A Metaprotocol For a Family of Rapid Multivocal Reviews of Generative AI in the Software Industry

Sabrina Rocha
Program Systems Engineering and Computer Science
Federal University of Rio de Janeiro
Rio de Janeiro, RJ - Brazil
sabrinarocha@cos.ufrj.br

Rodrigo Feitosa
Program Systems Engineering and Computer Science
Federal University of Rio de Janeiro
Rio de Janeiro, RJ - Brazil
rfeitosa@cos.ufrj.br

Larissa Galeno
Program Systems Engineering and Computer Science
Federal University of Rio de Janeiro
Rio de Janeiro, RJ - Brazil
galeno@cos.ufrj.br

Guilherme H. Travassos
Program Systems Engineering and Computer Science
Federal University of Rio de Janeiro
Rio de Janeiro, RJ - Brazil
ght@cos.ufrj.br

ABSTRACT

Context: With the growing interest in solutions based on Generative Artificial Intelligence (GenAI) applied to SE, a need to identify solutions beyond conceptual proposals and prototypes arises. In this rapidly evolving landscape, exploring methodological strategies that enable the identification of GenAI-based solutions used in the industry becomes essential. **Goals:** In this paper, we report on the experience of developing a metaprotocol for a family of Rapid Multivocal Reviews (RMRs) to identify operational GenAI-based technologies being used in the software industry (i) across the phases of the software development lifecycle and (ii) in the planning and executing SE empirical studies. **Method:** A metaprotocol was developed through an interactive process combining GenAI's creative support with software engineers' feedback. GenAI assisted in defining the stages of the RMRs family, while discussion cycles with software engineers contributed to refining and validating it. **Results:** Six RMR instances have been organized by covering the perspectives of Software Requirements, Design, Test, Coding, Management, and Empirical Methods described in SWEBOK 4.0. It allowed guiding the execution of the RMRs and highlighted the challenges of organizing the metaprotocol and justifying the decision-making. **Conclusion:** Using a metaprotocol to support the observation of industrial experience using GenAI in software projects through RMRs will contribute to organizing a body of knowledge regarding GenAI-based solutions currently available and used in the software industry to support the SWEBOK 4.0 practices.

KEYWORDS

Generative Artificial Intelligence (GenAI), Software Engineering, Rapid Multivocal Review, Solutions

1 Introduction

In recent years, the integration of Generative Artificial Intelligence (GenAI) has been extensively investigated and applied across various domains [2], including Software Engineering (SE) — both in industrial contexts, throughout the software development lifecycle [4], and in academic settings, through research in the field [17]. Within the software development lifecycle, GenAI-based solutions have shown promise in automating repetitive and manual tasks and fostering creativity [10]. Its potential extends beyond automation in the academic realm, contributing to a deeper understanding of developers' experiences, team dynamics, and the socio-technical interactions involved in software system development [17].

This scenario has sparked growing interest among researchers and practitioners in proposing new solutions and analyzing the practical adoption of these tools. Despite the field's rapid growth, there is still a lack of systematized evidence regarding the use of operational GenAI solutions in both industry and academia—whether throughout the software development lifecycle or in the execution of empirical studies. Moreover, many existing studies use AI tools to generate and evaluate source code [14]. However, despite its relevance, coding constitutes only a small portion of SE activities across the lifecycle.

In this context, secondary studies play a fundamental role in synthesizing existing knowledge, particularly Multivocal Literature Reviews (MLRs) [5] and Rapid Reviews (RRs) [3]. In this article, we describe the experience of developing a research metaprotocol designed to guide a family of Rapid Multivocal Reviews (RMRs). This methodological adaptation combines principles from both approaches to investigate emerging phenomena based on scientific and grey literature. To date, no reports have been found of the RMRs family being systematically planned and conducted in a coordinated manner to address different dimensions of a single theme in SE. The proposed RMRs family aims to identify GenAI-based technologies employed in industry, drawing upon six knowledge areas from the Software Engineering Body of Knowledge (SWEBOK v4.0), a widely recognized document for systematizing the core domains and practices of SE [9]: software requirements, design, software testing, coding, software management, and empirical methods and experimental techniques.

