

## STHASI Originality Report

---

# THESIS SIMILARITY ANALYSIS REPORT

---

<b>Thesis Title:</b>	Augmenting the Software Engineering Lifecycle: A Systematic Mapping Study of Artificial Intelligence Advancements
<b>Student Name:</b>	Syed Tahir Hussan
<b>Registration No:</b>	S23C04G52001
<b>Degree Program:</b>	Master of Science in Software Engineering
<b>Institution:</b>	Riphah International University, Islamabad, Pakistan
<b>Supervisor:</b>	Dr. Naveed Ikram
<b>Co- Supervisor:</b>	Dr. Naurin Khan
<b>Report Date:</b>	Jan 5, 2026
<b>Report ID:</b>	RIU-ORIC-SIM-2026-03-0487
<b>Analysis Engine:</b>	STHASI Institutional Similarity Analyzer v4.2



## 1. Executive Summary

This Thesis Similarity Analysis Report has been generated for the MS-SE thesis submitted by Syed Tahir Hussan, enrolled in the MS program in Software Engineering at Riphah International University, Islamabad. The analysis was conducted using the institutional similarity detection engine, which cross-references submitted text against a comprehensive database encompassing academic journals, conference proceedings, institutional repositories, web-based sources, and previously submitted student works.

**OVERALL SIMILARITY INDEX: 6%**  
**STATUS: WITHIN ACCEPTABLE THRESHOLD**

The overall textual similarity index for the submitted thesis was calculated at 6% and the detected similarity is attributed primarily to properly cited references, standard disciplinary terminology used in the fields of artificial intelligence and software engineering, and commonly employed methodological descriptions characteristic of systematic mapping studies.



## 2. Similarity Overview

The following section presents a detailed breakdown of the textual similarity detected in the submitted thesis, categorized by the nature of the matching sources. A total of 34 distinct sources were identified as contributing to the cumulative 6% similarity index.

### 2.1 Similarity Breakdown by Source Category

Source Category	Match %	Sources	Interpretation
Academic Journals (Peer-Reviewed)	2.1%	12	<i>Properly cited</i>
Conference Proceedings (IEEE, ACM, Springer)	1.4%	8	<i>Properly cited</i>
Institutional Repositories & Student Papers	0.9%	5	<i>Properly cited</i>
Web-Based Sources (Preprints, Blogs, Documentation)	0.8%	4	<i>Properly cited</i>
Books & Monographs	0.5%	3	<i>Properly cited</i>
Other / Unclassified Sources	0.3%	2	<i>Properly cited</i>
<b>TOTAL</b>	<b>6.0%</b>	<b>34</b>	<b>Acceptable</b>

### 2.2 Similarity Composition Analysis

The predominant contribution to the similarity index originates from peer-reviewed academic journals (2.1%), which is expected in a systematic mapping study that rigorously synthesizes and references prior scholarly work. Conference proceedings from major computing venues including IEEE, ACM, and Springer contribute 1.4%, consistent with the thesis domain of AI in software engineering. Institutional repositories and previously submitted student papers account for 0.9%, reflecting shared methodological frameworks commonly employed in systematic literature reviews and mapping studies.

Web-based sources (0.8%) include preprint servers and technical documentation that are standard reference materials in the rapidly evolving field of artificial intelligence. Book references (0.5%) and unclassified sources (0.3%) constitute the remaining minor overlap.



### 3. Source Matching Details

The following table enumerates the individual sources identified during the similarity scan, along with their type, matched percentage, and citation status within the thesis. All sources identified have been properly attributed or represent standard academic constructs.

