

What Does It Take to Research with AI? A Rapid Review of Competencies to Train LLM-Literate Researchers

Danilo Monteiro Ribeiro¹, Ronnie de Souza Santos^{1,4}, Rodrigo Siqueira¹,
Breno Andrade¹, Rafael Batista Duarte^{1,2}, Gilberto Hida¹,
Julia Alencar¹, Gustavo Pinto³

¹CESAR School – Recife, PE – Brazil

²Universidade de Pernambuco (UPE) – Recife, PE – Brazil

³Federal University of Pará (UFPA) – Belém, PA – Brazil

⁴University of Calgary – Calgary, AB – Canada

{dmr,ress,rms9,baa3,gsh,rbd}@cesar.school,

juqueirozalencar@gmail.com, gpinto@ufpa.br

Abstract. *The growing adoption of Large Language Models in scientific research has created a need to understand what competencies researchers and graduate students require to use these tools critically and responsibly. This rapid review analyzed 194 articles retrieved from Elicit and Google Scholar (2022–2025), from which 40 were selected for competency extraction and thematic analysis following independent dual screening (Gwet AC1: 0.76–0.83). Eight competencies were identified. The most prevalent was domain expertise and oversight of AI outputs ($\Sigma n = 123$), encompassing subject-matter mastery, systematic skepticism, source verification, and researcher accountability. Other key competencies include metacognition and decision making about AI use ($\Sigma n = 55$), ethics and academic integrity ($\Sigma n = 53$), prompt engineering for research ($\Sigma n = 38$), and reproducibility of AI use ($\Sigma n = 29$). AI literacy and technical knowledge ($\Sigma n = 16$) was explicitly identified as a risk factor when absent, with domain expertise treated as a prerequisite for meaningful critical evaluation. The findings suggest that preparing researchers to use LLMs goes beyond technical instruction, requiring an integrated set of epistemic, ethical, and methodological competencies centered on human accountability for the knowledge produced. These results have direct implications for the design of graduate programs and AI literacy initiatives.*

Resumo. *A crescente adoção de Large Language Models na pesquisa científica criou a necessidade de compreender quais competências pesquisadores e estudantes de pós-graduação precisam desenvolver para utilizar essas ferramentas de forma crítica e responsável. Esta rapid review analisou 194 artigos recuperados do Elicit e do Google Scholar (2022–2025), dos quais 40 foram selecionados para extração de competências e análise temática, após triagem dupla independente (Gwet AC1: 0,76–0,83). Oito competências foram identificadas. A mais prevalente foi o domínio do assunto e a supervisão das saídas da IA ($\Sigma n = 123$), que abrange conhecimento do tema, ceticismo sistemático,*

verificação de fontes e responsabilidade do pesquisador sobre o conteúdo produzido. Outras competências relevantes incluem metacognição e tomada de decisão sobre o uso da IA ($\Sigma n = 55$), ética e integridade acadêmica ($\Sigma n = 53$), engenharia de prompt para pesquisa ($\Sigma n = 38$) e reprodutibilidade do uso da IA ($\Sigma n = 29$). A alfabetização em IA e o conhecimento técnico ($\Sigma n = 16$) foram explicitamente identificados como fator de risco quando ausentes, com o conhecimento de domínio atuando como pré-requisito para uma avaliação crítica significativa. Os resultados sugerem que preparar pesquisadores para o uso de LLMs vai além da instrução técnica, exigindo um conjunto integrado de competências epistêmicas, éticas e metodológicas centradas na responsabilidade humana pelo conhecimento produzido. Esses resultados têm implicações diretas para o design de programas de pós-graduação e iniciativas de alfabetização em IA.

1. Introduction

Artificial Intelligence (AI) has supported teaching, learning, and assessment for years [Chen et al. 2020], stimulating discussions on digital literacy, AI literacy, critical thinking, and the skills needed in technology-mediated environments [Ng et al. 2021, Reyes and Morales 2021, Meerah et al. 2012]. Generative AI and Large Language Models (LLMs) extended this through natural language interfaces [Meyer et al. 2023, Rahman et al. 2023]. Unlike earlier educational technologies built for specific purposes, LLMs support a broad range of academic activities, from information retrieval and summarization to writing, programming, and problem solving [Meyer et al. 2023, Rahman et al. 2023, Lissack and Meagher 2024].

Software engineering education is one domain where this impact has drawn particular attention, with reported use in programming, requirements, design, testing, documentation, and learning activities [Kozov et al. 2024, Kirova et al. 2024]. Concerns have followed regarding academic integrity, overreliance on generated content, learning outcomes, and the preparation of professionals able to work responsibly in AI-assisted environments [Santos et al. 2026, Kirova et al. 2024, Kozov et al. 2024].

