Improving Human-Computer Interaction through domain-specific Usability Heuristics for assessing Requirements Engineering and Management tools

PhD Thesis Summary

Politehnica University of Timisoara
Doctoral domain of
Electronics. Telecommunication and Information Technology
Author: Ing. Rotaru Oana Alexandra
Coordinator: Prof. Dr. Ing. Radu Vasiu
Month: 06 Year:2026

1. Motivation

Information Technology (IT) has fundamentally transformed the automotive industry, shifting it from purely mechanical (ME) systems to highly complex, software (SW) driven, and increasingly autonomous systems. This transformation has significantly impacted engineering practices, particularly through the adoption of SW platforms and information systems that support development processes [1].

As modern vehicles rely heavily on SW, the complexity and diversity of requirements—covering safety, functionality, and regulatory compliance—have increased substantially. This evolution has led to the need for structured software lifecycle frameworks, specifically Application Lifecycle Management tools, with a strong focus on Requirements Engineering and Management (REM) [2].

REM is a core discipline within systems and SW engineering, supported by international standards such as ISO / IEC / IEEE 29148 [3], and closely linked to broader lifecycle standards (ISO / IEC / IEEE 15288 [4] and 12207 [5]). In the automotive domain, REM activities are critical due to the development of safety-critical systems. To support these activities, specialised tools such as Polarion [6], Jama Connect [7], and IBM DOORS Next [8] (DNG) are widely used.

These tools are used by various stakeholders with varying workflows and objectives. As a result, their effectiveness depends not only on functional capabilities, but also on how efficiently users can interact with them. This brings the research into the domain of Human-Computer Interaction (HCI), which focuses on designing and evaluating user-centred systems. Key HCI concepts include User Experience (UX), usability, and interface design, as defined by standards such as ISO 9241 [9]. Although general usability evaluation methods—such as Nielsen's heuristic evaluation [10]—provide a solid foundation, their generic nature can limit their effectiveness in addressing domain-specific challenges found in complex REM tools.

This thesis explores the intersection between REM and HCI, with the aim of balancing advanced engineering functionality with effective user interaction. It proposes domain-specific usability evaluation methods tailored to REM tools in the automotive domain. Additionally, it investigates the potential of Large Language Models (LLMs) as Artificial Inteligence (AI) based tools to conducting heuristic evaluations (HE) [11].

The research is guided by the following key questions:

1. How do the HCI considerations reflect in the assessment of REM tools?
2. Are generic heuristics effective in identifying domain-specific usability issues related to specific workflows of automotive REM tools?
3. Compared to generic heuristics, how effective are domain-specific heuristics, tailored to REM tools, in identifying usability issues?
4. How accurately can LLMs perform HE of REM tools, in terms of issue detection, severity, and overall coverage?

To address these questions, the thesis first reviews the state of the art in HCI applications within REM tools and then proposes a set of domain-specific heuristics designed for usability evaluation of REM tools used within the automotive context.

The thesis is structured in seven chapters as follows:

Chapter 1 "Motivation" exposes the rationale for selecting the current topic of the research and is structured into three sections: a general overview of the subject, where the interconnected domains are explained, the list of published articles resulting during the thesis, and the presentation of the structure of the PhD thesis.

Chapter 2 "Foundations of Requirements Engineering in the Automotive Industry" offers the theoretical foundation of the research, focused on Automotive Engineering (AE) and Requirements Engineering (RE). The main purpose of this chapter is to present the domain of RE, used in automotive industry, including the processes used, the tools and their functionalities and capabilities. In addition, the chapter addresses the application of AI within the domain of RE, with a stronger emphasis on LLMs and a prompt engineering role in shaping the outputs. The chapter presents the main domain concepts and motivation for investigating UX in REM tools.

Chapter 3 "Human Computer Interaction for Requirements Engineering and Management Tools" addresses the review of the literature on usability and UX in context of REM tools. The chapter begins with a theoretical base on the evolution of UX, followed by a discussion on the differences between the usability and UX concepts; followed by a review of usability evaluation methods with an accent on the HE method. The chapter also presents a literature related work on the existing domain-heuristics, ultimately presenting the applicability of UX in several broader digital platforms, in REM tools and the applicability of LLMs in usability evaluation.

Chapter 4 "Proposal of REM specific usability heuristics for Automotive REM tools" consists of the proposal of REM tool specific heuristics, obtained after applying the Prometheus methodology [12] dedicated to developing domain-specific heuristics. The stages of the methodology are explained in theory and applied in the current case, with the last stage of the methodology being presented in the next chapter, as part of the heuristics validation.

Chapter 5 "Validation of the proposed Requirements Engineering and Management specific usability heuristics" presents the validation of the REM-specific heuristics proposal. To validate the proposed domain heuristics, the final stage of Prometheus methodology is applied and two HEs are conducted using Nielsen's heuristics as a control set and the REM-specific heuristics as the domain set. The evaluation environment and procedure are presented, and the results are structured as follows: a summary of all issues (the issues identified using Nielsen and the issues identified using domain heuristics), the evaluators suggestions gathered during the assessments, a cross-comparing of the resulted issues, the computed quality indicators, and finally, the results summary and interpretation.

Chapter 6 "Exploratory analysis: Heuristics validation through Large Language Models" represents an exploratory chapter focused on the HE using LLMs. Across this chapter, the motivation behind this idea is presented, the applied evaluation methodology and the results and discussions are also included here. The main purpose of this chapter is to explore the LLM-based HEs and to compare the results with those obtained by human-based evaluation.

