

Article

Advanced Interdisciplinary Science and Technology
2025, Vol. 1(2), 73-76. DOI: 10.54117/aist.2025.v2i1.121



Prompts for Writing Scientific Research Papers Using Large Language

Mingzheng Ma*

Hebei Normal University, China

**Corresponding Author: 2660925202@qq.com*

Received: 18 December 2025 | Revised: 22 December 2025 | Accepted: 23 December 2025 | Published online: 26 December 2025

Abstract: The rapid advance of artificial intelligence has endowed Large Language Models (LLMs) with formidable capabilities in natural-language generation, textual comprehension and knowledge integration. Consequently, these systems are increasingly deployed to support scholarly paper writing. Prompts—the principal interface between user and model—exert a decisive influence on the quality and usability of the generated text. This study systematically examines design principles, strategies and optimization techniques for prompts employed in LLM-assisted academic writing, and evaluates their effectiveness across discrete research phases through illustrative cases. Findings indicate that prompts characterized by clear structure, sufficient context and unambiguous task specification significantly enhance the logical coherence and disciplinary professionalism of model output. Nevertheless, LLMs remain prone to knowledge hallucination and citation distortion; human verification and ethical safeguards are therefore indispensable. The paper proposes an operational framework for prompt engineering that can guide researchers and foster the responsible integration of AI into academic writing.

Keywords: Large Language Models; Research Articles; Prompts; Artificial Intelligence; Academic Writing

Academic writing has long been regarded as the “second laboratory” of science, serving not only as a repository for knowledge but as a reflexive process that exposes flaws in research design, data reliability and the limits of interpretation. Conventional writing is typically high-input and low-iteration: every loop—from topic selection to final manuscript—requires iterative literature retrieval, data reconciliation and argument adjustment, so that a single oversight can trigger cascading revisions. Doctoral candidates and non-native authors face a triple threshold of language barriers, logical conventions and disciplinary discourse, amplifying writing anxiety. Recently, transformer-based large language models (LLMs) with hundreds of billions of parameters have approached—or surpassed—human baselines in text generation, question answering and code synthesis, making scholarly writing a high-value application. Yet the stochastic and hallucinatory nature of LLMs renders fully autonomous drafting risky; instead, carefully engineered prompts act as “invisible rails” that suppress digressions at the macro level while activating domain knowledge and critical templates at the micro level, balancing efficiency with quality. Systematic inquiry into the mechanisms, risk boundaries and ethical governance of prompt engineering in academic writing can therefore offer researchers a reproducible playbook and provide policymakers with empirical evidence for regulating AI-assisted scholarship.

Over the past five years LLMs have evolved from “lexical probability calculators” to “open-domain reasoning engines.” GPT-3 first displayed in-context and few-shot generalization; GPT-4 introduced multimodal chains-of-thought and symbolic reflection; Claude adopted Constitutional AI for value alignment; domestic models such as ERNIE Bot and Tongyi Qianwen underwent secondary pre-training on Chinese scholarly corpora, cutting terminology errors. Parameter scaling from millions to hundreds of billions has triggered emergent abilities: beyond grammar correction, models can now generate testable hypotheses, auto-complete flow diagrams and draft IMRaD-compliant paragraphs. Meanwhile, scholarly writing itself is growing more complex: interdisciplinary convergence inflates reference lists, open-data mandates demand reproducible methods, and preregistration requires transparent hypotheses. The traditional “lone-wolf” model can no longer satisfy the “fast research–fast publication–fast iteration” ecology, making AI-assisted writing a necessity and turning prompt engineering into both an experimental field and a precision discipline. A prompt is essentially a semantic interface; its quality governs the depth of the model’s task comprehension and stylistic control. Academic writing’s stringent demands for accuracy, logic and verifiability require stricter control variables than everyday tasks. First, the principle of explicitness demands single-verb instructions (“Generate...”) rather than nested colloquial clauses. Second, contextual adequacy entails embedding disciplinary norms—e.g., “Use APA 7th” or “Follow CONSORT”—to curb format hallucinations. Third, structuralisation encourages numbered or hierarchical slots to elicit isomorphic output; experiments show that symbolic sequences such as “1.2.3.” raise list-generation probability by 47 %. Fourth, iterability reserves conversational windows for stepwise refinement—e.g., methods outline → sample-size calculation → ethics code—forming a “chain calibration” that reduces information loss. Finally, domain adaptation infuses prompts with field-specific lexicons and reasoning templates (e.g., “mechanism–context–policy” triads in

sociology or “composition–structure–property” chains in materials science) to activate specialist sub-networks and enhance credibility.

