Dear Editor and Reviewers,
We sincerely thank you and the reviewers for your thoughtful and constructive comments on our manuscript entitled "Towards the Next Generation of Software: Insights from Grey Literature on AI-Native Applications" (Manuscript ID: 143).
In response to your comments, we have thoroughly revised the manuscript. For convenience, the revised manuscript highlights significant changes in yellow. We hope this format facilitates your review, and we sincerely appreciate your time and constructive feedback.

Reviewer A

Reviewer Comment A.1 — The abstract section should more explicitly elaborate on the significance of the research topic (i.e., exploring native AI applications).

Reply: Thank you for this valuable suggestion. We agree that the abstract needed to better highlight why exploring AI-native applications is critical at this stage. We have revised the Context and Objective sections of the Abstract to explicitly highlight its significance.

- Revised Text: "Context: AI-native applications leverage generative AI as the core engine driving architecture. Despite substantial investment and rapid practical adoption, theoretical foundations remain underdeveloped, with interpretations varying across the industry and academic research still nascent. This gap creates inefficiencies and risks that hinder large-scale adoption.

  Objective: To facilitate the successful adoption of AI-native applications, ... "

Reviewer Comment A.2 — Why is gray literature review more suitable as a research method compared to other approaches (e.g., the Delphi method)?

Reply: This is an excellent question regarding methodological justification. We have added additional clarification in Section 1 (Introduction) to explicitly justify the choice of GLR.

- Revised Text: "Traditional Systematic Literature Review (SLR) [33, 34] and Systematic Mapping Study (SMS) [48, 49]are not adopted in this study. The reason is that SLR and SMS primarily rely on peer-reviewed white literature, while in the emerging domain of AI-native applications, academic research is still at an early stage and remains limited in scope and volume. Widely recognized experts and a stable conceptual framework in this domain have yet to emerge. Other methods, such as the Delphi technique [58], which rely on expert consensus, are less suitable in this context and may produce overly subjective and speculative results. Therefore, GLR represents a more appropriate and effective research method at the present stage, as it provides practitioners'contemporary perspectives on important topics in both practice and research, and presents empirical data derived from practitioners'real-world software development activities [22, 50]."

Reviewer Comment A.3 — In Chapter 2, the "Related Work" section, the discussion on the "industry perspective" is not comprehensive (currently only two definitions are listed), and similarly, the scope of the "academic perspective" is relatively narrow (currently only citing some papers focused on generative AI applications).

Reply: We accept this critique. The original draft provided only illustrative examples, which indeed made the landscape appear narrower than it is. To provide a stronger foundation for our study, we need to broaden the context. We have expanded Section 2 (Related Work).

- Industry perspectives: We have added references to industry development trends of AI-native applications and incorporated a broader range of industry perspectives on AI-native applications, such as those from Alibaba Cloud, frontline technical experts, and developer communities, beyond just Sapphire Ventures and Ericsson.
- Academic perspectives: We have significantly restructured this part to provide a more comprehensive evolutionary view. We have broadened the scope from underlying technologies, such as RLHF, RAG, and agent, to the AI-enabled software paradigm. We then contrast this with the emerging AI-Native paradigm, where AI serves as the core principle. We have incorporated the limited literature advocating for AI-native software engineering (SE 3.0) and the AI-native development lifecycle, thereby clarifying the academic context and highlighting the specific gap this study addresses.

Reviewer Comment A.4 — Chapter 2 might be merged with Section 5.2. After explaining the results of the research questions (RQs), the authors could discuss the similarities and differences between these findings and existing studies, and further explore their implications.

Reply: We appreciate this perspective on structuring the paper. While we understand the logic of merging them to foster a direct comparison, we have chosen to retain Section 2 (Related Work) as a separate section to adhere to the standard IMRAD (Introduction, Methods, Results, Discussion) structure, which allows us to clearly establish the research gap before introducing our methodology.

However, we fully agree with the reviewer's underlying point: the connection between our findings and the related work should be explicit in the discussion. Therefore, we have significantly strengthened Section 5.2.

