners engage with prompts, what challenges they face, and how social and institutional factors shape prompt outcomes. For instance, Martovytskyi et al. (2022) and Johnson et al. (2023) illustrate how ethnographic methods uncover latent usability and trust dynamics that purely quantitative methods may overlook. Peponakis et al. (2023) further demonstrate how computational-ethnographic hybrid models can yield actionable insights in applied contexts. Embedding such user-centered methods within prompt research would help contextualize performance metrics within lived interactional realities.

Lastly, while some speculative proposals in current literature lack empirical grounding, the emerging synergy between NLP and quantum computing merits early-stage exploration. Research in Quantum Natural Language Processing (QNLP) has shown promising developments in tasks such as sentiment analysis and semantic representation (Ruskanda et al., 2023; Yan et al., 2022). Studies by Zhou et al. (2022) and Lorenz et al. (2023) demonstrate that Noisy Intermediate-Scale Quantum (NISQ) systems can already perform meaningful NLP operations, while Bausch et al. (2021) and Yu et al. (2024) propose hybrid architectures that combine classical and quantum resources for increased modeling capacity. Although still nascent, these developments justify a measured but forward-looking research agenda into quantum-enhanced prompt optimization.

A rigorous and forward-looking research agenda for prompt engineering should prioritize the following directions: (1) broaden bibliometric data sources beyond Scopus to incorporate multilingual publications and preprint repositories, thereby capturing a more representative global research landscape; (2) develop standardized prompt auditing frameworks to evaluate bias, safety vulnerabilities, and transparency before deployment; (3) implement longitudinal validation strategies—such as Overall Prompting Effectiveness (OPE) assessments and A/B testing—in real-world settings to examine performance over time; (4) adopt mixed-methods approaches that integrate qualitative insights into user experience and contextual deployment; and (5) investigate the potential of emerging technologies, such as quantum computing, to support the next generation of prompt design and optimization. By advancing along these trajectories, the field can transition from descriptive bibliometric mapping toward empirically grounded, context-sensitive, and ethically rigorous methodologies that support the sustainable and responsible evolution of prompt-based AI systems.

CONCLUSION

Prompt engineering marks a fundamental shift in how we operationalize the capabilities of Large Language Models (LLMs)—transforming them from static, resource-intensive architectures into flexible, adaptable systems guided by nuanced user inputs. While the bibliometric review conducted by Fatawi et al. (2024) offers a valuable macroscopic overview of publication patterns, thematic trends, and collaborative networks, it stops short of addressing the methodological, ethical, and practical complexities underpinning this rapidly evolving field. This commentary extends their contribution by critically interrogating the epistemological blind spots of single-source bibliometric analysis, underscoring the importance of mixed-methods research, and proposing concrete strategies for developing secure, equitable, and context-sensitive prompt engineering practices.

Specifically, we have argued for broadening bibliometric inputs beyond Scopus to capture underrepresented perspectives, for integrating qualitative methodologies to better understand the lived realities of prompt deployment, and for instituting prompt audit standards modeled after rigorous software assurance practices. These recommendations are not merely procedural but foundational to ensuring that prompt engineering matures into a methodologically sound and socially responsible discipline.

Looking forward, the sustainable impact of prompt engineering will depend on the field's commitment to balancing innovation with critical reflection. Longitudinal validation studies, human-centered ethnographic research, and empirically grounded exploration of emerging technologies—such as quantum-enhanced NLP—will be essential to shaping a future in which prompt-based systems are not only performant but also accountable, inclusive, and adaptable. In doing so, the community can build on the structural map provided by bibliometric reviews and forge a richer, more reflexive understanding of prompt engineering's transformative potential in AI development.

LIMITATION

Fatawi et al.'s reliance on a single English-language database (Scopus) and exclusion of preprints or non-English sources risks underrepresenting emerging and regional work, while their purely bibliometric approach overlooks the practical "how" and "why" of prompt design. This focus obscures reproducibility challenges, variability in prompt efficacy, and domain-specific fragility, because it ignores qualitative insights into user practices, evaluation standards, and real-world deployment contexts.