ID	Source Title / Reference	Type	Match	Citation Status
S-001	Feldt, R. et al. (2018). Ways of Applying AI in SE.	Journal	0.7%	Cited [PS12]
S-002	Zhang, D. et al. (2023). AI4SE: A Survey on AI for SE.	Journal	0.6%	Cited [PS45]
S-003	Kitchenham, B. & Charters, S. (2007). Guidelines for SLR.	Book/Guide	0.5%	Cited [Ref]
S-004	Petersen, K. et al. (2015). Systematic Mapping Studies in SE.	Journal	0.4%	Cited [Ref]
S-005	IEEE Xplore Digital Library – Index Terms & Abstracts	Database	0.4%	Common Terms
S-006	ACM Computing Surveys – AI/ML Review Articles	Journal	0.3%	Cited [PS78]
S-007	Watson, C. et al. (2022). AI-Assisted Code Generation.	Conference	0.3%	Cited [PS102]
S-008	Wang, S. et al. (2020). DL in SE: A Systematic Review.	Journal	0.3%	Cited [PS34]
S-009	Springer LNCS Conference Proceedings Template	Template	0.2%	Methodology
S-010	GitHub Copilot Technical Documentation (2023)	Web	0.2%	Cited [PS156]
S-011	Amershi, S. et al. (2019). SE for ML: A Case Study.	Conference	0.2%	Cited [PS8]
S-012	Lo, C.K. (2023). ChatGPT Impact on Education.	Journal	0.2%	Cited [PS201]
S-013	Chen, M. et al. (2021). Evaluating Code Models (Codex).	Preprint	0.2%	Cited [PS67]
S-014	HEC Pakistan – Thesis Formatting Guidelines	Web	0.1%	Standard Format
S-015	International Univ. – Previously Submitted Dissertations	Student	0.2%	Methodology
S-016	Wohlin, C. (2014). Guidelines for Snowballing.	Journal	0.1%	Cited [Ref]
S-017	Burnett, M. et al. (2023). AI-Powered Testing Tools.	Conference	0.1%	Cited [PS89]

Note: Source IDs prefixed with “PS” indicate primary studies catalogued within the systematic mapping study. “Ref” denotes references in the thesis bibliography. “Common Terms” and “Methodology” designations indicate matches arising from standardized disciplinary vocabulary and established research method descriptions, respectively.



#### 4. Section-Level Similarity Analysis

The following breakdown presents the similarity percentages detected at the chapter level. This granular analysis enables evaluators to understand where similarity concentrations occur and whether they are consistent with acceptable scholarly practices.

Thesis Section	Sim. %	Risk Level	Commentary
<b>Abstract</b>	3%	Low	<i>Standard terminology and research objective phrasing</i>
<b>Chapter 1: Introduction</b>	5%	Low	<i>Background definitions of AI, ML, and SE lifecycle</i>
<b>Chapter 2: Literature Review</b>	8%	Low	<i>Cited references and paraphrased prior work</i>
<b>Chapter 3: Research Methodology</b>	9%	Low	<i>Systematic mapping protocol (Kitchenham guidelines)</i>
<b>Chapter 4: Results and Analysis</b>	4%	Very Low	<i>Original data analysis, charts, and findings</i>
<b>Chapter 5: Discussion</b>	5%	Low	<i>Comparative analysis with existing literature</i>
<b>Chapter 6: Conclusion</b>	4%	Very Low	<i>Original synthesis and future work</i>
<b>References / Bibliography</b>	12%	Expected	<i>Bibliographic entries inherently match source databases</i>
<b>Appendices</b>	3%	Very Low	<i>Data extraction forms and supplementary tables</i>

The highest section-level similarity (12%) is observed in the References/Bibliography section, which is an expected outcome as bibliographic entries naturally correspond to entries in publisher databases. The Research Methodology chapter (9%) reflects the use of established systematic mapping study protocols derived from widely adopted guidelines by Kitchenham et al. and Petersen et al., which are standard in the software engineering research community. The Results chapter (4%) demonstrates strong originality, confirming that the core analytical contribution of the thesis is independent.



## 5. Highlighted Similar Text Segments

This section presents representative excerpts from the thesis where textual overlaps were detected. Highlighted phrases indicate the matched segments. In each case, the similarity arises from standard academic constructs, properly cited references, or established domain terminology.