- 5.2.1. Implications to practitioners: "...**Capability-oriented system design.** Traditional applications are typically designed and delivered around a set of predefined and relatively stable functions. In contrast, the 8 key characteristics of AI-native applications identified in this study, such as being AI-centric and continuously learning and adaptive, imply that system design should no longer aim at one-off functional delivery, but instead emphasize the ability to evolve continuously. Therefore, from the beginning of design, space needs to be reserved for ongoing prompt optimization and data-driven feedback loops. The system architecture should adopt a modular, observable, and rollback-enabled design to support rapid iteration and reliable evolution of models, prompts, and data throughout the entire lifecycle.

  Decouple models and prompts from business logic to enable independent replacement and flexible composition. To address the non-determinism of generative models, introduce observability quality assurance mechanisms [G38]..."
- 5.2.2. Implications to researchers: "...**AI-native software engineering.** Hassan et al. proposed AI-native software engineering (SE 3.0), characterized by an intent-centered, conversation-driven development process that emphasizes collaboration between human developers and AI teammates [26]. However, the industry has yet to reach a consensus on a unified definition of AI-native

applications, and academic research on this domain remains scarce. Significant research focuses on using generative AI to optimize software maintenance, such as code review [66], architecture detection [6], etc. There is still a gap in understanding the engineering of systems where the AI model is the core logic engine. Future research could clarify the methodological and practical differences among traditional software engineering (SE 1.0), AI-enabled software engineering (SE 2.0), and AI-native software engineering (SE 3.0) by conducting comparative analyses of their architectural patterns, lifecycle evolution, and failure modes..."

Reviewer Comment A.5 — In Table 1, the exclusion criterion E2, "E2 The study may take the form of a scientific paper, book, video, job posting, training announcement, etc." appears somewhat ambiguous. An inclusion criterion should be added to clearly define the scope of acceptable gray literature formats.

Reply: Thank you for this valuable suggestion. We have revised Table 1 (Selection Criteria) to be more precise.

- Refined exclusion criteria: "E2 The primary purpose of the source is commercial promotion, recruitment, or training."
- Added inclusion criterion: "I4 The source is not formally peer reviewed nor formally published, such as a blog, post, white paper, project documentation or preprint."

Reviewer B

Reviewer Comment B.1 — It would be helpful to provide a clear description of AI-Native Applications in the Introduction section, as this concept is central to the paper and serves as the foundation for the grey literature review (GLR).

Reply: Thank you for this valuable suggestion. We have revised the Introduction section.

- Revised Text: "It no longer views AI as an auxiliary feature but positions generative AI technologies as the intelligent engines that drive its core logic, user interaction, and overall system architecture, with representative examples such as ChatGPT and Midjourney."

Reviewer Comment B.2 — The authors may clarify why a GLR was chosen instead of a systematic mapping study or systematic literature review, to better justify the methodological choice and highlight the motivation behind this work.

Reply: Thank you for this valuable suggestion. We have added additional clarification in Section 1 (Introduction) to explicitly justify the choice of GLR and highlight the motivation behind this work.

- Revised Text: "Traditional Systematic Literature Review (SLR) [33, 34] and Systematic Mapping Study (SMS) [48, 49]are not adopted in this study. The reason is that SLR and SMS primarily rely on peer-reviewed white literature, while in the emerging domain of AI-native applications, academic research is still at an early stage and remains limited in scope and volume. Widely recognized experts and a stable conceptual framework in this domain have yet to emerge. Other methods, such as the Delphi technique [58], which rely on expert consensus, are less suitable in this context and may produce overly subjective and speculative results. Therefore, GLR represents a more appropriate and effective research method at the present stage, as it provides practitioners'contemporary perspectives on important topics in both practice and research, and presents empirical data derived from practitioners'real-world software development activities [22, 50]."

Reviewer Comment B.3 — The overall goal of the study can be articulated at the beginning of Section 3, along with an explanation of how the three RQs contribute to achieving this goal.

Reply: Thank you for this valuable suggestion. We have inserted a new paragraph at the beginning of Section 3. This paragraph clearly articulates the study's overall goal. In Section 3.2, we have added an explanation of how these research questions contribute to achieving this goal.