RECOMMENDATION

Future research should draw on multiple data sources (for example, Web of Science, arXiv, and regional repositories) and employ mixed-method frameworks that combine bibliometric mapping with case studies, interviews, or ethnographic analysis. Scholars ought to develop standardized prompt-audit protocols—covering injection risks, bias checks, and privacy safeguards—and run longitudinal A/B tests in operational environments to track prompt performance over time. Finally, exploring emerging frontiers such as quantum-enhanced NLP could open new paths for optimizing prompts in resource-constrained and complex applications.

Author Contributions

The authors have sufficiently contributed to the study, and have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Acknowledgement

We thank the authors of the original bibliometric review, Fatawi and colleagues, for making their data and visualizations available, and we are grateful to the peer reviewers whose thoughtful feedback greatly improved this commentary.

Declaration of Interest
The authors declare no conflict of interest.

REFERENCES

Adeosun, S. (2024). Scopus indexing delays of articles published in major pharmacy practice journals. *Innovations in Pharmacy, 15*(4), 5. https://doi.org/10.24926/iip.v15i4.6322

An, X., He, J., Bi, B., Wu, G., Xu, J., Yu, W., … Ren, Z. (2024). The role of astrocytes in Alzheimer's disease: A bibliometric analysis. *Frontiers in Aging Neuroscience, 16*. https://doi.org/10.3389/fnagi.2024.1481748

Banda, F., Marivate, V., & Nakatumba-Nabende, J. (2025). A few-shot learning approach for a multilingual agro-information question answering system. *Applied AI Letters, 6*(2). https://doi.org/10.1002/ail2.122

Bausch, J., Subramanian, S., & Piddock, S. (2021). A quantum search decoder for natural language processing. *Quantum Machine Intelligence, 3*(1). https://doi.org/10.1007/s42484-021-00041-1

Botunac, I., Bakarić, M., & Matetić, M. (2024). Comparing fine-tuning and prompt engineering for multi-class classification in hospitality review analysis. *Applied Sciences, 14*(14), 6254. https://doi.org/10.3390/app14146254

Chen, D., Huang, R., Jomy, J., Wong, P., Yan, M., Croke, J., … Raman, S. (2024). Performance of multimodal artificial intelligence chatbots evaluated on clinical oncology cases. *JAMA Network Open, 7*(10), e2437711. https://doi.org/10.1001/jamanetworkopen.2024.37711

Dana, L., Chhabra, M., & Agarwal, M. (2023). A two-decade history of women's entrepreneurship research trajectories in developing economies context: Perspectives from India. *Journal of Management History, 30*(1), 6–28. https://doi.org/10.1108/jmh-11-2022-0064

Demirbaş, K., Saygılı, S., Yılmaz, E., Gülmez, R., Ağbaş, A., Taşdemir, M., … Canpolat, N. (2025). The potential of ChatGPT as a source of information for kidney transplant recipients and their caregivers. *Pediatric Transplantation, 29*(3). https://doi.org/10.1111/petr.70068

Dimitsaki, S., Natsiavas, P., & Jaulent, M. (2024). Applying AI to structured real-world data for pharmacovigilance purposes: Scoping review. *Journal of Medical Internet Research, 26*, e57824. https://doi.org/10.2196/57824

Fan, L. (2024). Artificial intelligence ethics: A dialogue between technological advances and human values. *International Journal of Education and Humanities, 14*(2), 260–265. https://doi.org/10.54097/tvqkkf40

Fatawi, I., Asy'ari, M., Hunaepi, H., Samsuri, T., & Bilad, M. R. (2024). Empowering language models through advanced prompt engineering: A comprehensive bibliometric review. *Indonesian Journal of Science & Technology, 9*(2), 441–462. https://doi.org/10.17509/ijost.v9i2.71481