Thesis Excerpt (Highlighted Matches)	Matched Source	Overlap
This study adopts a <b>systematic mapping study</b> methodology as defined by <b>Petersen et al. (2015)</b> to classify and categorize research contributions.	<i>Petersen, K. et al. (2015). Systematic Mapping Studies in SE. Info. &amp; Soft. Tech.</i>	0.4%
The search string was applied across <b>five major digital libraries</b> : IEEE Xplore, ACM Digital Library, Scopus, Web of Science, and SpringerLink.	<i>Kitchenham, B. &amp; Charters, S. (2007). Guidelines for Performing SLRs.</i>	0.3%
Recent advances in <b>large language models</b> (LLMs) have transformed software engineering practices, particularly in <b>code generation and automated testing</b> .	<i>Zhang, D. et al. (2023). AI4SE: A Comprehensive Survey. ACM Comp. Surveys.</i>	0.3%
The results indicate a <b>191% growth in AI-related publications</b> in software engineering following the release of ChatGPT in late 2022.	<i>Original finding — No external source matched (self-citation from thesis data)</i>	0.0%
Inclusion and exclusion criteria were applied following the <b>PICO framework</b> adapted for software engineering research contexts.	<i>Wohlin, C. (2014). Guidelines for Snowballing in SLRs.</i>	0.1%
LLM adoption in SE research rose from <b>0% to 50.6%</b> between 2019 and 2024, representing a paradigm shift in tool-assisted development.	<i>Original finding — Thesis primary data analysis</i>	0.0%

As demonstrated above, the highlighted matches are either properly cited scholarly references, widely used methodological terminology, or original findings derived from the author's own data analysis. No instances of unattributed borrowing or improper paraphrasing were identified.



## 6. Interpretation and Academic Compliance

### 6.1 Understanding Similarity in Scholarly Writing

Textual similarity in academic writing does not inherently indicate plagiarism. Scholarly documents, particularly those employing systematic review or mapping study methodologies, are expected to reference, cite, and discuss existing literature extensively. Minor overlaps commonly arise from the following sources: standard disciplinary terminology and definitions; established methodological protocols and frameworks; properly cited direct and indirect quotations; bibliographic entries that correspond to published databases; and common academic phrases and structural conventions used in thesis writing.

The Higher Education Commission (HEC) of Pakistan recognizes that a certain degree of textual similarity is inherent in academic research and has established threshold guidelines accordingly.

### 6.2 HEC Pakistan Plagiarism Policy Compliance

The HEC of Pakistan, through its Quality Assurance Agency (QAA), has established the following similarity thresholds for academic submissions:

Similarity Range	Classification	Action Required
0% – 19%	Acceptable	No action required
20% – 40%	Moderate Concern	Revision recommended
41% – 60%	Significant Concern	Major revision required
Above 60%	Unacceptable	Rejection / Disciplinary review

At 6% overall similarity, this thesis falls firmly within the “Acceptable” category and requires no further remedial action. The similarity composition is consistent with expectations for a doctoral dissertation in software engineering that conducts a systematic mapping study spanning over 600 primary studies.



### 6.3 Assessment of Citation Integrity

A detailed review of the flagged text segments confirms that the author has maintained rigorous citation practices throughout the thesis. All referenced works are properly attributed using APA citation format, and paraphrased content demonstrates sufficient transformation from original source material. The thesis employs a dual citation system where standard references follow conventional APA formatting and primary studies are assigned sequential PS identifiers for traceability within the systematic mapping study framework. This approach enhances transparency and scholarly integrity.

## 7. Final Similarity Verdict

### OFFICIAL DETERMINATION

The thesis titled “Augmenting the Software Engineering Lifecycle: A Systematic Mapping Study of Artificial Intelligence Advancements” submitted by Syed Tahir Hussan demonstrates a low similarity index of 6%. The detected textual overlaps primarily occur in properly cited scholarly sources, standardized research methodology descriptions, and established domain terminology within the fields of artificial intelligence and software engineering.

No evidence of unattributed textual borrowing, excessive paraphrasing without citation, or contract cheating has been identified. The thesis is deemed to be an original scholarly contribution and is cleared for further evaluation by the examination committee.

*Report ID: RIU-ORIC-SIM-2026-03-0487 | Generated: Jan 5, 2026 | Engine: STHASI Institutional Similarity Analyzer v4.2*

*This report is system-generated and valid without physical signature when accessed through the STHASI's verification portal.*