- Revised Text: "This study aims to overcome the fragmented understanding of AI-native applications by conducting a Grey Literature Review (GLR) [22] that synthesizes industry knowledge and practical experiences. It seeks to construct a comprehensive architectural perspective of AI-native applications, clarify their core characteristics, key quality attributes, and technical implementation elements, and provide a solid theoretical foundation for AI-native application practice and accelerating their engineering maturation process."

  "This study is structured around four research questions. The first three form the core content of a comprehensive architectural perspective on AI-native applications together. The fourth, in turn, examines the opportunities and challenges brought by AI-native applications from a macro perspective, highlighting the practical significance of the constructed comprehensive understanding."

Reviewer Comment B.4 — The implications of the results for practitioners and researchers can be further strengthened and discussed in more depth - ideally presented as highlighted bullet points to make the key takeaways more accessible.

Reply: Thank you for this valuable suggestion. We have significantly restructured Section 5.2. We have converted the narrative paragraphs into highlighted bullet points to emphasize key takeaways.

- 5.2.1. Implications to practitioners: "...**Capability-oriented system design.** Traditional applications are typically designed and delivered around a set of predefined and relatively stable functions. In contrast, the 8 key characteristics of AI-native applications identified in this study, such as being AI-centric and continuously learning and adaptive, imply that system design should no longer aim at one-off functional delivery, but instead emphasize the ability to evolve continuously. Therefore, from the beginning of design, space needs to be reserved for ongoing prompt optimization and data-driven feedback loops. The system architecture should adopt a modular, observable, and rollback-enabled design to support rapid iteration and reliable evolution of models, prompts, and data throughout the entire lifecycle.

Decouple models and prompts from business logic to enable independent replacement and flexible composition. To address the non-determinism of generative models, introduce observability quality assurance mechanisms [G38]..."

- 5.2.2. Implications to researchers: "...**AI-native software engineering.** Hassan et al. proposed AI-native software engineering (SE 3.0), characterized by an intent-centered, conversation-driven development process that emphasizes collaboration between human developers and AI teammates [26]. However, the industry has yet to reach a consensus on a unified definition of AI-native applications, and academic research on this domain remains scarce. Significant research focuses on using generative AI to optimize software maintenance, such as code review [66], architecture detection [6], etc. There is still a gap in understanding the engineering of systems where the AI model is the core logic engine. Future research could clarify the methodological and practical differences among traditional software engineering (SE 1.0), AI-enabled software engineering (SE 2.0), and AI-native software engineering (SE 3.0) by conducting comparative analyses of their architectural patterns, lifecycle evolution, and failure modes..."

Reviewer Comment B.5 — Consider organizing the core elements related to the definition of AI-Native Applications in Figure 3 in descending order. The same adjustment could improve clarity in several other figures throughout the paper.

Reply: Thank you for this valuable suggestion. We have regenerated Figure 3 and all other relevant bar charts (including Figure 4, 5, 6, 7, 8) to display data in descending order of frequency.

Reviewer Comment B.6 — The font size in several figures is too small and difficult to read. The authors might enlarge the font in these figures to improve readability.

Reply: Thank you for this valuable suggestion. We have increased the font size of axis labels, legends, and data values across all figures to ensure they are legible when printed or viewed on standard screens.

Reviewer C

Reviewer Comment C.1 — The study's motivation relies almost entirely on two industry reports (Refs. 4 and 6), which weakens the academic grounding of the work.

Reply: We thank the reviewer for this comment. We would like to clarify that Refs. 4 and 6 were only cited as illustrative examples of industry observations and do not constitute the sole basis for our study's motivation. To address this concern, we have strengthened the discussion of the study's motivation in Section 1.