Therefore, this article aims to present the initial stages of this research: a proposal for a family of RMR to investigate GenAI in SE. By sharing the insights gained, methodological decisions, and challenges encountered during the development of the meta protocol, this article seeks to contribute to future secondary study initiatives supported by GenAI while also providing a structured foundation for rapid investigations in SE.

The remainder of this paper is organized as follows: section 2 depicts the method and protocol; section 3 shows the results of the study; a discussion on the results and contributions of the study is presented in section 4; section 5 presents related work; finally, section 6 presents the threats to validity, and section 7 concludes the manuscript with final remarks and future work.

2 Research Method

This study is grounded on three main pillars: (i) established methodologies for conducting RRs and MLRs, (ii) recent approaches to the use of Generative Artificial Intelligence (GenAI) in secondary studies, and (iii) emerging applications of GenAI in SE.

The first pillar is supported by the principles discussed by Cartaxo et al. [3], which explore the role of RR in SE, and by the guidelines proposed by Garousi et al. [5] for conducting MLRs, including the use of grey literature. The second pillar highlights the emerging role of GenAI as a support tool for planning, performing, and analyzing secondary studies. Recent research has shown that Large Language Models (LLMs) can act as creative and assistive agents in the methodological design of reviews, the formulation of research questions, the construction of search strings, and the initial screening of studies [6, 16]. The third pillar refers to the increasing adoption of GenAI-based solutions across different domains of SE. Tools for automated code generation, requirements analysis, testing, refactoring, and managerial decision support are becoming more prevalent in software development [17, 19]. Despite this growth, there is still a lack of synthesized studies that identify which tools are being used in the industry, how they are applied, and to what extent they demonstrate maturity, practical applicability, and coverage across the software development lifecycle.

In response to this gap, this study proposes a structured approach in the form of a family of RMRs to identify and analyze GenAI-based tools from multiple thematic perspectives—such as requirements, design, testing, coding, management, and empirical methods—as defined by SWEBOK v4.0. It is worth noting that quality is a transversal dimension across all these areas, directly impacting each stage of the software development process [9].

This research followed three steps to the structured metaprotocol to guide the family of RMRs: (i) build a first version of the metaprotocol with the assistance of GenAI (ChatGPT-4o[1]); (ii) iterate the first version with software engineers practitioners, collecting feedback and improve it; (iii) define the version of the metaprotocol to start the secondary studies. We followed eight steps to define the metaprotocol, as shown in Figure 1.

3 Family of Rapid Multivocal Reviews

In a context marked by the rapid evolution of GenAI and the growing interest of industry and academia in its applications, we identified a significant gap in the literature: the lack of structured and practical studies that reveal how GenAI-based solutions are being applied in SE. This observation motivated the proposal of this work — the development of a family of RMRs aimed at synthesizing and analyzing GenAI-based solutions applied across various SE subareas.

[1]ChatGPT: https://chatgpt.com/

The development of the RMR family began with creating a protocol focused on the domain of Requirements Engineering (RE). The initial protocol was defined by two researchers who structured an RR to investigate how GenAI has supported RE activities. From that point on, the protocol was gradually refined and expanded in collaboration with eleven additional researchers. The research team incorporated grey literature and additional subareas of Software Engineering, as outlined in SWEBOK v4.0 [9]. This process contributed to developing a metaprotocol to guide a family of RMRs.

The researchers were grouped to their familiarity with the instances, which were as follows: **software requirements** (three researchers), which encompasses both the expression of needs and constraints of a solution and the activities involved in defining and maintaining those requirements throughout the software lifecycle; **software design** (three researchers), which consists of transforming requirements into a technical architecture composed of components, interfaces, and interactions; **software testing** (three researchers), which focuses on the dynamic validation of expected system behaviors through selected test cases; **coding** (two researchers), which concerns the implementation and maintenance of source code based on design artifacts; **software management** (two researchers), which includes planning, measurement, control, coordination, and risk management to ensure the efficient delivery of software products and services; and **empirical methods** (two researchers), which refers to the use of experiments, observations, and data analysis to evaluate solutions and support evidence-based decision-making. All instances were delineated based on the structure of SWEBOK v4.0 and adapted to suit GenAI-focused investigations' goals.