Feng, H., Chen, J., Zhang, Z., Lou, Y., Zhang, S., & Yang, W. (2023). A bibliometric analysis of artificial intelligence applications in macular edema: Exploring research hotspots and frontiers. *Frontiers in Cell and Developmental Biology, 11*. https://doi.org/10.3389/fcell.2023.1174936

Gimeno, A., Krause, K., D'Souza, S., & Walsh, C. (2024). Completeness and readability of GPT-4-generated multilingual discharge instructions in the pediatric emergency department. *JAMIA Open, 7*(3). https://doi.org/10.1093/jamiaopen/ooae050

Gurung, B., Liu, P., Harris, P., Sagi, A., Field, R., Sochart, D., … Asopa, V. (2022). Artificial intelligence for image analysis in total hip and total knee arthroplasty. *The Bone & Joint Journal, 104-B*(8), 929–937. https://doi.org/10.1302/0301-620x.104b8.bjj-2022-0120.r2

Hauna, A., Yunus, A., Fukui, M., & Khomsah, S. (2025). Enhancing LLM efficiency: A literature review of emerging prompt optimization strategies. *International Journal on Robotics Automation and Sciences, 7*(1), 72–83. https://doi.org/10.33093/ijoras.2025.7.1.9

Heston, T., & Khun, C. (2023). Prompt engineering in medical education. *International Medical Education, 2*(3), 198–205. https://doi.org/10.3390/ime2030019

Huang, J., Yang, D., Rong, R., Nezafati, K., Treager, C., Chi, Z., … Xie, Y. (2024). A critical assessment of using ChatGPT for extracting structured data from clinical notes. *NPJ Digital Medicine, 7*(1). https://doi.org/10.1038/s41746-024-01079-8

Jiang, Y., Cai, Y., Zhang, X., Chen, L., Zhou, X., & Chen, Y. (2024). A two-decade bibliometric analysis of laser in ophthalmology: From past to present. *Clinical Ophthalmology, 18*, 1313–1328. https://doi.org/10.2147/opth.s458840

Johnson, A., Wang, L., Boucher, S., & Fulgoni, A. (2023). Quality by design pilot analysis of FDA regulatory guidance web data usability for innovators in non-Hodgkin lymphoma: Overcoming challenges to CAR-T development. *Blood, 142*(Suppl. 1), 7169. https://doi.org/10.1182/blood-2023-187280

Johnson-Eilola, J., Selber, S., & York, E. (2024). Can artificial intelligence robots write effective instructions? *Journal of Business and Technical Communication, 38*(3), 199–212. https://doi.org/10.1177/10506519241239641

Kim, K., Lee, M., Won, H., Kim, M., Kim, Y., & Lee, S. (2023). Multi-stage prompt tuning for political perspective detection in low-resource settings. *Applied Sciences, 13*(10), 6252. https://doi.org/10.3390/app13106252

Kim, T., Makutonin, M., Sirous, R., & Javan, R. (2025). Optimizing large language models in radiology and mitigating pitfalls: Prompt engineering and fine-tuning. *Radiographics, 45*(4). https://doi.org/10.1148/rg.240073

Kumar, R., Saxena, S., Kumar, V., Prabha, V., Kumar, R., & Kukreti, A. (2023). Service innovation research: A bibliometric analysis using VOSviewer. *Competitiveness Review: An International Business Journal Incorporating Journal of Global Competitiveness, 34*(4), 736–760. https://doi.org/10.1108/cr-01-2023-0010

Lai, Q., Spoletini, G., Mennini, G., Laureiro, Z., Tsilimigras, D., Pawlik, T., … Rossi, M. (2020). Prognostic role of artificial intelligence among patients with hepatocellular cancer: A systematic review. *World Journal of Gastroenterology, 26*(42), 6679–6689. https://doi.org/10.3748/wjg.v26.i42.6679