- Revised Text: "...Meanwhile, some technology visionaries in the mobile application domain argue that "AI-native apps are built with architectures that fully incorporate machine learning and AI models as fundamental components." [56]. Industry perceptions of AI-native applications remain fragmented. Meanwhile, academic research on AI-native applications remains in early stages,

with relatively limited literature. Existing studies focus more on generative AI technologies themselves and on the AI-enabled paradigm, such as Reinforcement Learning from Human Feedback (RLHF) framework [69], efficient fine-tuning [39], automated code reviews [66], and architecture consistency checking [6], rather than on AI-native applications as a distinct paradigm. This gap contributes to piecemeal exploration, inefficient resource utilization, and substantial trial-and-error costs, introducing unpredictable quality risks and ultimately constituting a key bottleneck to large-scale adoption. For example, conceptual discussions in industry white papers and strategic analyses often frame the AI-native applications paradigm in abstract business or visionary terms; the actual development, predominantly visible in open-source projects, focuses on exploring concrete implementation pathways independently through practice. This misalignment between perception and practice has resulted in a lack of systematic guidance and difficulties in knowledge consolidation. To facilitate the successful adoption of AI-native applications, it is necessary to establish a systematic architectural-level understanding. Similar to the evolution of other architectural paradigms in software engineering, during the early stages of a paradigm, it is necessary to clarify its core characteristics, key quality attributes, and technical implementation elements at the architectural level, thereby laying a theoretical foundation for its maturation. For example, following the introduction of the microservices paradigm, researchers led by Flower [18] systematically analyzed the core characteristics, key quality attributes, and technical aspects of microservices architecture [37, 67, 71], effectively promoting its successful practice and widespread adoption."

Reviewer Comment C.2 — The discussion is narrow and largely interpretive, offering commentary on the findings without engaging with relevant academic literature.

Reply: Thank you for this valuable suggestion. In response, we have expanded the Discussion section by incorporating relevant academic literature, and reorganized the content into clearly highlighted key points to make the main conclusions more accessible and easier to understand.

- Revised Text: Due to space constraints, please refer to Section 5 of the manuscript.

Reviewer Comment C.3 — Although the authors claim to have conducted a thematic analysis, the reporting lacks authenticity. Specifically, direct statements or quotations from the 106 GL sources are missing under each RQ. Moreover, it is unclear which GL sources contributed to which RQ. A table mapping GL sources to RQs would substantially strengthen the credibility and traceability of the findings.

Reply: Thank you for this valuable suggestion. We have revised the description in the Section 4 by adding references to GL sources within the analysis for each research question. In addition, we have included tables mapping GL sources to the respective research questions (Table 4, 5, 6, 7, 8, 9), in order to enhance the credibility and traceability of the findings.

- Revised Text: Due to space constraints, please refer to Section 4 of the manuscript.

Reviewer Comment C.4 — Abstract/Introduction: Although defining "AI-native applications" is one of the study's objectives, the authors do not provide their own understanding or interpretation of this term. They could also highlight the diversity of interpretations that complicate the establishment of a unified definition.

Reply: Thank you for this valuable suggestion. We have added our understanding of AI-native applications separately in the Abstract and the Introduction sections.

- Abstract: "AI-native applications leverage generative AI as the core engine driving architecture."
- Introduction: "It no longer views AI as an auxiliary feature but positions generative AI technologies as the intelligent engines that drive its core logic, user interaction, and overall system architecture, with representative examples such as ChatGPT and Midjourney."

Reviewer Comment C.5 — Abstract: Referring to "106 GL sources" as "106 studies" is misleading. These are not academic studies but rather could be documents, blogs, videos, or other GL sources. The foundational GL review literature (e.g., Garousi et al.) avoids calling them "studies."

Reply: Thank you for this valuable suggestion. We have standardized the terminology throughout the manuscript, referring to all gray literature collectively as "GL sources."

- Revised Text: such as "106 GL sources"

Reviewer Comment C.6 — Introduction: The writing style is inconsistent across paragraphs (e.g., compare the second and fourth paragraphs and the bulleted list of contributions).

Reply: We thank the reviewer for pointing this out. We have carefully reviewed the Introduction and revised it to ensure a consistent writing style across all paragraphs, including the bulleted list of contributions.

- Revised Text: such as "This study reviews 106 GL sources and identify the defining characteristics and concepual boundaries of AI-native applicaions."