Beyond education, LLMs are increasingly used across the research lifecycle for literature exploration, idea generation, academic writing, coding, peer review, and data analysis [Rahman et al. 2023, Meyer et al. 2023, Reddy and Shojaee 2025, Trinkenreich et al. 2025, Trinkenreich et al. 2026]. This creates opportunities but also raises concerns regarding transparency, reproducibility, hallucinations, bias, source verification, scientific integrity, and researcher accountability [Lissack and Meagher 2024, Blau et al. 2024, Meyer et al. 2023, Rahman et al. 2023]. In software engineering research, guidelines have begun to support the rigorous use and reporting of LLM-based studies [Baltes et al. 2025, Trinkenreich et al. 2025, Trinkenreich et al. 2026].

Research traditionally requires competencies related to critical evaluation of evidence, methodological rigor, information management, ethical conduct, knowledge synthesis, and scientific communication [Meerah et al. 2012, Reyes and Morales 2021, Wohlin 2007, Pizard et al. 2021, Kitchenham et al. 2004, Kitchenham et al. 2009], while AI literacy research emphasizes understanding AI capabilities and limitations, evaluating outputs, and making informed decisions about their use [Ng et al. 2021]. How

these competencies intersect in AI-assisted research remains underexplored, and those required for the effective, critical, and responsible use of LLMs remain dispersed across the literature [Meyer et al. 2023, Rahman et al. 2023, Lissack and Meagher 2024, Trinkenreich et al. 2026].

To address this gap, this study identifies and consolidates the competencies required for the use of LLMs in scientific research, with particular attention to software engineering research, analyzing studies published between 2022 and 2025 that document real practices mediated by generative AI tools. It addresses the following research question: **RQ: What competencies are required for the effective, critical, and responsible use of large language models throughout the scientific research process?** The paper makes three contributions: it synthesizes empirical evidence on how LLMs are used across the research process, identifies and consolidates competencies associated with AI-assisted research, and discusses implications for software engineering research, graduate education, and researcher preparation.

2. Background

Scientific research is fundamentally a process of producing, evaluating, interpreting, and communicating knowledge. In software engineering, evidence-based approaches emphasize systematic, transparent, and reproducible procedures [Kitchenham et al. 2004, Kitchenham et al. 2009, Wohlin 2007], which are especially important in secondary studies such as systematic literature reviews, where conclusions depend on rigorous identification, selection, analysis, and synthesis of evidence [Kitchenham et al. 2004, Kitchenham et al. 2009, Pizard et al. 2021]. Research quality therefore depends not only on the evidence but also on the decisions researchers make throughout the process.

LLMs have introduced new possibilities for supporting research, with reported use for literature exploration, idea generation, scientific discovery, coding, academic writing, peer review, and data analysis [Rahman et al. 2023, Meyer et al. 2023, Reddy and Shojaee 2025, Trinkenreich et al. 2025, Trinkenreich et al. 2026]. In software engineering research, these tools have been incorporated into exploratory investigations, empirical studies, and evidence synthesis [Trinkenreich et al. 2025, Trinkenreich et al. 2026], helping researchers navigate increasingly large bodies of literature [Rahman et al. 2023, Meyer et al. 2023, Reddy and Shojaee 2025].

At the same time, this growing role has prompted discussions about researcher oversight, accountability, and the evaluation of outputs whose internal decision processes are difficult to inspect [Lissack and Meagher 2024, Blau et al. 2024]. Reported concerns include hallucinations, inaccurate references, unsupported claims, bias, and limited transparency [Meyer et al. 2023, Rahman et al. 2023, Lissack and Meagher 2024]. Such issues are particularly relevant in evidence synthesis, where judgments made during study selection, quality assessment, data extraction, coding, and synthesis directly shape the resulting body of evidence, and errors may propagate to subsequent analysis and interpretation [Kitchenham et al. 2004, Kitchenham et al. 2009, Baltes et al. 2025].

Recent guidance for LLM-based software engineering studies reflects these concerns by emphasizing transparent reporting of AI use, reproducibility, documentation of prompts and interactions, and careful evaluation of outputs [Baltes et al. 2025]. Rather than treating LLMs solely as productivity tools, the literature increasingly

examines their influence on retrieval, interpretation, synthesis, and communication [Lissack and Meagher 2024, Blau et al. 2024, Trinkenreich et al. 2026]. Consequently, AI-assisted research must preserve scientific rigor, maintain accountability for research decisions, and keep conclusions grounded in evidence.

3. Method

This study aims to identify and consolidate the competencies required for the use of LLMs in scientific research, with particular attention to software engineering research. To achieve this objective, we conducted a rapid review, an evidence synthesis approach that follows the main principles of systematic reviews while enabling timely investigation of emerging topics [Cartaxo et al. 2020]. The review comprised four phases: study identification, study selection, data extraction, and thematic analysis.