The metaprotocol was constructed iteratively and collaboratively, involving 15 participants: one senior researcher, five doctoral-level researchers, seven master's researchers, and two final-year undergraduate students in computer science. Additionally, a GenAI tool (ChatGPT-4o) was employed as a creative support throughout the development stages. To ensure the quality and practical applicability of the decisions, all AI-generated suggestions were subjected to interactive validation sessions with software engineers. Four sessions were held, each lasting approximately three hours, dedicated to the analysis and validation of all instances of the metaprotocol before the execution of the RMRs on the defined sources. During these sessions, the AI's contributions were critically discussed, and the participants applied their technical and practical expertise to interpret and refine the suggestions whenever necessary. The effectiveness of GenAI in this process was directly dependent on the researchers' ability to understand the principles of RRs and MLRs, as well as the characteristics of the defined instances.

The resulting metaprotocol, a product of this collaborative process among researchers of software engineering and GenAI, is publicly available in *<https://figshare.com/s/3344a6c730aa181c0b9d>* as an open, reusable resource aligned with best practices for conducting secondary studies in SE. We hope this sharing and the information in the following subsections, which describe each stage of the metaprotocol, can contribute to the organization of a family of RMRs in the SE community.

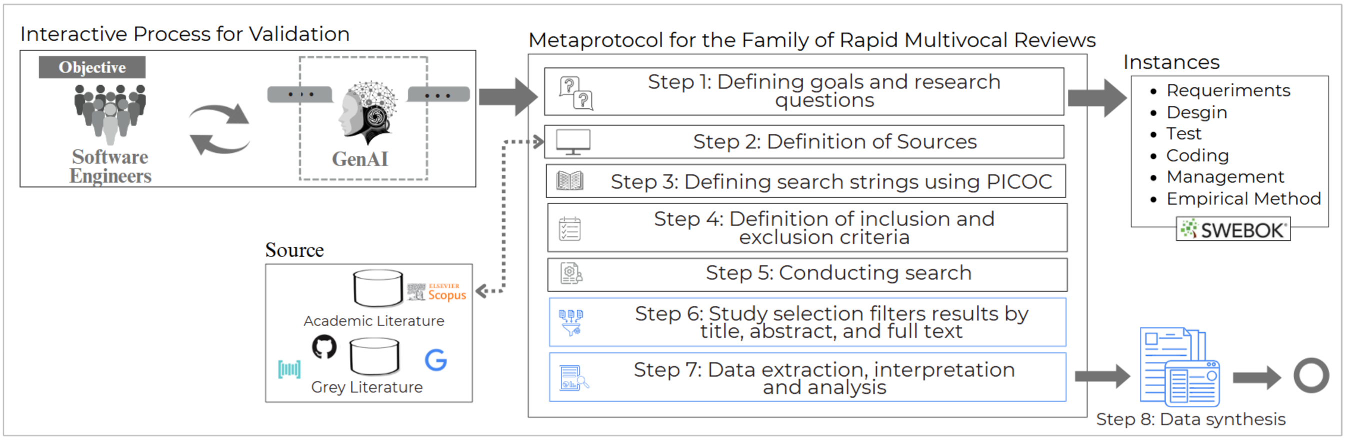

Figure 1: Interactive process for defining the metaprotocol.

3.1 Defining Goals and Research Questions

In this initial phase, GenAI was employed as a creative support tool through two complementary approaches: (i) the formulation of contextual prompts based on the scope and objectives of the study. These prompts included information such as the context, motivation, research problem, and review objectives. The aim was to guide GenAI in generating relevant suggestions aligned with the RMRs' focus. (ii) The simulation of GenAI as a researcher in Software Engineering, committed to developing an RMR. An example of this process can be accessed at https://figshare.com/s/3344a6c730aa181c0b9d.

This combination proved particularly valuable for broadening the repertoire of ideas and accelerating the exploratory phase of the study. GenAI introduced perspectives not previously considered by the software engineers, contributing to formulating research questions aligned with the central objective — identifying GenAI-based tools and understanding their use, maturity, accessibility, and adoption in the industry. The suggestions were analyzed, discussed collaboratively, and refined, with final validation being the responsibility of the researchers.

This experience corroborates the findings of Shaer et al. [16], which indicate that LLMs significantly contribute to the divergence phase of the ideation process by mitigating cognitive barriers and fostering creativity by introducing new investigative perspectives.