Lee, L., Hall, R., Stanley, J., & Krebs, J. (2024). Tailored prompting to improve adherence to image-based dietary assessment: Mixed methods study. *JMIR mHealth and uHealth, 12*, e52074. https://doi.org/10.2196/52074

Leung, C. (2024). Promoting optimal learning with ChatGPT: A comprehensive exploration of prompt engineering in education. *Asian Journal of Contemporary Education, 8*(2), 104–114. https://doi.org/10.55493/5052.v8i2.5101

Li, W., Chen, X., Deng, X., Wen, H., You, M., Liu, W., … Li, J. (2024). Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. *NPJ Digital Medicine, 7*(1). https://doi.org/10.1038/s41746-024-01029-4

Liu, Y., Du, H., Niyato, D., Kang, J., Cui, S., Shen, X., … Zhang, P. (2024). Optimizing mobile-edge AI-generated everything (AIGX) services by prompt engineering: Fundamental, framework, and case study. *IEEE Network, 38*(5), 220–228. https://doi.org/10.1109/mnet.2023.3335255

Lorenz, R., Pearson, A., Meichanetzidis, K., Kartsaklis, D., & Coecke, B. (2023). QNLP in practice: Running compositional models of meaning on a quantum computer. *Journal of Artificial Intelligence Research, 76*, 1305–1342. https://doi.org/10.1613/jair.1.14329

Maharjan, J., Garikipati, A., Singh, N., Cyrus, L., Sharma, M., Ciobanu, M., … Das, R. (2024). OpenMedLM: Prompt engineering can outperform fine-tuning in medical question-answering with open-source large language models. *Scientific Reports, 14*(1). https://doi.org/10.1038/s41598-024-64827-6

Martel, C., Rathje, S., Clark, C., Pennycook, G., Bavel, J., Rand, D., … Linden, S. (2024). On the efficacy of accuracy prompts across partisan lines: An adversarial collaboration. *Psychological Science, 35*(4), 435–450. https://doi.org/10.1177/09567976241232905

Martovytskyi, V., Сєвєрінов, O., Liashenko, O., Koltun, Y., Liashenko, S., Kis, V., … Yevstrat, D. (2022). Devising an approach to the identification of system users by their behavior using machine learning methods. *Eastern-European Journal of Enterprise Technologies, 3*(3(117)), 23–34. https://doi.org/10.15587/1729-4061.2022.259099

Mwogosi, A. (2025). Ethical and privacy challenges of integrating generative AI into EHR systems in Tanzania: A scoping review with a policy perspective. *Digital Health, 11*. https://doi.org/10.1177/20552076251344385

Nyaaba, M., & Zhai, X. (2024). Generative AI professional development needs for teacher educators. *Journal of AI, 8*(1), 1–13. https://doi.org/10.61969/jai.1385915

Odero, B., Nderitu, D., & Samuel, G. (2024). The ubuntu way: Ensuring ethical AI integration in health research. *Wellcome Open Research, 9*, 625. https://doi.org/10.12688/wellcomeopenres.23021.1

Park, C. (2025). Ethical artificial intelligence in nursing workforce management and policymaking: Bridging philosophy and practice. *Journal of Nursing Management*. Advance online publication. https://doi.org/10.1155/jonm/7954013

Parras, J., Almodóvar, A., Apellániz, P., & Zazo, S. (2022). Inverse reinforcement learning: A new framework to mitigate an intelligent backoff attack. *IEEE Internet of Things Journal, 9*(24), 24790–24799. https://doi.org/10.1109/jiot.2022.3194694

Patil, R., Heston, T., & Bhuse, V. (2024). Prompt engineering in healthcare. *Electronics, 13*(15), 2961. https://doi.org/10.3390/electronics13152961