Reviewer Comment C.7 — Introduction: The distinction between "industry stakeholders" and "practitioners" (mentioned in paragraph 2) is unclear. Did practitioners author all 106 GL sources? If not, what differentiates the two categories?

Reply: We thank the reviewer for pointing this out. We initially used the terms "industry stakeholders" and " practitioners" to highlight the distinction between industry leaders and developers. However, we recognized that this caused some confusion. We have therefore unified the terminology to "practitioners".

All 106 GL sources are authored by practitioners.

Reviewer Comment C.8 — Related Work: Section 2.2 reads as a series of one-sentence summaries rather than a synthesized academic review that highlights conceptual gaps.

Reply: We thank the reviewer for pointing this out. We have significantly restructured this part to provide a more comprehensive evolutionary view. We have broadened the scope from underlying technologies, such as RLHF, RAG, and agent, to the AI-enabled software paradigm. We then contrast this with the emerging AI-Native paradigm, where AI serves as the core principle. We have incorporated the limited literature advocating for AI-native software engineering (SE 3.0) and the AI-native development lifecycle, thereby clarifying the academic context and highlighting the specific gap this study addresses.

- Revised Text: "Currently, research on the AI-native applications paradigm is still in its infancy, with existing studies primarily focusing on generative AI technologies themselves and their enabling application scenarios. Much of this body of research seeks to advance generative AI technologies, including enhancing domain-specific knowledge in LLMs through fine-tuning [39, 24], precisely controlling LLMs outputs via prompt engineering [40, 63, 53], aligning LLMs behavior through Reinforcement Learning from Human Feedback (RLHF) [69, 72], and improving performance on downstream tasks by leveraging retrieval-augmented generation (RAG) [51, 4] and LLMs-based agent [38, 57] approaches. Building on this foundation, researchers have further investigated the feasibility of AI-enabled paradigm. For example, generative AI techniques have been applied to various stages of the software development lifecycle, such as software architecture design [47, 10, 11], code generation [31, 32, 41], and software testing [61, 65, 25], in order to support developers and improve productivity. In addition, some studies have explored the quality attributes, technical challenges, key concerns and design principles of AI-enabled applications [54, 42, 15, 29, 23, 19, 62]. Such applications generally follow traditional software architecture design principles, treating AI as an integrated module to enhance product functionality"

Reviewer Comment C.9 — Research Methodology: The motivation for choosing a GL review is not sufficiently stated. While the method is appropriate, the rationale should be explicitly grounded in established guidelines, particularly those from V. Garousi, M. Felderer, and M.V. Mäntylä, "Guidelines for including grey literature and conducting multivocal literature reviews in software engineering," Information and Software Technology. The authors should justify their approach following that work.

Reply: Thank you for this valuable suggestion. We have added additional clarification in Section 1 (Introduction) to explicitly justify the choice of GLR and highlight the motivation behind this work.

- Revised Text: "Traditional Systematic Literature Review (SLR) [33, 34] and Systematic Mapping Study (SMS) [48, 49]are not adopted in this study. The reason is that SLR and SMS primarily rely on peer-reviewed white literature, while in the emerging domain of AI-native applications, academic research is still at an early stage and remains limited in scope and volume. Widely recognized experts and a stable conceptual framework in this domain have yet to emerge. Other methods, such as the Delphi technique [58], which rely on expert consensus, are less suitable in this context and may produce overly subjective and speculative results. Therefore, GLR represents a more appropriate and effective research method at the present stage, as it provides practitioners'contemporary perspectives on important topics in both practice and research, and presents empirical data derived from practitioners'real-world software development activities [22, 50]."

Reviewer Comment C.10 — Research Methodology: There is an inconsistent use of tense (past, present, and even future). Past tense should be used consistently when describing this section.

Reply: Thank you for this valuable suggestion. We have carefully revised Section 3 and ensured that past tense is used consistently when describing our methods.

- Revised Text: such as "This study aimed to overcome the fragmented understanding of AI-native applications by conducting a Grey Literature Review (GLR)..."

Reviewer Comment C.11 — Research Methodology: The rationale for selecting only the top 10 results from Google and Bing searches is not explained. Also, why were GitHub projects added? The process of merging GitHub projects with Bing and Google search results is unclear. Please refer to the corresponding figure.