3.2 Definition of Sources

In the second stage of defining the metaprotocol, a search strategy was adopted combining indexed sources and grey literature, aiming to broaden the scope and capture academically validated solutions and emerging practices related to using GenAI in SE. Initially, the protocol included the use of the Scopus[2] database, which was recognized for its extensive coverage of scientific publications in the field of computing. During the metaprotocol definition process with software engineers, two additional sources were incorporated: Google[3], aimed at retrieving grey literature, such as technical blogs and non-indexed reports, and GitHub[4], a widely used platform for hosting open-source GenAI tools and agents. Furthermore, a key change in the collaborative definition process was the inclusion of the Papers With Code[5] platform, suggested by the GenAI tool. Previously unknown to some team members, this source was highly relevant for linking scientific publications with their practical implementations. It is widely adopted within the Artificial Intelligence and Machine Learning communities. Its inclusion was motivated by its direct alignment with the goals of the RMRs family: to identify GenAI solutions that are effectively operationalized in practice.

This protocol evolution underscores the importance of integrating grey literature into secondary studies, particularly in contexts marked by rapid technological transformation, such as GenAI. As discussed by Cartaxo et al. [3], overlooking sources such as code repositories, technical blogs, and market reports may undermine the timeliness and applicability of the findings. The combination of scientific and practical sources enables more robust data triangulation. It helps bridge the gap between academic research and professional practice, fostering outcomes that are more useful and aligned with the needs of the SE community.

3.3 Defining Search Strings

To guide the formulation of the search strings, we adopted the PICOC model (Population, Intervention, Comparison, Outcome, and Context), as proposed by Petticrew and Roberts [13]. This approach is widely used in systematic reviews because it ensures both comprehensiveness and conceptual alignment between the selected terms and the research objectives.

Starting from an initial protocol, we sought to refine the terms with the support of a Generative AI tool. Software engineers reviewed and discussed the suggestions generated, leading to a collectively validated and adjusted set of search strings. This process respected the particularities of each RMR instance and was guided by the thematic categories defined in SWEBOK v4.0. Table 1 presents an example of the search strings used in the Empirical Methods instance, using PICOC.

[2]Scopus: https://www-scopus-com
[3]Google: https://www.google.com/
[4]Github: https://github.com/
[5]Paper With Code: https://paperswithcode.com/

Table 1: Example of the final search strings

PICOC	Search Strings
Population	("experimental" OR "empirical" OR "evidence-based" OR "replication stud*" OR "qualitative" OR "quantitative")
Intervention	("Generative AI" OR "Generative Artificial Intelligence" OR "Generative Model*" OR "Large Language Model*" OR "Language Model*" OR "Small Language Model*" OR "LLM*" OR "RAG" OR "Retrieval Augmented Generation" OR "Natural Language Processing" OR "NLP" OR "AI Agent" OR "AI Multi-Agent")
Comparison	Not applicable.
Outcome	("Application" OR "Technolog*" OR "Approach*" OR "Method*" OR "Tool*" OR "Framework*" OR "Solution*" OR "Strateg*" OR "Model*" OR "System*" OR "Platform*" OR "Technique*")
Context	("Software Engineering")

Regarding the Intervention, we started from a consolidated set of terms representative of the GenAI field, including expressions such as "Generative AI," "LLM," and "RAG", among others. Interaction with the GenAI tool and iterative refinement cycles with software engineers led to enhancements that broadened the coverage and flexibility of the search string. Notably, wildcard operators were used to generalize terms (e.g., "Generative Model*" instead of "Generative Models" and "LLM*" instead of "LLM"), allowing for the capture of a broader range of terminological variations. Additionally, relevant emerging terms were incorporated, such as "Natural Language Processing", "NLP", "AI Agent", and "AI Multi-Agent", reflecting current trends in the application of intelligent agents and language models in SE. Since the aim of the RMRs family does not involve comparative analysis, the "Comparison" component was deemed not applicable. As for the Outcome, the initial version of the string included terms oriented toward the characterization of impacts and limitations of the solutions, such as "Limitation", "Benefit", "Challenge", "Quality", and "Productivity". However, to better align the search with the focus of this study — identifying operational solutions and applied technologies — we prioritized terms more directly associated with technical artifacts, such as "Tool", "Framework", "Solution", "Model", "System", and "Platform". This reformulation aimed to increase the precision of the search in identifying concrete solutions, implemented frameworks, and technologies effectively used in practice while also reducing ambiguity and the risk of overlap with other instances of the RMRs family, such as those focused on quality assessment. Lastly, the Context was maintained as "Software Engineering", as it aligns with the overarching scope of the research.

