Investigating the Use of LLMs for Evidence Briefings Generation in Software Engineering

* Indicates required question

Consent Form

This form concerns your participation in the study *"Evaluating the Use of LLMs to Automatically Generate Evidence Briefings in Software Engineering."* Your participation is voluntary, and all responses will be anonymous and used solely for research purposes.

To ensure ethical and transparent data collection, this form follows the best practices outlined by Badampudi et al. (2017). You may withdraw at any time without penalty.

Please read the following before continuing.

1. **1. Title and Purpose of the Study** *

You are invited to participate in a research study titled *“Investigating the Use of LLMs for Evidence Briefings Generation in Software Engineering”*.
The goal of this study is to explore how Large Language Models (LLMs) can be used to automatically generate Evidence Briefings—short, practitioner-oriented summaries of research papers. As a software practitioner, you will be asked to evaluate whether the summaries are easy to understand and useful for real-world application.

2. What Will Participation Involve?

If you agree to participate, you will complete a short online questionnaire. You will be presented with two Evidence Briefings and asked to assess how clear and useful each one is for your daily practice. This activity is expected to take approximately 10 to 15 minutes.

3. Voluntary Participation

Your participation is completely voluntary. You may withdraw at any time without any consequences.

4. Risks and Benefits

There are no known or anticipated risks associated with this study. While you will not receive any direct benefits, your participation will contribute to ongoing research efforts aimed at improving knowledge transfer between academia and industry.

5. Anonymity and Confidentiality

We will not collect any personally identifiable information such as your name or email. All data will be analyzed in aggregate form. No individual responses will be reported.

6. Data Storage and Usage

Your responses will be stored securely and used exclusively for academic research purposes. Anonymized data may be included in publications or made available through open-access research repositories. No personally identifiable information will ever be disclosed.

7. Contact Information

If you have any questions about this study, please contact the researchers at [mauro.mmo93@gmail.com].

8. Consent Statement

By selecting "I agree" below, you confirm that:

Are at least 18 years old;

Have read and understood the information presented in this consent form;

Understand that your participation is voluntary and you may withdraw at any time;

Voluntarily agree to participate in this research study.

Mark only one oval.

- ◯ I agree
- ◯ I disagree

Characterization Form

This section collects background information to help us analyze responses according to your profile. It includes questions about your academic degree and professional experience. Your answers will remain anonymous and are essential for ensuring the validity of our analysis.

2. What is your highest academic degree? *

Mark only one oval.

- ◯ High School
- ◯ Bachelor's Degree
- ◯ Master's Degree
- ◯ Ph.D.

3. What is your current professional role? *(Select the one that best describes your position)* *

Mark only one oval.

- ◯ Business Analyst / Product Owner / Product Manager
- ◯ Scrum Master / Agile Coach
- ◯ Software Architect / Technical Lead
- ◯ Software Developer
- ◯ UI/UX Designer
- ◯ QA / Test Engineer
- ◯ Other: _____

4. How many years of professional experience do you have? *

Mark only one oval.

- ◯ 0-2
- ◯ 2-4
- ◯ 4-6
- ◯ 6-10
- ◯ More than 10

5. In which country are you based? *

_____

6. How often do you read software engineering research papers? *

Mark only one oval.

- ◯ Daily
- ◯ Weekly
- ◯ Montly
- ◯ I have read only a few papers, but it is not common
- ◯ I never read a software engineering research papers

7. Have you ever read a systematic review paper? *

Mark only one oval.

- ◯ Yes
- ◯ No

8. If you answered Yes for the last question, for what reason you read a systematic review paper?

Mark only one oval.

- ◯ Research purposes
- ◯ Decision-making on practice
- ◯ Policy-making
- ◯ Teaching
- ◯ Other: _____

Main Section

In the following section, you will be presented with two Evidence Briefings, followed by a few questions about usefulness and ease of understanding.

Understanding the Impact of Pair Programming on Software Development

INTRODUCTION:

This briefing summarizes the findings from a comprehensive meta-analysis conducted by Hannay et al. (2009) on the effectiveness of pair programming compared to solo programming. It aims to provide software engineering practitioners with insights into how pair programming affects quality, duration, and effort in software development tasks.