3.1. Study Selection

Initially, 194 articles were selected for analysis, of which 100 were obtained from the first 10 pages of Google Scholar results and 94 were retrieved using the Elicit tool, already excluding duplicates also returned by Google Scholar. In Elicit, an AI-assisted semantic search was conducted in scientific databases, with studies ranked according to their thematic relevance. In Google Scholar, the first 10 pages of results were considered to broaden the coverage of the literature, including possible grey literature, preprints, and relevant publications not indexed in traditional databases. We used Elicit and Google Scholar rather than curated databases such as Scopus, Web of Science, the ACM Digital Library, and IEEE Xplore because, as a rapid review of a fast-moving topic, we prioritized timely retrieval over exhaustive coverage. The complete search strings and the protocol adopted in this stage are reported in the artifacts (Section 7).

After the initial selection of articles from the respective search sources, the research team, composed of four researchers, met to define the screening strategy. The screening process was organized into two distinct groups, each composed of two researchers. Group 1 was responsible for analyzing the first 97 articles, while Group 2 was responsible for analyzing the remaining 97 articles. The methodological execution of the selection and screening process is summarized in Figure 1(a).

The screening of studies was conducted based on previously defined inclusion and exclusion criteria. Studies were included if they: (I1) were published between 2022 and 2025 and documented scientific research practices mediated by generative AI tools; (I2) reported at least one concrete use of AI within a stage of the research process; (I3) described AI tools together with their purpose, research context, or level of autonomy; (I4) provided sufficient information to infer methodological adaptations or competencies associated with AI use; and (I5) were written in English or Portuguese, the paper that not attend the inclusion criteria are excludes.

In the first screening stage, the researchers assessed the titles and abstracts of the selected studies to determine their relevance to the use of artificial intelligence in scientific research. This stage resulted in 48 articles selected for full-text analysis, 37 from Group 1 and 11 from Group 2. In the second stage, these articles were analyzed in greater depth according to the inclusion and exclusion criteria defined for the study. As a result, 40 articles were selected for data extraction.

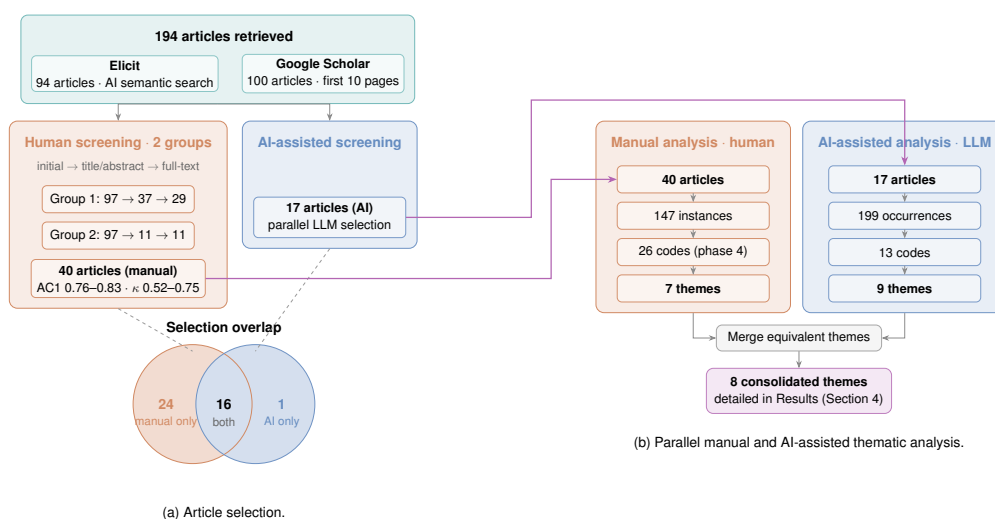


Figure 1. Overview of the review process

Note. (a) Article selection: 194 retrieved articles were screened independently by humans (40 selected) and by an AI-assisted procedure (17 selected). (b) Thematic analysis: 7 manual and 9 AI-assisted themes were merged into 8 consolidated competencies.

To complement the manual screening, an AI-supported validation was conducted on a set of 194 articles using the same inclusion and exclusion criteria. The application identified 17 articles already included in the manual review and 1 additional article not previously selected by the researchers. However, after reviewing this article, the researchers decided not to include it in the final corpus, as it did not fully meet the acceptance criteria. Therefore, the final set of studies remained composed of 40 articles.

To evaluate the reliability of the screening process, three chance-corrected agreement coefficients were computed over the reviewers' independent include/exclude classifications: Gwet's AC1 [Gwet 2008], Krippendorff's Alpha [Krippendorff 2018], and Weighted Cohen's Kappa [Cohen 1968]. Gwet's AC1 values ranged from 0.76 to 0.83, indicating substantial to almost perfect agreement between reviewers, following the interpretation benchmarks proposed by Landis and Koch [Landis JRKoch 1977]. Krippendorff's Alpha and Weighted Cohen's Kappa values ranged from 0.52 to 0.75, indicating moderate to substantial agreement. In this study, Gwet's AC1 was adopted as the primary agreement indicator because it is less sensitive to prevalence effects and category imbalance than traditional kappa-based statistics [Gwet 2008].