3.4 Definition of Selection Criteria

At this stage, the software engineers initially proposed the selection criteria based on the objectives of the RMRs family. GenAI was employed as a supportive tool during both the definition and revision of these criteria, contributing to the development of more precise, more objective descriptions aligned with the scope of the study. Although it did not autonomously propose new criteria, the tool assisted in identifying ambiguities and refining the textual formulation, making the criteria more precise and comprehensible. The criteria were established using December 2021 as a temporal milestone due to the significant advancements in the field following the release of the ChatGPT model—an event that marked a turning point in the adoption of GenAI solutions and the surge in publications within SE [19].

3.5 Conducting Search

In this fifth stage, we conducted the search execution process manually and iteratively, considering each source's specificities. In Scopus, we applied filters by subject area and a time restriction starting from December 2021, as established in the inclusion criteria.

For the grey literature, searches were performed on Google with automation support using a Python library[6], allowing us to retrieve URLs related to the formulated queries quickly. During this process, we encountered a practical limitation regarding the maximum length allowed for search strings by Google's engine. When using Boolean operators in combination with filters, we observed that overly long queries or those containing multiple keywords connected by `OR` exceeded the accepted limits for the search field, resulting in errors or truncated strings. To address this issue, we split the queries into smaller blocks and executed them separately, ensuring that all relevant terms were adequately covered. This fragmentation required additional post-processing curation to remove duplicates and consolidate the results.

During the search process in non-traditional sources such as GitHub and Papers With Code, we observed that these platforms do not support the direct execution of complex queries, such as those structured according to the PICOC model. As an alternative, we adopted a strategy of conducting advanced searches via Google, restricting the results to specific domains using the `site:` operator. This approach allowed us to circumvent the technical limitations of these sources and adapt the search execution to the realities of the RMRs without compromising methodological consistency or alignment with the study's objectives. It is also important to note that, for each query, we considered the top 100 returned results — rather than the first 100 pages — to capture the most relevant material as prioritized by Google's ranking algorithms. The searches were documented and refined interactively, with the support of software engineers, based on ongoing assessment and feedback regarding the relevance of the retrieved materials.

Throughout this stage, Generative AI was also employed as a complementary support tool to review the formulation of queries targeting grey literature. The AI suggested adjustments to Boolean operators and provided alternative terms to segment the strings into smaller blocks. This proved particularly useful in adapting the queries to the character length constraints of Google's search field. Although the execution was manual or semi-automated, the AI was a facilitator in designing more effective and compatible search strategies for the available mechanisms.

[6]Python library: googlesearch-python

3.6 Studies Selection

The study selection process is conducted manually based on predefined inclusion and exclusion criteria. As complementary support, we are exploring using GenAI to condense abstracts and facilitate the preliminary screening of potentially relevant studies. This approach is inspired by the work of Huotala et al. [6], who employed LLMs to simplify scientific abstracts and compared the screening decisions generated by the models with those made by human reviewers. Similarly, we are conducting a parallel analysis, comparing the suggestions provided by GenAI with the decisions of software engineers to assess the degree of alignment between them. We are also investigating using more advanced prompting strategies [15]—such as one-shot, few-shot, and especially chain-of-thought—that encourage the model to articulate its reasoning before making a judgment, thereby approximating the human decision-making process [7]. Structured prompts with explicit reasoning steps are being tested as a potential methodological enhancement to improve the consistency and transparency of GenAI contributions during the screening process.

3.7 Data Extraction, Interpretation, and Analysis

The data extraction and analysis phase is currently in progress. It was conducted in a structured manner, combining deductive coding strategies based on predefined categories with exploratory approaches for contextual interpretation guided by patterns and thematic clusters emerging from the reading of the studies. Data collection is based on a form that gathers metadata such as title, abstract, authors, year, and place of publication. In both approaches, Generative AI has served as a valuable methodological support.

Predefined categories guide deductive coding. We adopted a set of categories grounded in the SWEBOK v4.0 and aligned with our research questions.

