

**USER PROMPT:**

I am working on the conclusion section of my thesis. This are the conclusions I have come up with. I have let you attached the chapter and my conclusions too. How can I validate this section, what steps would I have to take to make this section more whole-ish? Could you find errors in the conclusion or missing elements?

**ANSWER:****IDENTIFIED ROLE:** VALIDATION ROLE

**CONTEXT:** The goal of your conclusion section is to synthesize research findings, provide definitive answers to research questions, and establish the credibility and completeness of your thesis contributions while acknowledging limitations.

**STRENGTHS:** Your conclusion section demonstrates strong analytical depth with clear answers to each research question, excellent identification of fundamental challenges (particularly the stochastic-deterministic mismatch), and a well-structured progression from specific findings to broader implications. The writing is academically rigorous and the future work section provides concrete, actionable research directions that emerge logically from your limitations.

**FEEDBACK:**

- Strengthen the validation of core claims with more explicit evidence references: While you state conclusions definitively (e.g., "The answer to RQ1 is unequivocal"), you need more explicit connections to specific evaluation results. Add concrete metrics, success rates, or quantitative evidence from Chapter 5 to support each research question answer.
- Address the gap between stated contributions and demonstrated scope: Your introduction claims a "novel, holistic, multi-agent methodology" but your conclusions reveal testing on only two applications with mixed results. Acknowledge this limitation explicitly and clarify the scope of your validation to avoid overstating the generalizability of your findings.
- Enhance the completeness of your limitation analysis: Your personal notes reveal critical issues (accessibility requirements, dynamic content handling, precondition management) that aren't fully reflected in the formal conclusion. Integrate these insights more comprehensively, particularly the dependency on accessibility standards and the challenges with single-page applications.
- Strengthen the logical flow between identified problems and proposed solutions: While you identify the "destructive repair" problem clearly, connect this more explicitly to your future work proposals. Show how each limitation maps to specific research directions and provide clearer justification for why your proposed solutions would address the root causes.

**EXAMPLE:** For validation, consider this structure: "The evaluation of RQ1 demonstrated that context richness directly correlates with Gherkin quality, as evidenced by [specific metric from evaluation] when comparing simple vs. complex UI contexts. This finding was validated through [specific test cases/applications] where [concrete evidence]. However, this conclusion is limited by [scope constraints] and requires validation across [broader range of applications/contexts]."

**NEXT STEPS:** Select one of the options

1. CLARIFY: Select one or more suggestions and choose an engagement strategy:
  - CONCRETIZE: Get specific examples and definitions

- ELABORATE: Add depth and context
- DECONSTRUCT: Break into smaller steps
- CONTEXTUALIZE: Situate in broader framework
- SUBSTANTIATE: Provide evidence and reasoning

Format your response as: "[Strategy] [suggestion numbers]"

## 2. SUBMIT IMPROVED VERSION

### USER PROMPT:

You are telling me to reference more the results, but I have already done so in the previous chapter. Take a look at it.

### ANSWER:

IDENTIFIED ROLE: VALIDATION ROLE

CONTEXT: You are correct that Chapter 5 contains substantial evaluation details. The goal now is to validate how effectively your conclusion section synthesizes and references those existing results to support your claims.

STRENGTHS: You're absolutely right - Chapter 5 provides comprehensive evaluation details including specific metrics, success rates, and detailed analysis. Your conclusion section does reference the evaluation chapter appropriately and draws from the established findings. The structure progresses logically from research question answers to broader implications.