Peng, X., Bian, H., Zhao, H., Jia, D., Li, M., Li, W., … Xu, P. (2025). Research hotspots and trends in lung cancer STAS: A bibliometric and visualization analysis. *Frontiers in Oncology, 14*. https://doi.org/10.3389/fonc.2024.1495911

Peponakis, M., Kapidakis, S., Doerr, M., & Tountasaki, E. (2023). From calculations to reasoning: History, trends and the potential of computational ethnography and computational social anthropology. *Social Science Computer Review, 42*(1), 84–102. https://doi.org/10.1177/08944393231167692

Que, X., Zhu, S., & Han, B. (2025). Data-driven visualization of the dynamics of geothermal energy and hot dry rock research. *Energies, 18*(9), 2342. https://doi.org/10.3390/en18092342

Roozenbeek, J., Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances, 8*(34), eabo6254. https://doi.org/10.1126/sciadv.abo6254

Ruskanda, F., Abiwardani, M., Syafalni, I., Larasati, H., & Mulyawan, R. (2023). Simple sentiment analysis ansatz for sentiment classification in quantum natural language processing. *IEEE Access, 11*, 120612–120627. https://doi.org/10.1109/access.2023.3327873

Sacca, L., Lobaina, D., Burgoa, S., Lotharius, K., Moothedan, E., Gilmore, N., … Kitsantas, P. (2024). Promoting artificial intelligence for global breast cancer risk prediction and screening in adult women: A scoping review. *Journal of Clinical Medicine, 13*(9), 2525. https://doi.org/10.3390/jcm13092525

Saeed, N., Ridzuan, M., Majzoub, R., & Yaqub, M. (2023). Prompt-based tuning of transformer models for multi-center medical image segmentation of head and neck cancer. *Bioengineering, 10*(7), 879. https://doi.org/10.3390/bioengineering10070879

Samayoa, J., & Albarracín, D. (2025). Bypassing versus correcting misinformation: Efficacy and fundamental processes. *Journal of Experimental Psychology: General, 154*(1), 18–38. https://doi.org/10.1037/xge0001687

Saraiva, L. (2024). Moral agency and responsibility in AI systems. *International Journal of Philosophy, 3*(1), 25–36. https://doi.org/10.47941/ijp.1867

Skryd, A., & Lawrence, K. (2024). ChatGPT as a tool for medical education and clinical decision-making on the wards: Case study. *JMIR Formative Research, 8*, e51346. https://doi.org/10.2196/51346

Stephan, D., Bertsch, A., Burwinkel, M., Vinayahalingam, S., Al-Nawas, B., Kämmerer, P., … Thiem, D. (2024). AI in dental radiology—Improving the efficiency of reporting with ChatGPT: Comparative study. *Journal of Medical Internet Research, 26*, e60684. https://doi.org/10.2196/60684

Su, M., & Rungruang, P. (2023). Mapping the knowledge base and theoretical evolution of workplace conflict outcomes: A bibliometric and qualitative review, 1972–2022. *International Journal of Conflict Management, 35*(2), 360–386. https://doi.org/10.1108/ijcma-02-2023-0025

Sutar, P., Kolte, G., Yamini, S., & Mathiyazhagan, K. (2024). Food supply chain resilience in the digital era: A bibliometric analysis and development of conceptual framework. *Journal of Business and Industrial Marketing, 39*(9), 1863–1893. https://doi.org/10.1108/jbim-10-2023-0587

Takemoto, K. (2024). All in how you ask for it: Simple black-box method for jailbreak attacks. *Applied Sciences, 14*(9), 3558. https://doi.org/10.3390/app14093558

Tan, W., Liao, J., Wang, Y., & Li, W. (2023). A global bibliometric analysis on Kawasaki disease research over the last 5 years (2017–2021). *Frontiers in Public Health, 10*. https://doi.org/10.3389/fpubh.2022.1075659