The extraction is being carried out manually but with strategic support from GenAI, which suggests that relevant excerpts from the full texts should be associated with the predefined categories. One of the practices currently under evaluation involves using example-guided prompts—through one-shot and few-shot approaches—to identify information such as the type of activity supported by the GenAI solution, its level of maturity, and the application context (academic or industrial). Inspired by the approach proposed by Petersen and Gerken [12], we are also assessing how Retrieval-Augmented Generation (RAG)-based architectures can optimize the process by analyzing the whole PDF document, prioritizing the most relevant passages, and only then forwarding these segments to the OpenAI API. This technique aims to enhance the performance of GenAI in the extraction task by reducing noise and focusing on the most meaningful evidence. Nonetheless, the final validation and categorization remain under human researchers' responsibility, ensuring the extracted data's quality and accuracy.

Contextual Interpretation. Beyond extraction, GenAI is being experimentally employed to support the contextual interpretation of claims found in the studies. For instance, when an author states that their work includes empirical validation, we prompt the AI to verify the presence of methodological descriptions or concrete evidence that substantiates this claim. This approach is inspired by the work of Petersen [11], who demonstrated how GPT-4 can be used to validate classifications made by authors in primary studies. In our case, we aim to assess whether GenAI can assist in critical analysis by acting as a second lens—capable of enhancing the consistency and depth of researchers' interpretations.

3.8 Data Synthesis

The data synthesis stage represents a pivotal moment in the execution of the RMRs. At this point, we intend to employ a combination of deductive analysis and an exploratory approach grounded in thematic clustering to construct a comprehensive view of GenAI solutions applied to SE. Furthermore, the synthesis will compare perspectives drawn from scientific and grey literature, enabling a robust triangulation that deepens the understanding of real and emerging GenAI applications across industry and academia. It is worth noting that using LLMs to support data synthesis has been gaining traction in other domains, such as medicine [8], helping researchers navigate vast volumes of information and identify key themes more efficiently. This trend also points to a promising path for SE, suggesting that GenAI-based tools may become powerful allies in constructing systematic reviews.

4 Discussion

Recent studies have shown that models such as GPT can achieve — and, in specific contexts, even surpass — human performance in research tasks within SE [6, 19]. Throughout the development of our metaprotocol, it became evident that integrating GenAI tools into SE research is far more than a technical endeavor — it represents a paradigm shift.

At each stage in which GenAI was involved — from the definition of objectives to the selection of studies — we observed an acceleration of tasks and a more creative flow. This observation aligns with the evidence presented by Trinkenreich et al. [17], who highlights how GenAI tools can enhance creativity and alleviate repetitive tasks that often consume researchers' time and energy. In our experience, GenAI did not replace the researcher — instead, it served as a creative companion, suggesting directions, proposing terms, and sometimes challenging our way of thinking. However, its use demands caution. GenAI tools carry the biases and limitations inherent in the data on which they were trained, and with that come gaps, inconsistencies, and the risk of inaccurate responses. There is also the potential for the inadvertent use of incorrect information and, at a deeper level, the danger that overreliance may weaken researchers' critical thinking and analytical skills [17]. For this reason, we argue that using GenAI in scientific processes must always be subject to oversight. The academic community ensures transparency, validates sources, applies clear criteria at every stage, and maintains a critical stance throughout the process. GenAI can be a catalyst for accelerating discovery — but only if employed with awareness and responsibility.

Our account of the development of this metaprotocol also aims to raise awareness within the SE research community regarding both the potential and the risks associated with using GenAI. As these tools become increasingly integrated into the daily practices of research, it will be essential to establish clear guidelines, invest in ongoing training, and promote systematic evaluations that ensure quality, reproducibility, and scientific integrity. Moreover, this work draws attention to a significant gap: we still know little about using GenAI-based solutions in the software industry. The use of GenAI for coding has been widely validated in the literature [10]. However, although this application is relevant, coding accounts for only a small fraction of engineering activities throughout the software lifecycle. Therefore, this study investigates the use of GenAI in other stages of the engineering process, aiming to understand its potential in less explored tasks.

Finally, the development of this metaprotocol demonstrated that GenAI tools can serve as valuable allies in scientific research. However, context, motivation, problem, and objectives must be clearly defined and embedded within structures that ensure transparency, traceability, and methodological rigor. The future of SE research with GenAI will not be determined solely by the capabilities of the models but by the responsibility with which we employ them.