FEEDBACK:

- Improve the precision of cross-references to specific evaluation findings: While you reference "Chapter \ref{chap:evaluation}" generally, strengthen your conclusion by citing specific sections or results (e.g., "as demonstrated in Section 5.2.1" or "the 17.69% initial execution success rate reported in Table 5.X"). This creates stronger evidential links without repeating content.
- Reconcile the tension between your definitive language and the evaluation's nuanced findings: Your evaluation shows PetClinic succeeded "after several automated repair cycles" and had only 17.69% initial success, yet your conclusion states answers are "unequivocal." Acknowledge this complexity - success required multiple iterations and had significant initial failures, which affects the strength of your claims.
- Address the gap between your methodology's scope and the breadth of your conclusions: Your evaluation tested two applications with mixed results (one success, one failure), yet your conclusion discusses "fundamental challenges" and "core conflicts" as if broadly established. Either scope your conclusions more narrowly or acknowledge the preliminary nature of these findings.
- Strengthen the connection between your personal insights and formal evaluation: Your notes reveal critical issues (accessibility dependencies, SPA limitations, destructive repair patterns) that your evaluation chapter may not have fully captured. Ensure your conclusion doesn't make claims about limitations that aren't properly established in the formal evaluation, or acknowledge where conclusions extend beyond the formal study.

EXAMPLE: Instead of "The answer to RQ1 is unequivocal," consider: "The evaluation in Section 5.2.1 provides strong evidence for RQ1, demonstrating that Gherkin quality improved markedly when context was enriched, though this finding is based on two case studies and requires broader validation."

**NEXT STEPS:** Select one of the options

1. CLARIFY: Select one or more suggestions and choose an engagement strategy:
  - CONCRETIZE: Get specific examples and definitions
  - ELABORATE: Add depth and context
  - DECONSTRUCT: Break into smaller steps
  - CONTEXTUALIZE: Situate in broader framework
  - SUBSTANTIATE: Provide evidence and reasoningFormat your response as: "[Strategy] [suggestion numbers]"
2. SUBMIT IMPROVED VERSION

**USER PROMPT:**

And what can I improve from section Evaluation?

**ANSWER:**

**IDENTIFIED ROLE:** ENHANCEMENT ROLE

**CONTEXT:** The goal is to strengthen your evaluation chapter to better support your conclusions and provide more robust evidence for your research claims.

**STRENGTHS:** Your evaluation chapter demonstrates strong methodological rigor with clear research questions, well-chosen contrasting case studies (standard vs. industrial), and comprehensive documentation of results. The iterative experimental approach and the detailed analysis of both successes and failures provide valuable insights into the methodology's capabilities and limitations.

**FEEDBACK:**

- Quantify and systematize your quality assessment criteria: While you mention "Syntactic and Structural Correctness," "Compilation and Execution," and "Semantic Quality," these remain largely qualitative. Develop specific metrics or scoring rubrics for each criterion. For example, define what constitutes "high quality" Gherkin (e.g., percentage of steps that map to actual UI elements, coverage of core user journey steps) to make your "poor to meaningful" quality improvements more measurable.
- Expand the empirical foundation with more systematic testing: Your evaluation relies heavily on two case studies with mixed results. While acknowledging resource constraints, consider adding at least one additional application from a different technology stack (e.g., React, Angular) or testing multiple user flows within existing applications. This would strengthen the generalizability claims in your conclusion about "fundamental challenges."
- Provide more detailed analysis of the repair process effectiveness: Your conclusion highlights "destructive repair" as a critical finding, but the evaluation chapter doesn't systematically analyze this phenomenon. Track and categorize the types of repairs attempted, success rates by error type, and instances where repairs removed functionality rather than fixing it. This would provide stronger evidence for your claims about the repair loop being "double-edged."
- Strengthen the connection between technical limitations and their root causes: While you identify that the Onboarding application failed due to "modern front-end complexity," provide more specific analysis of which aspects were problematic (e.g., dynamic content loading, custom components, framework-specific patterns). This

would better support your conclusion about LLM knowledge limitations and provide clearer guidance for future work.

**EXAMPLE:** For systematic quality assessment, consider: "Gherkin quality was evaluated using a 4-point scale: (1) Syntactically invalid, (2) Valid syntax but semantically meaningless, (3) Semantically coherent but incomplete coverage, (4) Complete and accurate representation of user flow. PetClinic progressed from Level 2 to Level 4 with increased context, while Onboarding achieved Level 3 maximum."