Reply: Thank you for this valuable suggestion. We have added a clarification in Section 3.3 explaining the rationale for selecting the top 10 pages of Google and Bing search results. GitHub projects were included because they provide many real-world examples that reflect current practices and are considered a type of grey literature. The URLs of the GitHub projects were combined with the Google and Bing results to form the full dataset for the study.

- Revised Text: "We adopted an effort bounded criteria [21], i.e., including only the top N search engine hits, as the stopping criterion for this GLR search. Specifically, we collected the top 10 pages of results from Google and Bing with 10 URLs per page, following prior GLR in the software engineering domain [21, 5]."

  "As the world's largest open-source community, GitHub hosts a wealth of project documentation, issue reports, and practical experiences, and is also regarded as an important source of grey literature [68]."

  "To ensure consistency during merging, all entries were standardized by URL and title. Cross-platform duplicates were identified and removed using URL comparisons."

Reviewer Comment C.12 — Research Methodology: The criteria for including GitHub projects with more than 1000 stars and 50 forks need justification.

Reply: Thank you for this valuable suggestion. We have added a clarification in Section 3.3 explaining the rationale for the criteria for including GitHub projects with more than 1000 stars and 50 forks.

- Revised Text: "Considering the varying quality and impact of open-source projects, and referring to existing empirical GitHub-based project selection criteria [70, 27, 60], we filtered for projects with at least 1,000 stars, over 50 forks..."

Reviewer Comment C.13 — Research Methodology: The classification of GL sources into "title" and "full text" is confusing: are these referring to academic papers or GL items? Do we have full text or abstract in GL sources?

Reply: Thank you for this valuable suggestion. We realized that our wording was ambiguous. We have revised the related text in Section 3.3.

- Revised Text: "We subsequently screened these GL sources by examining their actual content in accordance with the inclusion and exclusion criteria..."

Reviewer Comment C.14 — Research Methodology: As mentioned earlier, in Table 1, please reconsider the terminology (e.g., "study" vs. "source") in accordance with Garousi et al. Also, note that item E2 is presented in a different color.

Reply: Thank you for this valuable suggestion. We have updated Table 1 to replace the term "study" with "source" in accordance with Garousi et al. Regarding item E2, in our version, we did not observe any color difference. Please let us know if there is a specific formatting issue we should address.

- Revised Text: such as "I1 The source primarily focuses on AI-native applications."

Reviewer Comment C.15 — Research Methodology: Section 3.5 discusses several quality criteria (given in Appendix A), but these are not reflected in the GitHub repository. While the authors' effort to share their repository is respected, I recommend further transparency.

Reply: Thank you for this valuable suggestion. We have updated the GitHub repository to include the quality assessment data and reorganized the files to improve readability and transparency. Detailed information can be found at https://github.com/llc202jy/GLR-on-AI-native-applications.

Reviewer Comment C.16 — Research Methodology: Section 3.7 provides only a brief overview of the thematic analysis but does not describe how it was actually conducted. No traceability is available to understand how themes were derived. Including a table of themes, subthemes, and example quotations from GL sources would substantially enhance comprehensibility.

Reply: Thank you for this valuable suggestion. We have added a detailed table of themes, subthemes, and representative examples from the GL sources in Section 3.7 (Table 3) to improve transparency and traceability of the thematic analysis process.

- Revised Text: "Table 3 presents the thematic analysis for AI-Native application definitions, including the identified themes, candidate themes, and example quotations."

Reviewer Comment C.17 — Research Methodology: The statement "All data were derived strictly from the original content of the papers, without any subjective interpretation" is unclear—what "papers" are being referred to? Upon examining the last two sources (GitHub links) in the provided repository-extraction sheet, I found no evidence of data extraction related to this article.

Reply: Thank you for this comment. We acknowledge that the original wording was imprecise. The term "papers" has been revised to "GL sources" to accurately reflect that the data were derived from the original content of the grey literature sources included in our study.