3.2. Data Extraction and Data Analysis

The extracted competency instances were analyzed using thematic analysis [Cruzes and Dyba 2011]. We adopted an inductive approach in which themes emerged from the extracted data. Coding of the 147 competency instances followed an iterative process of comparison, refinement, and merging of codes representing similar skills, knowledge areas, responsibilities, or practices, which produced 26 codes in the fourth coding phase. These codes were subsequently consolidated into 7 higher-level themes describing the competencies reported in the literature for the effective and responsible use of LLMs in scientific research, which are available in the repository at zenodo.21313656. All coding decisions, code revisions, theme definitions, and final interpretations were performed and validated by the researchers.

In addition to the manual analysis, a complementary thematic analysis was conducted with an LLM-based approach, whose implementation, including the model, version, parameters, and prompts, is documented in the artifacts (Section 7). Operating independently over the 17 articles it had selected, the software tool extracted 199 competency occurrences and generated 9 themes, as summarized in Figure 1(b). The AI-assisted analysis served as a complementary coding lens rather than an autonomous source of conclusions. Its themes were compared against the codes defined manually by the researchers, who reviewed and validated them and retained full responsibility for every coding decision, theme definition, and final interpretation.

Finally, by comparing the themes generated through the AI-supported analysis with those defined manually by the researchers, conceptual similarities and overlaps between categories were identified. The researchers refined the nomenclature produced by both analyses, grouping semantically related themes under a single label. This procedure reduced redundancies, standardized the terminology, and strengthened the coherence of the final thematic structure, consolidating the results into 8 final themes, which are presented in the Results.

The merging process was conducted by the two senior researchers, who independently reviewed the themes produced by both analyses and proposed equivalences based on conceptual overlap in definition and scope. Cases of disagreement were resolved through discussion until consensus was reached. No formal inter-rater agreement metric was calculated for this phase, as the decisions involved interpretive judgment about theme boundaries rather than binary classification.

The criterion for consolidation was conceptual equivalence. Themes from the manual and AI-assisted analyses that described the same competency were merged under a single label, while themes that appeared in only one of the analyses were kept as separate competencies. Applying this criterion to the seven manual themes and the nine AI-assisted themes resulted in eight final competencies. The complete mapping between the original themes and the eight consolidated competencies is available in the repository at [zenodo.21313656](https://zenodo.org/record/21313656).

3.3. Threats to Validity

Our findings should be interpreted considering the limitations inherent to the adopted method. First, as a rapid review, this study relied on Elicit and Google Scholar rather than curated bibliographic databases such as Scopus, Web of Science, the ACM Digital Library, and IEEE Xplore. Excluding these sources may have reduced both coverage and reproducibility, since peer-reviewed studies indexed only in them could have been omitted, and the semantic ranking performed by Elicit is not fully transparent, which further limits the replicability of the retrieval. Second, competency identification and coding involved qualitative interpretation, although independent screening, agreement assessment, and iterative discussions among the researchers were used to mitigate this threat. Third, the manual and AI-assisted analyses extracted competencies at different densities, approximately 3.7 instances per article in the manual analysis (147 across 40 articles) and 11.7 per article in the AI-assisted analysis (199 across 17 articles), and covered different numbers of studies. In addition, the pooled counts were not de-duplicated across the two analyses, so an instance identified in both contributed to each source. The pooled frequencies (Σn) should therefore be read as an indication of the relative salience of each

competency rather than a precise measure of prevalence or a count of unique instances. Finally, the identified competencies reflect a rapidly evolving body of literature and should be interpreted within the context of publications available between 2022 and 2025.

4. Results

Our analysis identified eight competencies associated with the use of AI in research. These competencies describe the knowledge, skills, responsibilities, and decision making processes reported in the literature as necessary for the effective, critical, and responsible use of AI throughout the research lifecycle. Table 1 presents the identified competencies, their frequencies, definitions, and representative evidence extracted from the reviewed studies. Because the manual and AI-assisted analyses were pooled without de-duplication, the reported frequencies (Σn) should be read as indicators of the relative salience of each competency rather than absolute counts of unique instances. The following sections describe each competency and discuss how they were characterized across the reviewed studies.

4.1. Domain Expertise and Oversight of AI Outputs.

Domain expertise and oversight of AI outputs was the most frequently reported competency. This competency refers to the ability to critically assess AI generated outputs, verify information against external sources, identify inaccuracies and hallucinations, and retain responsibility for research decisions and conclusions. The reviewed studies suggest that this evaluation depends on sufficient domain expertise, since detecting subtle errors, superficial reasoning, or misleading claims requires subject-matter knowledge that may exceed what the AI tool itself provides. Researchers without an adequate command of the topic may accept plausible but flawed outputs, as the ability of the tools can exceed the skills of the person using them. Across the reviewed studies, domain knowledge was portrayed as essential for distinguishing meaningful contributions from superficial ones, recognizing which hypotheses merit further exploration, and arriving at ground-truth conclusions from AI responses. For example, a researcher using an LLM to summarize studies for a literature review would still need enough domain expertise to verify claims against the original sources, detect misrepresented findings, and judge whether the synthesis meaningfully advances the field before incorporating it into a manuscript.