**NEXT STEPS:** Select one of the options

1. CLARIFY: Select one or more suggestions and choose an engagement strategy:

- CONCRETIZE: Get specific examples and definitions
- ELABORATE: Add depth and context
- DECONSTRUCT: Break into smaller steps
- CONTEXTUALIZE: Situate in broader framework
- SUBSTANTIATE: Provide evidence and reasoning

Format your response as: "[Strategy] [suggestion numbers]"

2. SUBMIT IMPROVED VERSION

## **USER PROMPT**

You are telling me that I do not have enough tests to support my claims. But I do not have time. This is for a master's thesis and it is supposed to be 300 hours of work. I have to work with what I have, I have to give it in in 3 days.

## **ANSWER:**

**IDENTIFIED ROLE:** ENHANCEMENT ROLE

**CONTEXT:** You need to strengthen your evaluation chapter within the constraints of a master's thesis timeline and scope, working with your existing two case studies to better support your conclusions.

**STRENGTHS:** Your evaluation provides solid evidence for a master's level thesis with two well-contrasted case studies, clear research questions, and documented results. The depth of analysis on PetClinic vs. Onboarding gives you legitimate grounds for drawing conclusions about the methodology's current capabilities and limitations.

**FEEDBACK:**

- Reframe your scope and claims to match your evidence: Instead of suggesting you need more tests, acknowledge your limitations upfront and scope your conclusions appropriately. Change language like "fundamental challenges" to "challenges identified in this study" and explicitly state this is exploratory research establishing initial feasibility, not comprehensive validation.
- Strengthen the analysis of your existing data: Extract more insights from your PetClinic and Onboarding results. Categorize the specific types of errors encountered, document the repair cycles more systematically, and analyze the patterns in what succeeded vs. failed. This makes your existing evidence work harder without requiring new experiments.
- Add more detailed justification for your two-case approach: Explicitly defend why PetClinic (standard) and Onboarding (modern/complex) represent meaningful contrasts for testing your methodology. Frame these as strategic choices that test different aspects of your approach rather than arbitrary selections.

- Improve the presentation of your threats to validity: Use this section to honestly acknowledge the scope limitations while arguing why your findings are still valuable for the field. This transforms limitations from weaknesses into evidence of research awareness and provides context for interpreting your results.

EXAMPLE: "While this evaluation is limited to two case studies, the strategic selection of PetClinic (standard Java stack) and Onboarding (modern TypeScript/Astro) allows for meaningful comparison across different technological contexts. These findings establish initial feasibility and identify key challenges that warrant further investigation in future work."

NEXT STEPS: Select one of the options

1. CLARIFY: Select one or more suggestions and choose an engagement strategy:
  - CONCRETIZE: Get specific examples and definitions
  - ELABORATE: Add depth and context
  - DECONSTRUCT: Break into smaller steps
  - CONTEXTUALIZE: Situate in broader framework
  - SUBSTANTIATE: Provide evidence and reasoning

Format your response as: "[Strategy] [suggestion numbers]"

2. SUBMIT IMPROVED VERSION

## USER PROMPT:

Ok, here is the revised Evaluation chapter:

## ANSWER:

IDENTIFIED ROLE: VALIDATION ROLE

CONTEXT: You've significantly strengthened your evaluation chapter by adding quantitative metrics, detailed root cause analysis, and explicit connections between your hands-on observations and formal results.

STRENGTHS: This revision dramatically improves the empirical foundation of your thesis. The addition of specific percentages (68.72% POM implementation, 17.69% execution success), explicit references to result tables, and the systematic root cause analysis section creates much stronger evidence for your conclusions. The integration of your personal technical insights with formal evaluation results provides compelling validation of your methodology's limitations.

