

Codebook for Qualitative Data Analysis of NLP Tool Explainability in Requirements Engineering

Category 1: Participant Background and Experience

Captures professional history, role evolution, and initial perceptions of NLP tools.

Main Code	Sub-Code	Definition
Professional Experience	Career Duration	Number of years in the field.
	Field Focus	Specific expertise area (e.g., RE, safety-critical industries).
Early Career	Manual Processes	Description of RE workflows before adopting NLP tools.
	Traditional Methods	Manual tasks performed before automation (e.g., documentation, validation).
Role Evolution	Shift to Automation	Transition from manual methods to AI-assisted RE tasks.
	New Responsibilities	New duties related to NLP tool adoption (e.g., reviewing AI-generated outputs).
Initial Expectations	Efficiency Promise	Anticipated benefits of using NLP tools in RE.

Category 2: Current Use of NLP Tools (RQ1)

Addresses how NLP tools are integrated into RE workflows, their benefits, and their limitations.

Main Code	Sub-Code	Definition
Tool Usage	Tool Name	Name of the NLP tool used (e.g., Eraser.io, DiagramGPT).
	Tool Function	Primary purpose of the tool (e.g., design artifact generation, requirement validation).
Design Artifact Generation	Artifact Type	Type of outputs generated (e.g., UML diagrams, sequence diagrams).
Efficiency Gains	Speed Improvement	Reduction in task completion time due to NLP tools.
	Scale Benefit	Ability to handle large volumes of textual requirements more efficiently.

Tool Adoption Timeline	Adoption Time	When NLP tools were first introduced in the organization.
	Evolution Context	Broader industry trends influencing NLP tool adoption.

Category 3: Challenges of Non-Explainability (RQ2)

Identifies issues related to the opacity of NLP-generated outputs and their impact on RE workflows.

Main Code	Sub-Code	Definition
Non-Explainable Outputs	-	Vague, unclear, or misleading NLP-generated artifacts.
Manual Validation	-	Need for extensive manual verification of tool outputs.
Trust Issues	Tool Reliability Doubt	Hesitation to fully rely on NLP-generated outputs.
	Misinterpretation Risk	Different team members interpreting the same output inconsistently.
	Error Potential	Increased risk of design mistakes due to ambiguous or incorrect NLP outputs.
Domain Limitations	Context Ignorance	NLP tools failing to account for domain-specific terminology and constraints.
Time Delays	Review Delay	Additional time required to verify tool-generated outputs.
	Project Impact	Broader project timeline disruptions due to NLP limitations.
Cost Increases	Labor Cost	Increased costs due to manual validation efforts.
	Rework Expense	Financial burden caused by fixing errors introduced by NLP tools.
Frequency of Issues	Occurrence Rate	How often unclear outputs are generated by NLP tools.
Specific Example	Case Description	Detailed incident demonstrating a challenge.
	Resolution Effort	Steps taken to correct or mitigate the issue.

Category 4: Strategies to Mitigate Challenges (RQ3)

Documents workarounds and adaptive strategies used by practitioners to handle NLP tool limitations.

Main Code	Sub-Code	Definition
Source Tracing (Manual)	-	Manually linking tool outputs to input requirements.

Team Review Meetings	Group Discussion	Collaborative analysis of NLP-generated outputs.
	Consensus Building	Resolving discrepancies in tool-generated artifacts through team discussions.
Documentation	Issue Logging	Systematic recording of tool-related issues.
	Decision Tracking	Documentation of decisions made during validation.
Hybrid Approach	-	Combining NLP-generated outputs with manual validation to improve reliability.
Traceability Matrix	Mapping	Using structured traceability matrices to validate NLP outputs.

Category 5: Impacts on Workflow, Collaboration, and Projects (RQ4)

Explores how NLP tools affect RE workflows, teamwork, and project execution.

Main Code	Sub-Code	Definition
Workflow Disruption	-	Delays and inefficiencies caused by NLP-generated artifacts requiring extensive review.
Team Misalignment	Interpretation Variance	Different stakeholders interpreting NLP-generated outputs inconsistently.
	Conflict Increase	Increased disagreements within teams due to unclear or incorrect tool-generated artifacts.
Stakeholder Communication	Regulatory Challenge	Difficulty explaining NLP-generated outputs to regulatory auditors and external stakeholders.
Project Risk	Safety Threat	Potential for safety-critical errors caused by non-explainable NLP-generated designs.

Category 6: Desired Features for Improvement (RQ5)

Identifies practitioner-recommended enhancements to improve NLP tool adoption and effectiveness.

Main Code	Sub-Code	Definition
Source Tracing (Automated)	Output Linking	Automatic linking of generated outputs to specific input requirements.
Contextual Justifications	Domain Relevance	Providing explanations tied to industry-specific standards and terminology.
	Decision Clarity	Offering clear justifications for tool-generated modifications or omissions.

Validation/Compliance Support	-	Automated checks to ensure NLP-generated artifacts align with industry regulations.
	Compliance Report	Generating summaries that document regulatory adherence.
Real-Time Feedback	Parameter Adjustment	Allowing practitioners to tweak settings and receive immediate feedback on outputs.
	Dynamic Output	Instant reflection of modifications in NLP-generated designs.
Collaborative Features	Annotation Tools	Enabling teams to add shared comments on NLP-generated outputs.
	Version Control	Tracking changes and maintaining revision history.
	Discussion Threads	Integrating team-based discussion features within NLP tools.
Feedback Loop	User Correction	Allowing practitioners to override incorrect NLP-generated outputs.
	Algorithm Refinement	Using user feedback to improve tool performance over time.