4.2. Metacognition and Decision Making Regarding AI Use

The second most recurrent competency concerns the ability to determine when, where, and how AI should be used during research activities. This competency positions researchers as active decision makers who evaluate the appropriateness of AI support according to task characteristics, methodological requirements, and research goals. The literature frequently portrayed AI as a collaborator rather than a replacement for researchers. Studies highlighted the importance of maintaining humans in the loop, combining human and AI contributions, and deciding which tasks should remain under direct human control. Effective AI use therefore depends on researchers' ability to judge the suitability of AI support in different contexts. For example, a researcher might use AI to organize interview transcripts or generate preliminary themes while reserving theoretical interpretation and conclusion development for themselves.

4.3. Ethics, Privacy, and Academic Integrity

Ethics, privacy, and academic integrity emerged as another highly recurrent competency. This competency encompasses responsible AI use, protection of sensitive information, transparency regarding AI involvement, appropriate attribution of authorship, prevention of plagiarism, and awareness of legal and societal implications. The reviewed studies discussed ethical considerations from multiple perspectives, including data privacy, bias, ownership of AI generated content, responsible disclosure of AI use, and academic honesty. Collectively, these studies suggest that ethical considerations should accompany AI use throughout the entire research process rather than being addressed only at publication time. For instance, a researcher analyzing proprietary organizational data with AI tools would need to ensure that confidential information is protected, AI use is disclosed appropriately, and final interpretations remain under human responsibility.

4.4. Prompt Engineering for Research

Prompt engineering refers to the ability to formulate effective instructions that guide AI systems toward useful and reliable outputs. This competency includes defining objectives, providing context, specifying expected outputs, and iteratively refining interactions according to task requirements. Evidence related to this competency emphasized that prompt quality directly influences output quality. Studies described prompt engineering as a structured activity involving clear instructions, contextual information, iterative refinement, and explicit specification of desired outcomes. Researchers were frequently encouraged to refine prompts when initial outputs were unsatisfactory. For example, a researcher seeking themes in interview data may initially obtain generic results and subsequently refine the prompt by specifying the study context, participant population, and analytical objective.

4.5. Reproducibility and Reporting of AI Use

Reproducibility and reporting of AI use concerns the documentation and disclosure of how AI tools are employed during research activities. This competency includes reporting prompts, model versions, configurations, and AI assisted tasks to promote transparency, accountability, and reproducibility. The reviewed studies frequently emphasized the importance of documenting AI use. Reporting practices were discussed as mechanisms for enabling replication, supporting transparency, and helping readers understand how AI contributed to research outcomes. Several studies also referred to emerging institutional expectations regarding disclosure of AI use in scientific work. A research team using an LLM to assist with article screening in a literature review, for example, would be expected to document the prompts, model version, and validation procedures used throughout the process.

4.6. Methodological and Experimental Design with AI

Methodological and experimental design with AI refers to the ability to incorporate AI into research planning and study design. This competency includes defining research objectives, specifying constraints, describing datasets adequately, and designing workflows that account for the strengths and limitations of AI systems. The reviewed studies suggest that successful AI use depends substantially on the quality of the research

design surrounding it. Evidence highlighted the importance of clearly formulated research goals, well defined datasets, structured workflows, and explicit methodological constraints. These elements influence the quality and reliability of AI generated outputs. For example, a researcher designing an AI assisted study would need to define research questions, describe the available data, and establish validation procedures before introducing AI into the workflow.

4.7. AI Literacy and Technical Knowledge

AI literacy and technical knowledge refers to understanding how AI systems function, including their capabilities, limitations, risks, and technical characteristics. This competency enables researchers to make informed decisions regarding tool selection, interpretation of outputs, and appropriate use of AI systems. The literature suggests that effective AI use requires more than operational familiarity with AI tools. Researchers were expected to understand model limitations, recognize potential risks, evaluate available tools, and determine whether specific systems were suitable for particular research tasks. Several studies also emphasized the need for formal training and AI related competencies within research education. For example, a researcher selecting an AI tool for qualitative coding would need to understand how the model operates, its limitations, and the potential implications of its outputs for the validity of the study.

4.8. Data Analysis and Interpretation with AI

The final competency concerns the use of AI to support quantitative, qualitative, and exploratory data analysis activities. It includes generating analytical code, supporting statistical analyses, identifying patterns, processing research materials, and assisting with interpretation while preserving human responsibility for final conclusions. Although less frequently reported than other competencies, several studies described AI as a useful tool for supporting analytical activities. Reported applications included generating code, supporting statistical procedures, identifying patterns in datasets, and assisting with exploratory analyses. At the same time, studies consistently emphasized that human researchers remain responsible for evaluating analytical quality and interpreting results. For example, a researcher using AI generated code to analyze software repository data would still need to verify the correctness of the analysis, assess the suitability of the analytical approach, and determine whether the resulting findings are supported by the evidence.