5 Related Work

The use of GenAI has significantly expanded in supporting scientific research, including in the field of SE, with studies highlighting its benefits, limitations, and methodological challenges. Andersen et al. [1] identified, through a survey of 2.534 researchers, a growing adoption of LLMs across various stages of the research process—particularly among early-career researchers—emphasizing the need for responsible guidelines and human oversight. Similarly, Petersen and Gerken [12] proposed an interactive and modular approach for conducting systematic mapping studies using LLMs, structuring the process with specialized agents, and adhering to the human-in-the-loop principle to ensure traceability and methodological quality. This approach aligns with the goals of this study by combining GenAI-generated suggestions with expert validation. In contrast, Trinkenreich et al. [17] adopts a critical stance, applying McLuhan's Tetrad to examine the impact of LLMs on the research pipeline, warning of risks such as diminished creativity, topic homogenization, and reduced analytical capacity, and reinforcing the need for ethical guidelines and training for their responsible use. In contrast to these perspectives, this study presents a practical application of GenAI in conducting a family of RMRs, offering a concrete response to recent literature by integrating responsible GenAI use with scientific rigor.

6 Threats to Validity

Despite our efforts to ensure rigor in designing the metaprotocol, we acknowledge threats to validity and categorize them according to Wohlin et al. [18].

Regarding the validity of the conclusion, the reliability of the suggestions generated by the GenAI was considered a threat to the definition of metaprotocol. To mitigate this risk, three expert with experience in secondary studies was involved to review and discuss the proposed content. Two main risks concerning internal validity were identified: (i) bias resulting from using GenAI and (ii) subjectivity in interpreting and validating the suggestions. The first was mitigated through validation sessions with software engineers who did not participate in the initial protocol design. The second was addressed using a structured protocol—including context, motivation, problem, and objectives—to guide the analysis and support consensus among reviewers. Finally, regarding external validity, two threats were considered: (i) the generalizability of the results, as this is the first study to propose a family of RMRs in the SE domain, which was mitigated by aligning each step with the goals of the RMRs; and (ii) technological dependency on GenAI tools, which was mitigated through detailed documentation of interactions, enabling replication with other tools in the future.

7 Conclusion

In a context where GenAI is rapidly evolving and capturing the interest of both industry and academia, we identified a gap: a lack of studies reveals how these solutions are being applied in practice. From this concern, the proposal of this work emerged — a family of RMRs to organize the body of knowledge about GenAI solutions applied in SE. The development of the protocol was both collaborative and interactive: we combined the analytical perspective of software engineers with the creative potential of GenAI. This collaboration enabled us to structure each stage of the RMRs while respecting the specificities of areas, such as requirements, design, testing, coding, management, and empirical methods, as defined in the SWEBOK V4. In the following steps, we plan to complete all instances of the RMRs family, explore advanced prompting techniques, and investigate how this protocol can be adapted to other contexts, including developing a tool that facilitates its adoption by other researchers. We also intend to compare the creative support of different technologies based on GenAI. Above all, this research is an invitation to rethink how we conduct secondary studies and responsibly integrate the use of GenAI into these processes.

ARTIFACT AVAILABILITY

The raw data and all the steps necessary to reproduce the study are detailed in the supplementary material located at https://figshare.com/s/3344a6c730aa181c0b9d.

ACKNOWLEDGMENTS

We sincerely thank all software engineers (Alexandre Esteves, André Gheventer, Arthur Sasse, Carlos Brito, Felipe Souza, João Pedro Wieland, Júlio Guimarães, Luísa Marschhausen, Patrícia Gonçalves, Pedro Gonçalves, Rafael Meirelles) who contributed to the evolution of the metaprotocol and are undertaking the RMRs. We acknowledge using ChatGPT tools to support the creative process, translation, and rewriting to improve text clarity. We reviewed all outputs to ensure alignment with our original intent. Prof. Travassos is a CNPq Researcher and FAPERJ CNE. CAPES, FAPERJ, and CNPq partially fund this work.