MAIN FINDINGS:

The meta-analysis of 18 studies revealed several key insights about pair programming:

1. **Quality:** Pair programming has a small but significant positive effect on the quality of code produced. Specifically, when comparing pair programming to solo programming, the evidence suggests that pairs tend to deliver higher-quality solutions, especially in complex tasks. However, the improvement in quality comes with an increase in the effort required.

2. **Duration:** Pair programming tends to be faster than solo programming for simpler tasks, with a moderate positive effect on duration. However, for complex tasks, while quality improves, the time taken to complete tasks can increase.

3. **Effort:** There is a medium negative effect on effort when using pair programming. This means that while pairs may produce better quality code, they often require significantly more effort to do so, especially for complex tasks.

4. **Task Complexity:** The effectiveness of pair programming is influenced by the complexity of the programming task. It is more beneficial for complex tasks in terms of quality but can lead to increased effort and potentially longer durations. Conversely, for simpler tasks, it can reduce completion time but may compromise quality.

5. **Expertise Levels:** The expertise of programmers also plays a crucial role in the effectiveness of pair programming. Junior programmers may achieve performance levels comparable to more experienced individuals when paired together, particularly on complex tasks.

6. **Publication Bias:** The analysis identified signs of publication bias, indicating that studies showing positive effects of pair programming are more likely to be published. This suggests that the true effectiveness of pair programming may be less favorable than reported.

WHO IS THIS BRIEFING FOR?

This briefing is intended for software engineering practitioners, project managers, and team leads who are considering implementing pair programming in their development processes. It provides a research-based overview to help inform decisions about whether and when to use pair programming.

WHAT IS INCLUDED IN THIS BRIEFING?

- A summary of the main findings from the original meta-analysis on pair programming.
- Insights into how task complexity and programmer expertise affect the outcomes of pair programming.

WHAT IS NOT INCLUDED IN THIS BRIEFING?

- Detailed statistical metrics and methodologies used in the original meta-analysis.
- In-depth descriptions of individual studies included in the analysis.

FOR ADDITIONAL INFORMATION ABOUT THIS BRIEFING:

Visit cin.ufpe.br/eseg/briefings.

REFERENCE:

Hannay, J. E., Dybå, T., Arisholm, E., & Sjøberg, D. I. K. (2009). The effectiveness of pair programming: A meta-analysis. *Information and Software Technology*, 51(11), 1110–1122. doi:10.1016/j.infsof.2009.02.001.

9. How familiar are you with the practice of pair programming?

Mark only one oval.

	1	2	3	4	5	
Not	◯	◯	◯	◯	◯	Very familiar

10. To what extent do you agree with the following statement: *

Mark only one oval per row.

	Strongly disagree	Disagree	Slightly disagree	I neither agree or disagree	Slightly agree	Agree	Strongly agree
The language used in the evidence briefing was clear and accessible.	◯	◯	◯	◯	◯	◯	◯

11. Please justify your choice

12. To what extent do you agree with the following statement: *

Mark only one oval per row.

	Strongly disagree	Disagree	Slightly disagree	I neither agree or disagree	Slightly agree	Agree	Strongly agree
The information in the evidence briefing was organized in a logical and coherent way.							

13. Please justify your choice

14. To what extent do you agree with the following statement: *

Mark only one oval per row.

	Strongly disagree	Disagree	Slightly disagree	I neither agree or disagree	Slightly agree	Agree	Strongly agree
The briefing presented the main ideas clearly and with the right amount of detail.							

15. Please justify your choice

16. To what extent do you agree with the following statement: *

Mark only one oval per row.

	Strongly disagree	Disagree	Slightly disagree	I neither agree or disagree	Slightly agree	Agree	Strongly agree
The content of the evidence briefing is relevant to challenges or tasks I encounter in my professional practice.							

17. Please justify your choice

18. To what extent do you agree with the following statement: *

Mark only one oval per row.