To improve traceability, Table 1 is complemented by a competency-to-study matrix that maps each of the eight consolidated competencies to the primary studies from which its instances were extracted; the full matrix and the study identifiers are available in the artifacts (Section 7).

4.9. Answer to the Research Question

Our findings indicate that the effective, critical, and responsible use of LLMs in scientific research requires competencies spanning the entire AI-assisted process, from understanding AI capabilities and limitations, designing workflows, and formulating prompts, to evaluating outputs, conducting AI-assisted analyses, ensuring transparency, and maintaining ethical responsibility. The eight competencies suggest that AI-assisted research is not simply about adopting new tools, but about developing expertise that combines technical

Table 1. Competencies for AI-Assisted Research

Competency	Σn	Description	Representative Evidence
Domain Expertise and Oversight of AI Outputs	123	Ability to critically assess AI-generated outputs, verify information, and identify inaccuracies, supported by sufficient domain expertise to recognize errors and retain responsibility for research conclusions.	“Researchers should always evaluate critically the generated results and compare them with existing knowledge or expert opinions to ensure the validity of conclusions” (ID41). “Detecting such misleading statements can be challenging, especially when the abilities of the tools exceed the skills of the person using them” (ID56).
Metacognition and Decision Making Regarding AI Use	55	Ability to determine when, where, and how AI should be used during research activities and when human expertise remains necessary.	“Its scientist-in-the-loop design prioritizes cognitive collaboration over full automation” (ID151). “We envision these technologies as augmenting rather than replacing human researchers” (ID67).
Ethics, Privacy, and Academic Integrity	53	Ability to use AI responsibly while considering privacy, authorship, transparency, plagiarism, and broader ethical implications.	“Ethical concerns are well-documented in existing studies that highlight issues such as data privacy, misinformation, and unintended biases” (ID49). “Such initiatives can ensure the benefits of AI augmentation without compromising academic integrity” (ID40).
Prompt Engineering for Research	38	Ability to formulate, refine, and contextualize prompts to obtain reliable and relevant AI outputs.	“Prompt engineering is the process of creating clear, concise and easily understandable prompts” (ID12). “GPT-4’s performance can be improved by asking questions in an iterative manner or providing additional context” (ID41).
Reproducibility and Reporting of AI Use	29	Ability to document and disclose AI use, including prompts, tools, and configurations, to support transparency and reproducibility.	“Authors are advised to retain the dialogue records to facilitate reviewers, editors and readers in replicating and understanding the process” (ID12). “The use of AI to generate texts needs to be explicitly acknowledged” (ID63).
Methodological and Experimental Design with AI	20	Ability to integrate AI into research design through clear objectives, constraints, workflows, and validation procedures.	“Research goal: Scientist describes a research goal along with preferences, experiment constraints, and other attributes” (ID142). “The process reliability depends on the formulation of the research goal and the description of the dataset” (ID5).
AI Literacy and Technical Knowledge	16	Understanding of AI capabilities, limitations, risks, models, and technical characteristics required for informed use.	“Researchers should be equipped with proper training to utilize chatbots and other AI tools effectively and ethically” (ID40). “The selection of a suitable base LLM model for finetuning is of utmost importance” (ID93).
Data Analysis and Interpretation with AI	7	Ability to employ AI to support analytical activities while preserving human responsibility for interpretation and conclusions.	“Researchers consider using GenAI to support statistical analysis and help recognize patterns in data as good research practices” (ID13). “For data analysis and visualization tasks, we also report on the quality of the analysis and visualization” (ID56).

knowledge, methodological rigor, critical thinking, and scientific judgment. Preparing researchers for this therefore represents an educational challenge requiring explicit training and guidance. In software engineering education, these findings point to the need to complement traditional research and technical competencies with AI literacy, human oversight, ethical decision making, transparency, and the critical evaluation of AI-generated outputs, so that future researchers and practitioners can use AI in ways that preserve research quality, integrity, and trustworthiness.

5. Discussion

Although LLMs introduce new forms of assistance throughout the research lifecycle, our findings suggest that they do not alter the epistemic foundations of research. Rather than transferring responsibility for knowledge production to AI systems, the identified competencies indicate that researchers remain responsible for evaluating evidence, assessing

validity, interpreting results, and justifying conclusions. AI-assisted research thus appears less as a replacement of traditional research expertise and more as a redistribution of where and how that expertise is exercised. The prominence of competencies related to domain expertise, oversight, metacognitive decision making, ethics, and reproducibility suggests that the challenge is not obtaining outputs from AI systems, but governing, evaluating, and appropriately integrating them into scientific work.

5.1. Comparing Results with the Literature

Our findings align with existing discussions regarding AI-assisted research, which frequently emphasize concerns related to transparency, reproducibility, hallucinations, bias, source verification, scientific integrity, and researcher accountability [Meyer et al. 2023, Rahman et al. 2023, Lissack and Meagher 2024, Blau et al. 2024, Baltes et al. 2025]. Similarly, our results are consistent with traditional conceptions of research competence that emphasize critical evaluation of evidence, methodological rigor, ethical conduct, and scientific judgment as essential components of trustworthy research [Kitchenham et al. 2004, Kitchenham et al. 2009, Wohlin 2007, Meerah et al. 2012, Reyes and Morales 2021]. The predominance of competencies related to domain expertise, oversight, ethics, and decision making suggests that these foundational research competencies remain relevant even as AI becomes increasingly integrated into research workflows.