REFERENCES

[1] Jens Peter Andersen, Lise Degn, Rachel Fishberg, Ebbe K Graversen, Serge PJM Horbach, Evanthia Kalpazidou Schmidt, Jesper W Schneider, and Mads P Sørensen. 2025. Generative Artificial Intelligence (GenAI) in the research process–A survey of researchers' practices and perceptions. *Technology in Society* 81 (2025), 102813.
[2] Callen Anthony, Beth A Bechky, and Anne-Laure Fayard. 2023. "Collaborating" with AI: Taking a system view to explore the future of work. *Organization Science* 34, 5 (2023), 1672–1694.
[3] Bruno Cartaxo, Gustavo Pinto, and Sergio Soares. 2018. The role of rapid reviews in supporting decision-making in software engineering practice. In *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018*. 24–34.

[4] Christof Ebert and Panos Louridas. 2023. Generative AI for software practitioners. *IEEE Software* 40, 4 (2023), 30–38.

[5] Vahid Garousi, Michael Felderer, and Mika V Mäntylä. 2019. Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and software technology* 106 (2019), 101–121.

[6] Aleksi Huotala, Miikka Kuutila, Paul Ralph, and Mika Mäntylä. 2024. The promise and challenges of using LLMs to accelerate the screening process of systematic reviews. In *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering*. 262–271.

[7] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. *Advances in neural information processing systems* 35 (2022), 22199–22213.

[8] Xuân-Lan Lam Hoai and Thierry Simonart. 2023. Comparing meta-analyses with ChatGPT in the evaluation of the effectiveness and tolerance of systemic therapies in moderate-to-severe plaque psoriasis. *Journal of Clinical Medicine* 12, 16 (2023), 5410.

[9] William Leara. 2024. Guide to the SWEBOK v4.0 Has Been Released. https://www.basicinputoutput.com/2024/10/guide-to-swebok-v40-has-been-released.html [accessed on Accessed: May 16, 2025].

[10] MDB Modi. [n. d.]. Transforming Software Development Through Generative AI: A Systematic Analysis of Automated Development Practices. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* 10, 6 ([n. d.]).

[11] Kai Petersen. 2024. Case study identification with GPT-4 and implications for mapping studies. *Information and Software Technology* 171 (2024), 107452. doi:10.1016/j.infsof.2024.107452

[12] Kai Petersen and Jan M. Gerken. 2025. On the road to interactive LLM-based systematic mapping studies. *Information and Software Technology* 178 (2025), 107611. doi:10.1016/j.infsof.2024.107611

[13] Mark Petticrew and Helen Roberts. 2008. *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons.

[14] Patricia de Oliveira Santos, Allan Chamon Figueiredo, Pedro Nuno Moura, Bruna Diirr, Adriana CF Alvim, and Rodrigo Pereira Dos Santos. 2024. Impacts of the Usage of Generative Artificial Intelligence on Software Development Process. In *Proceedings of the 20th Brazilian Symposium on Information Systems*. 1–9.

[15] Yuya Sasaki, Hironori Washizaki, Jialong Li, Nobukazu Yoshioka, Naoyasu Ubayashi, and Yoshiaki Fukazawa. 2025. Landscape and Taxonomy of Prompt Engineering Patterns in Software Engineering. *IT Professional* 27 (2025), 41–49. doi:10.1109/MITP.2024.3525458

[16] Orit Shaer, Angelora Cooper, Osnat Mokryn, Andrew L Kun, and Hagit Ben Shoshan. 2024. AI-Augmented Brainwriting: Investigating the use of LLMs in group ideation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Article 1050, 17 pages. doi:10.1145/3613904.3642414

[17] Bianca Trinkenreich, Fabio Calefato, Geir Hanssen, Kelly Blincoe, Marcos Kalinowski, Mauro Pezzè, Paolo Tell, and Margaret-Anne Storey. 2025. Get on the Train or be Left on the Station: Using LLMs for Software Engineering Research [Manuscript under review]. (2025).

[18] Claes Wohlin, Per Runeson, Martin Höst, Magnus C Ohlsson, Björn Regnell, and Anders Wesslén. 2012. *Experimentation in software engineering*. Springer Science & Business Media.

[19] Quanjun Zhang, Chunrong Fang, Yang Xie, Yaxin Zhang, Yun Yang, Weisong Sun, Shengcheng Yu, and Zhenyu Chen. 2023. A survey on large language models for software engineering. *arXiv preprint arXiv:2312.15223* (2023).