	Strongly disagree	Disagree	Slightly disagree	I neither agree or disagree	Slightly agree	Agree	Strongly agree
I can apply the main insights or recommendations of this briefing in my daily work.	⬭	⬭	⬭	⬭	⬭	⬭	⬭

19. Please justify your choice

_____

_____

_____

_____

_____

The use of Definition of Done in agile projects

Introduction:

This briefing reports scientific evidence on the Definition of Done criteria used in agile projects based on scientific evidence from a systematic review.

Main Findings:

The main findings of this paper considered that there are different types of Definition of Done (DoD) criteria, such as:

- Activity (e.g., peer code review);
- Metrics (e.g., localization defect density);
- Targets (e.g., product committed to CVS);
- Standards (e.g., coding standards);
- Checklist (e.g., design review checklist).

The Done Criteria is categorized based on each activity given its end goal. The defined criteria was software verification and validation, deploy, regulatory compliance, code inspection, test process quality, software architecture design, process management, configuration management, and Non-functional requirements check. After the process a total of 8 studies were selected. It shows that 3 studies applied multilevel DoD, meaning that to be done, a backlog item must be checked during more than one development step, such as story, sprint, release, etc.

Regarding the context of the reported projects, most of the studies (7 papers) used Scrum as the agile method, and were performed in the industry (6 papers).

The studies indicate they were done as experience papers and solution proposal since most of them were used in industry. Based on quality assessment performed, we concluded that the quality of the studies from an evidence-based perspective is low.

Nevertheless, some studies show the use of DoD as a means of complementing the agile process and comply with external requirements such as for ISO 9001, CMMI audits, as well as the addition of activities for assuring the product quality.

Only unit testing criteria was present in 5 studies, while peer code review and acceptance test were present in 4 studies, and system test, integration test and static code review were found in 3 studies. This disagreement is expected, because the DoD is defined by team according to its context.

Studies indicate that the use of DoD brings some benefits to the projects. However, those benefits are not clear, since they differ among studies. The main benefits described are:

- Improved collaboration between teams;
- Increase productivity;
- Reduce technical debt;
- Reduce defects deferred and defects reopened.

Who is this briefing for?

Software engineering practitioners who want to make decisions about definition of done in agile projects based on scientific evidence.

Where the findings come from?

All findings of this briefing were extracted from the systematic review conducted by Silva et al.

What is included in this briefing?

The main findings of the original systematic review. Evidence characteristics through a brief description about the original systematic review and the studies it analized.

What is not included in this briefing?

Details about the process performed to achieve results presented in this briefing report.

20. How familiar are you with the concept of definition of done?

Mark only one oval.

	1	2	3	4	5	
Not	◯	◯	◯	◯	◯	Very familiar

21. To what extent do you agree with the following statement: *

Mark only one oval per row.

	Strongly disagree	Disagree	Slightly disagree	I neither agree or disagree	Slightly agree	Agree	Strongly agree
The language used in the evidence briefing was clear and accessible.	◯	◯	◯	◯	◯	◯	◯

22. Please justify your choice

23. To what extent do you agree with the following statement: *

Mark only one oval per row.

	Strongly disagree	Disagree	Slightly disagree	I neither agree or disagree	Slightly agree	Agree	Strongly agree
The information in the evidence briefing was organized in a logical and coherent way.							

24. Please justify your choice

25. To what extent do you agree with the following statement: *

Mark only one oval per row.

	Strongly disagree	Disagree	Slightly disagree	I neither agree or disagree	Slightly agree	Agree	Strongly agree
The briefing presented the main ideas clearly and with the right amount of detail.							

26. Please justify your choice

27. To what extent do you agree with the following statement: *

Mark only one oval per row.

	Strongly disagree	Disagree	Slightly disagree	I neither agree or disagree	Slightly agree	Agree	Strongly agree
The content of the evidence briefing is relevant to challenges or tasks I encounter in my professional practice.							

28. Please justify your choice

29. To what extent do you agree with the following statement: *

Mark only one oval per row.

	Strongly disagree	Disagree	Slightly disagree	I neither agree or disagree	Slightly agree	Agree	Strongly agree
I can apply the main insights or recommendations of this briefing in my daily work.	⬭	⬭	⬭	⬭	⬭	⬭	⬭

30. Please justify your choice