At the same time, our findings differ from portions of the AI literacy literature that place stronger emphasis on understanding AI technologies and their technical operation [Ng et al. 2021]. Although AI literacy, technical knowledge, and prompt engineering emerged as relevant competencies, they were less recurrent than competencies associated with domain expertise, oversight, accountability, and methodological decision making. This pattern suggests that effective AI-assisted research may depend less on technical mastery of AI systems and more on researchers' ability to exercise judgment regarding when AI should be used, how outputs should be validated, and under what conditions generated content can be incorporated into scientific knowledge claims.

The novelty of this study lies in integrating discussions that frequently appear dispersed across different streams of literature. Previous work has addressed AI literacy, research integrity, responsible AI, prompt engineering, transparency, reproducibility, and methodological rigor largely as separate concerns. Our findings suggest that these topics collectively describe a broader competency profile required for AI-assisted research. By consolidating them into a single evidence-based framework, this study provides an integrated understanding of the knowledge, skills, responsibilities, and decision making processes needed to conduct research effectively, critically, and responsibly in AI-assisted environments.

5.2. Implications

Research. This study contributes an evidence-based consolidation of competencies associated with AI-assisted research. Existing literature has discussed opportunities, risks, applications, and guidelines for LLM use, but limited attention has been given to systematically identifying the knowledge, skills, responsibilities, and decision making processes researchers require to engage with these technologies responsibly. The identified competencies provide a foundation for future work on how they are developed, applied,

and manifested across research contexts, disciplines, and methodological approaches, and may support competency models, researcher assessment approaches, and guidelines for responsible practice. More broadly, they contribute to discussions on the role of human judgment, accountability, and expertise in increasingly AI-mediated knowledge production.

Software Engineering Education. Preparing future researchers and graduate students for AI-assisted environments requires more than technical instruction on AI tools. By clarifying the knowledge, skills, responsibilities, and decision making processes associated with AI-assisted research, these findings may support instructors in designing research methods courses, graduate seminars, and AI literacy initiatives, and in defining learning objectives, structuring training activities, and setting expectations for responsible AI use across the research lifecycle. For students, particularly at the graduate level, they offer a clearer understanding of what is expected in increasingly AI-assisted environments. Incorporating these competencies into research training may help prepare researchers who engage with AI while maintaining the rigor, integrity, and accountability expected of scientific research.

6. Conclusion

This study investigated the competencies required for the effective, critical, and responsible use of Large Language Models throughout the scientific research process. Through a rapid review of studies published between 2022 and 2025, we identified eight competencies associated with AI-assisted research, spanning domain expertise and oversight, decision making regarding AI use, ethics and academic integrity, prompt engineering, reproducibility, methodological design, AI literacy, and data analysis. Our findings suggest that effective use of LLMs in research depends not only on understanding AI systems but also on maintaining the forms of judgment, accountability, and methodological rigor traditionally associated with scientific research. The prominence of competencies related to domain expertise, oversight, ethics, and decision making indicates that researchers remain responsible for assessing evidence, validating outputs, and justifying conclusions even when AI systems participate throughout the research lifecycle. Beyond identifying individual competencies, this study consolidates discussions that frequently appear dispersed across the literature on AI literacy, responsible AI, research integrity, reproducibility, and AI-assisted research practices. The resulting synthesis provides an evidence-based view of the knowledge, skills, responsibilities, and decision making processes associated with conducting research in increasingly AI-assisted environments. Findings also suggest that preparing researchers for these environments requires attention not only to technical aspects of AI use but also to the development of competencies that support critical engagement with AI-generated outputs and responsible integration of AI into research workflows. **Future Work.** Future studies may investigate how these competencies manifest across different disciplines, research methodologies, and stages of the research lifecycle. Additional research may also explore how researchers develop these competencies over time and how they can be assessed and supported through training initiatives and institutional guidance. Such investigations may contribute to the development of competency models, researcher preparation strategies, and guidelines that support the responsible use of AI in scientific research.

7. Artifacts

All materials underlying this review are available in the replication package at zenodo.21313656, including the complete search strings for Elicit and Google Scholar, the screened studies, the coding files of both analyses, a competency-to-study traceability matrix, and the full specification of the AI-assisted procedures (model, version, parameters, and prompts).

Declaration on the Use of Artificial Intelligence

Generative AI tools supported parts of the qualitative data analysis, as described in the methodology, and assisted with grammar and language revision of the manuscript. All AI-generated outputs and suggestions were reviewed, verified, and edited by the authors, who retain full responsibility for all analytical decisions, findings, and conclusions.