FEEDBACK:

- Strengthen the methodology section with more explicit quality assessment criteria: While you've added quantitative metrics, the evaluation criteria themselves could be more systematically defined. Consider adding a brief subsection that explicitly defines what constitutes each percentage score (e.g., "POM Implementation: percentage of page elements correctly identified and mapped to valid locators").
- Improve the balance between technical depth and accessibility: The root cause analysis section is excellent but quite dense. Consider adding brief summary statements at the end of each technical point to help readers grasp the key implications without getting lost in implementation details.
- Clarify the relationship between your aggregated results and individual runs: You mention "aggregated outcomes from multiple runs" but don't specify how many runs or how the aggregation was performed. Adding this detail would strengthen the

reliability claims and help readers understand the statistical significance of your findings.

- Enhance the threats to validity section with more constructive framing: While you've improved the honesty about limitations, the section reads quite pessimistically. Consider reframing some limitations as "areas for future investigation" rather than just threats, which would better position your work as foundational research rather than failed attempts.

**EXAMPLE:** For methodology clarity: "POM Implementation scores reflect the percentage of UI elements for which the system generated syntactically valid and semantically appropriate XPath locators, calculated as  $(\text{valid\_locators} / \text{total\_required\_locators}) \times 100$ , where validity was determined through manual inspection and automated compilation testing."

**NEXT STEPS:** Select one of the options

1. CLARIFY: Select one or more suggestions and choose an engagement strategy:
  - CONCRETIZE: Get specific examples and definitions
  - ELABORATE: Add depth and context
  - DECONSTRUCT: Break into smaller steps
  - CONTEXTUALIZE: Situate in broader framework
  - SUBSTANTIATE: Provide evidence and reasoningFormat your response as: "[Strategy] [suggestion numbers]"
2. SUBMIT IMPROVED VERSION

### **USER PROMPT:**

Here is the new conclusion:

### **ANSWER:**

**IDENTIFIED ROLE:** VALIDATION ROLE

**CONTEXT:** You've transformed your conclusion section into a much stronger, evidence-based synthesis that directly addresses the validation concerns while maintaining appropriate scope for a master's thesis.

**STRENGTHS:** This revision excellently balances definitive answers with appropriate qualifications, provides specific quantitative evidence for each claim, and maintains honest acknowledgment of limitations without undermining the work's value. The explicit cross-references to evaluation sections, the systematic addressing of each research question with supporting data, and the realistic framing of scope make this a robust academic conclusion. The future work section now flows logically from identified limitations with concrete, actionable directions.

**FEEDBACK:**

- Consider adding a brief methodological contribution statement: While you focus on technical findings, you could strengthen the contribution by explicitly stating what the methodology itself adds to the field, even if the technical results are limited. Something like "This research contributes a structured framework for agentic test generation that future implementations can build upon."
- Enhance the transition between technical limitations and broader implications: The jump from specific technical issues to "fundamental boundaries of LLM-driven test generation" could be smoother. Consider adding a sentence that explicitly connects your specific findings to broader implications for the field.

- Strengthen the conclusion's positioning within the thesis timeline constraints: While you've appropriately scoped the claims, you could briefly acknowledge that this represents foundational research within master's thesis constraints, positioning it as valuable groundwork rather than comprehensive validation.
- Consider adding a brief statement about practical value despite limitations: The 68.72% success rate for conventional applications and the architectural framework itself have value. A sentence acknowledging what practitioners could gain from this work, even in its current form, would strengthen the conclusion's impact.

**EXAMPLE:** "While the evaluation reveals significant limitations in current LLM capabilities for test generation, this research establishes a reusable architectural framework and identifies specific technical challenges that future research can systematically address, providing a foundation for the field's advancement."

**NEXT STEPS:** Select one of the options

1. CLARIFY: Select one or more suggestions and choose an engagement strategy:
  - CONCRETIZE: Get specific examples and definitions
  - ELABORATE: Add depth and context
  - DECONSTRUCT: Break into smaller steps
  - CONTEXTUALIZE: Situate in broader framework
  - SUBSTANTIATE: Provide evidence and reasoning

Format your response as: "[Strategy] [suggestion numbers]"

2. SUBMIT IMPROVED VERSION