Tardin, M., Perin, M., Simões, C., & Braga, L. (2024). Organizational sustainability orientation: A review. *Organization & Environment, 37*(2), 298–324. https://doi.org/10.1177/10860266231226144

Velásquez, J., Franco, C., & Cadavid, L. (2023). Prompt engineering: A methodology for optimizing interactions with AI-language models in the field of engineering. *Dyna, 90*(230), 9–17. https://doi.org/10.15446/dyna.v90n230.111700

Vemprala, S., Bonatti, R., Bucker, A., & Kapoor, A. (2024). ChatGPT for robotics: Design principles and model abilities. *IEEE Access, 12*, 55682–55696. https://doi.org/10.1109/access.2024.3387941

Wang, L., Bhanushali, T., Huang, Z., Yang, H., Badami, S., & Hightow-Weidman, L. (2025). Evaluating generative AI in mental health: Systematic review of capabilities and limitations. *JMIR Mental Health, 12*, e70014. https://doi.org/10.2196/70014

Wang, L., Bi, W., Zhao, S., Ma, Y., Lv, L., Meng, C., … Lv, H. (2024). Investigating the impact of prompt engineering on the performance of large language models for standardizing obstetric diagnosis text: Comparative study. *JMIR Formative Research, 8*, e53216. https://doi.org/10.2196/53216

Wang, W., & Jacobson, S. (2022). Effects of health misinformation on misbeliefs: Understanding the moderating roles of different types of knowledge. *Journal of*

Information, Communication and Ethics in Society, 21(1), 76–93. https://doi.org/10.1108/jices-02-2022-0015

Wei, L., Griffiths, L., & Tollefsbol, T. (2025). CeRNA networks in ischemic stroke: A bibliometric analysis. *Egyptian Journal of Medical Human Genetics, 26*(1). https://doi.org/10.1186/s43042-025-00687-7

Wilbers, S., Espinosa-Leal, L., Sand, R., & Reiff-Stephan, J. (2023). Overall prompting effectiveness for optimizing human-machine interaction in cyber-physical systems. *Journal of Integrated Design and Process Science, 27*(3–4), 211–220. https://doi.org/10.1177/10920617241295859

Wilder, E., & Walters, W. (2021). Using conventional bibliographic databases for social science research: Web of Science and Scopus are not the only options. *Scholarly Assessment Reports, 3*(1), 4. https://doi.org/10.29024/sar.36

Wu, H. (2024). Large language models capsule: A research analysis of in-context learning (ICL) and parameter-efficient fine-tuning (PEFT) methods. *Applied and Computational Engineering, 43*(1), 327–331. https://doi.org/10.54254/2755-2721/43/20230858

Wu, L., Jin, L., Li, L., Yu, K., Wu, J., Lei, Y., … He, J. (2023). An examination of Alzheimer's disease and white matter from 1981 to 2023: A bibliometric and visual analysis. *Frontiers in Neurology, 14*. https://doi.org/10.3389/fneur.2023.1268566

Yan, Y., Qiu, D., & Yan, R. (2022). A quantum language-inspired tree structural text representation for semantic analysis. *Mathematics, 10*(6), 914. https://doi.org/10.3390/math10060914

Yang, J., Chi, Q., Wen-qiang, X., & Yu, H. (2024). Research on adversarial attack and defense of large language models. *Applied and Computational Engineering, 93*(1), 105–113. https://doi.org/10.54254/2755-2721/93/20240922

Yu, W., Yin, L., Zhang, C., Chen, Y., & Liu, A. (2024). Application of quantum recurrent neural network in low-resource language text classification. *IEEE Transactions on Quantum Engineering, 5*, 1–13. https://doi.org/10.1109/tqe.2024.3373903

Zhou, X., Pang, J., Feng, Y., Liu, F., Guo, J., Liu, W., … Shan, Z. (2022). A new method of software vulnerability detection based on a quantum neural network. *Scientific Reports, 12*(1). https://doi.org/10.1038/s41598-022-11227-3