References

- Baltes, S., Angermeir, F., Arora, C., Barón, M. M., Chen, C., Böhme, L., Calefato, F., Ernst, N., Falessi, D., Fitzgerald, B., et al. (2025). Guidelines for empirical studies in software engineering involving large language models. *arXiv preprint arXiv:2508.15503*.
- Blau, W., Cerf, V. G., Enriquez, J., Francisco, J. S., Gasser, U., Gray, M. L., Greaves, M., Grosz, B. J., Jamieson, K. H., Haug, G. H., et al. (2024). Protecting scientific integrity in an age of generative ai.
- Cartaxo, B., Pinto, G., and Soares, S. (2020). Rapid reviews in software engineering. In *Contemporary empirical methods in software engineering*, pages 357–384. Springer.
- Chen, L., Chen, P., and Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE access*, 8:75264–75278.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4):213.
- Cruzes, D. S. and Dyba, T. (2011). Recommended steps for thematic synthesis in software engineering. In *2011 international symposium on empirical software engineering and measurement*, pages 275–284. IEEE.
- Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, 61(1):29–48.
- Kirova, V. D., Ku, C. S., Laracy, J. R., and Marlowe, T. J. (2024). Software engineering education must adapt and evolve for an llm environment. In *Proceedings of the 55th ACM technical symposium on computer science education v. 1*, pages 666–672.
- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., and Linkman, S. (2009). Systematic literature reviews in software engineering—a systematic literature review. *Information and software technology*, 51(1):7–15.
- Kitchenham, B. A., Dyba, T., and Jorgensen, M. (2004). Evidence-based software engineering. In *Proceedings. 26th International Conference on Software Engineering*, pages 273–281. IEEE.

- Kozov, V., Ivanova, G., and Atanasova, D. (2024). Practical application of ai and large language models in software engineering education. *International Journal of Advanced Computer Science & Applications*, 15(1).
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage publications.
- Landis JRKoch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159174.
- Lissack, M. and Meagher, B. (2024). Navigating the future of large language models in scientific research: Opportunities, challenges, and ethical considerations. *Challenges, and Ethical Considerations (September 02, 2024)*.
- Meerah, T. S. M., Osman, K., Zakaria, E., Ikhsan, Z. H., Krish, P., Lian, D. K. C., and Mahmod, D. (2012). Measuring graduate students research skills. *Procedia-Social and Behavioral Sciences*, 60:626–629.
- Meyer, J. G., Urbanowicz, R. J., Martin, P. C., O'Connor, K., Li, R., Peng, P.-C., Bright, T. J., Tatonetti, N., Won, K. J., Gonzalez-Hernandez, G., et al. (2023). Chatgpt and large language models in academia: opportunities and challenges. *BioData mining*, 16(1):20.
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., and Qiao, M. S. (2021). Conceptualizing ai literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2:100041.
- Pizard, S., Acerenza, F., Otegui, X., Moreno, S., Vallespir, D., and Kitchenham, B. (2021). Training students in evidence-based software engineering and systematic reviews: a systematic review and empirical study. *Empirical Software Engineering*, 26(3):50.
- Rahman, M., Terano, H. J. R., Rahman, N., Salamzadeh, A., and Rahaman, S. (2023). Chatgpt and academic research: A review and recommendations based on practical examples. *Journal of Education, Management and Development Studies*, 3(1):1–12.
- Reddy, C. K. and Shojaee, P. (2025). Towards scientific discovery with generative ai: Progress, opportunities, and challenges. In *Proceedings of the AAAI conference on artificial intelligence*, volume 39, pages 28601–28609.
- Reyes, C. E. G. and Morales, L. D. G. (2021). Research competencies mediated by technologies: A systematic mapping of the literature. *Education in the Knowledge Society (EKS)*, 22:e23897–e23897.
- Santos, R. d. S., Santos, I., Bento, M., Destefanis, G., Magalhães, C., and Wessel, M. (2026). Llm use, cheating, and academic integrity in software engineering education. *International Conference on the Foundations of Software Engineering (FSE)*.
- Trinkenreich, B., Calefato, F., Blincoe, K., Wivestad, V. T., Alves, A. P. S., Araújo, J. C., Araújo, M. C., Tell, P., Kalinowski, M., Zimmermann, T., et al. (2026). Taking a pulse on how generative ai is reshaping the software engineering research landscape. *arXiv preprint arXiv:2604.11184*.
- Trinkenreich, B., Calefato, F., Hanssen, G., Blincoe, K., Kalinowski, M., Pezzè, M., Tell, P., and Storey, M.-A. (2025). Get on the train or be left on the station: Using

llms for software engineering research. In *Proceedings of the 33rd ACM International Conference on the Foundations of Software Engineering*, pages 1503–1507.

Wohlin, C. (2007). Empirical software engineering: Teaching methods and conducting studies. In *Empirical Software Engineering Issues. Critical Assessment and Future Directions: International Workshop, Dagstuhl Castle, Germany, June 26-30, 2006. Revised Papers*, pages 135–142. Springer.