Regarding the availability of the extracted data, the previous version of the extraction sheet only contained the synthesized results. We have now added the raw extraction sheets for each stage of the methodology to the GitHub repository, following the steps described in the paper. These additions provide the detailed evidence of data extraction and improve the transparency and reproducibility of our study. Detailed information can be found at https://github.com/llc202jy/GLR-on-AI-native-applications.

- Revised Text: "All data were derived strictly from the original content of the selected GL sources, without any subjective interpretation."

Reviewer Comment C.18 — Results/Methodology: The process of identifying themes is not described. The analysis appears to rely primarily on frequency counts rather than a genuine thematic analysis.

Reply: Thank you for this comment. We have added the theme identification process and an example in Section 3.7 (Paragraph 1, Table 3).

In addition, the Results section has been reorganized to present and discuss the findings primarily around the identified themes rather than simple frequency counts.

- Revised Text: Due to space constraints, please refer to Section 4 of the manuscript.

Reviewer Comment C.19 — Results: It appears that 71 out of 106 GL sources originate from GitHub. This observation could be interesting to highlight and interpret in the results or discussion section.

Reply: Thank you for this valuable suggestion. We have added a discussion of this observation in Section 5, highlighting the large proportion of GitHub-based sources and interpreting it as evidence of the strongly practice-driven and implementation-oriented nature of current AI-native applications. We also emphasize this finding in the abstract to reflect its importance and its implications for understanding the characteristics and potential biases of the available grey literature.

- Abstract: "It is noteworthy that more than 60open-source project documentation, empirically highlighting that AI-native applications are largely practice-driven, while their underlying theoretical foundations remain underdeveloped."
- 5.1. Interpretation of Results: "In the results for RQ1, a notable pattern in data source distribution emerged. High-level, conceptual definitions of AI-native applications primarily originate from technical blogs and industry articles. In contrast, detailed descriptions of the key characteristics of AI-native applications are predominantly derived from open-source project documentation on GitHub (71 out of 106 items, accounting for over 60It reveals the dynamic process through which AI-native applications evolve from theoretical conception to practical implementation..."

Reviewer Comment C.20 — Discussion: Footnote 3 links to a simple Google search (what is "charlesshen.com"), which merely redirects to Google. Consider whether such a link is necessary; the same applies to Footnote 4.

Reply: Thank you for this valuable suggestion. The original intention of these footnotes was to provide readers with direct access to the two search engines for comparison. However, we agree that these links add little value in their current form. Therefore, we have removed Footnotes 3 and 4 in the revised version.

Reviewer Comment C.21 — Section 6: could be integrated into the Discussion section as a subsection (minor suggestion).

Reply: Thank you for this valuable suggestion. We have integrated the original Section 6 into the Discussion section as a subsection in the revised manuscript.

- Revised Text: "5.3. Threats to Validity Following the guidelines for validity analysis [55, 64], this section evaluates potential threats to the validity of the research findings..."

Reviewer Comment C.22 — Threats to Validity: Please cite Braun and Clarke (p. 24, line 4).

Reply: Thank you for this valuable suggestion. We have added the citation to Braun and Clarke in the Threats to Validity section as requested.

- Revised Text: "...we followed the structured, multi-stage methodology proposed by Braun and Clarke [9]."

Reviewer Comment C.23 — References: For Reference 4, this made me think whether the cited source (which is critical to motivate this work) has actually been read by the authors. Because it seems that the reference is taken from ChatGPT? Please Explain

Reply: Thank you for pointing this out. To quickly convert webpage content into BibTeX format, we used ChatGPT to generate the reference entry. However, this process was not thoroughly reviewed, and an inappropriate suffix appeared in the reference. We have since manually reviewed and corrected all references in the paper to ensure they are accurate, properly formatted, and directly based on the original sources.

Reviewer Comment C.24 — References: Reference 14 does not follow the same format as other online references with the URLs.

Reply: Thank you for pointing this out. We have corrected this and ensured that all references are now consistent and adhere to the journal's guidelines.

Reviewer Comment C.25 — References: Include DOIs where available.

Reply: Thank you for this valuable suggestion. We have added DOIs to all relevant references where available, ensuring the completeness and accuracy of the citations.