At the topic-selection stage prompts must balance “trend tracking” with “gap spotting.” One can instruct the model to map a three-year citation network, detect burst keywords and then filter candidates by data availability, ethical risk and budget, converting “inspiration” into “testable questions.” For literature reviews, prompts should trigger three-tier reasoning: chronological mapping, methodological clustering and critical appraisal of instruments or theories, preventing mere laundry lists. The methods section demands the highest reproducibility; prompts must explicitly request sample-size formulas, randomization schemes and missingness strategies, plus executable R or Python snippets, synchronizing “method–code” generation. In results and discussion, where data-dumping or over-interpretation looms, prompts can embed counterfactual templates: report main effects → robustness checks → “had the opposite been observed” discussions, cultivating reflexivity. Even abstracts and titles—short yet visibility-critical—benefit from a three-step prompt: information compression → selling-point extraction → SEO optimization, balancing scholarly norms with discoverability.

To validate prompt efficacy, we recruited 30 Ph.D. candidates and randomly assigned them to a vague-prompt group or a structured-prompt group to draft a methods paragraph on “short-video use and adolescent sleep quality.” The vague group received only “Write a methods section”; the structured group received a hierarchical prompt specifying participants, manipulations, data collection and analysis within 400 words. The structured cohort scored 42 % higher on information completeness and 35 % higher on terminological accuracy, while omitting frequent lapses such as unreported sample sizes. Eye-tracking further revealed longer fixations on numerals and bullets, suggesting that “visual scaffolds” activate format memory. Crucially, reproducibility ratings by independent reviewers were nearly one standard deviation higher, indicating that prompts influence not just surface quality but the verifiability core of academic prose.

LLMs have graduated from “efficiency tools” to “cognitive partners.” Efficiency: a thousand-word draft in seconds compresses weeks of cold-start labor. Knowledge: cross-disciplinary associations surpass individual memory, supplying multi-perspective evidence. Language: real-time grammar correction and terminology recommendation lowers publication barriers for non-native scholars. Yet limitations are systemic. Knowledge hallucinations are not random noise but over-confident extrapolations from high-frequency co-occurrences, especially in citations and historical facts. Citation distortion arises from noisy training corpora and mirror-site pollution, yielding plausible but non-existent references. Lack of critical thinking surfaces when models merely echo prompt cues instead of interrogating assumptions or statistical flaws. Ethical risks are tangible: journals have already detected undisclosed AI-generated passages that evade traditional similarity checks. Without governance, AI could shift from “accelerator” to “black hole” of integrity.

The core of research ethics is traceability and accountability. AI hybrid authorship demands layered disclosure: low-risk language polishing can be footnoted; high-risk methodological or interpretative contributions require detailed statements of model version, prompt strategy and human verification. Journals and funders should co-develop AI-use declaration templates that archive

prompts, generated text and revision logs, enabling end-to-end auditing. The community must promote “verifiable writing” standards in which data, code and prompts form a triad subject to replication and reverse audit. Universities should embed AI ethics in research-integrity training and institute pre-screening workflows, coupling technical and institutional safeguards to defend public trust. Next-generation LLMs will target factuality and explainability. By integrating knowledge graphs and live citation databases, models can insert DOIs or PMIDs during generation, driving hallucination rates below journal tolerance. Prompt engineering will evolve from hand-crafting to auto-optimization: reinforcement-learning-based “Prompt Bots” will dynamically generate discipline- or journal-specific prompts, delivering personalized writing assistants. Multi-agent orchestration—one agent retrieving literature, another cleaning data, a third drafting results and a fourth checking compliance—will let authors act as directors via natural-language commands. In the long run, prompt engineering may become the third universal methodological skill after statistics and coding, propelling scholarly writing from an experience-driven craft to an algorithm-driven paradigm.