Chapter 7 "Contributions and Conclusions" is the final chapter of the thesis, focused on the contributions and conclusions. It includes a summary of the contributions, a brief discussion on the theoretical contributions to literature, the implications for the industry, and, lastly, the potential future work intended.

2. Foundations of Requirements Engineering in the Automotive Industry

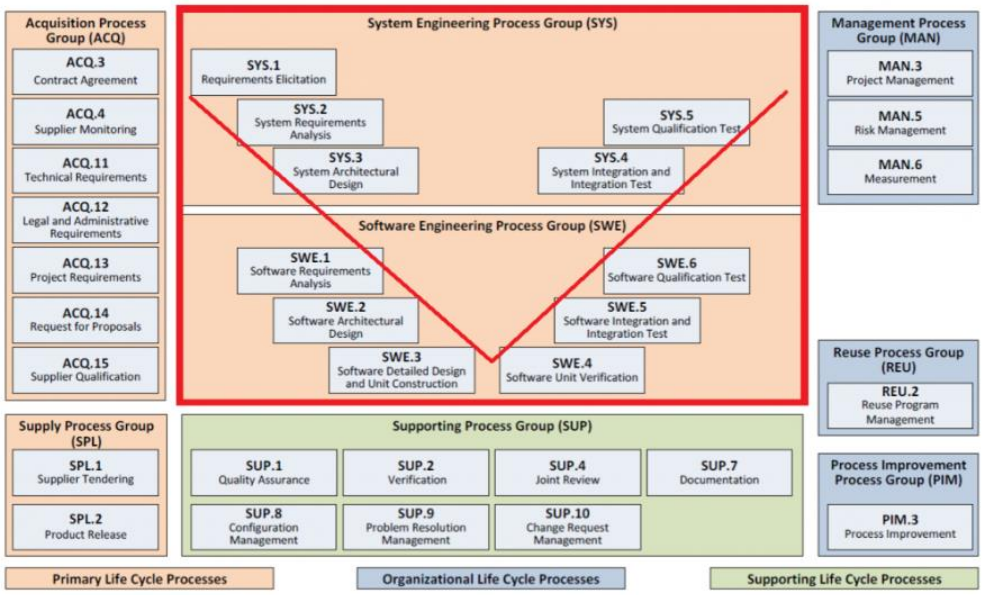


Figure 1. Automotive SPICE process reference model [13]

Modern automotive development relies on well-defined lifecycle processes to ensure safety, quality and compliance, supported by the V-model [14], which establishes a direct relationship between requirements definition and corresponding verification and validation activities. Development approaches can include traditional Waterfall methodologies [15], providing structure and traceability, Agile methodologies, enabling flexibility and iterative delivery, or hybrid models combining both to address HW constraints and SW dynamics [16].

In this context, RE represents a core discipline responsible for eliciting, analysing, specifying, validating, and managing system requirements throughout the lifecycle [17]. As seen in Figure 1 RE ensures alignment between stakeholder needs and technical implementation, based on principles such as clarity, completeness, consistency, traceability, feasibility, and verifiability.

In the automotive domain, RE is essential for managing safety-critical systems, supporting regulatory compliance and handling both functional and non-functional requirements in complex environments. RE activities extend throughout the development lifecycle, from concept definition to system design, implementation, testing, and post-production support, closely aligned with the V-model structure [14].

Traceability plays a fundamental role in linking requirements across abstraction levels and lifecycle artifacts, ensuring coverage, transparency and compliance, while change management provides structured workflows for handling requirement evolution through impact analysis, approval and verification [18].

Automotive RE is strongly governed by standards such as ISO 26262 for functional safety [19], ASPICE for process capability assessment [20], ISO / IEC / IEEE 29148 for RE practices [3] and ISO/SAE 21434 for cybersecurity requirements [21].

To support these activities, specialised REM tools are used, enabling structured authoring, hierarchical organisation, bidirectional traceability, version control, workflow automation, collaboration and integration with other engineering systems [22]. Examples of such tools include IBM DOORS [23], DNG [8], Siemens Polarion [6], Jama Connect [7], and ReqView [24], each offering different levels of scalability, collaboration, and integration capabilities. Table 1 offers an analysis of different popular REM tools in terms of features offered.

Table 1 Comparative Feature Table of REM tools.

Feature	DOORS	DNG	Siemens Polarion	Jama Connect	ReqView
Structured Authoring & Hierarchy	Yes	Yes	Yes	Yes	Yes
Traceability	Yes	Yes	Yes	Yes	Yes
Versioning & Baselines	Yes	Yes	Yes	Yes	Yes
Configuration Management	Partial	Yes	Yes	Partial	Partial
Collaboration and Review	No	Yes	Yes	Yes	No
Integration with Test, MBSE tools	Partial	Yes	Yes	Partial	No
ReqIF Import/export support	Yes	Yes	Yes	Yes	Yes
Web-based interface	No	Yes	Yes	Yes	Partial

The selection of REM tools depends on project complexity, team distribution, process maturity, integration needs, regulatory requirements, and cost considerations. In recent years, AI, particularly LLM [25], has emerged as a supporting technology in RE, enabling tasks such as requirements elicitation, classification, quality assessment, traceability support, and change impact analysis. Despite these opportunities, AI integration remains limited by challenges such as accuracy, domain specificity, and the risk of unreliable output.

Prompt engineering plays a critical role in effectively leveraging LLMs, involving structured input design through clear instructions, contextual information, and defined constraints to improve output quality. Different prompting strategies, including zero-shot, few-shot, and chain-of-thought approaches, enable more controlled and goal-oriented AI interactions [26].

In general, the chapter emphasises that RE is a fundamental pillar in automotive development, supported by structured processes, strict standards, and specialised tools, while emerging AI technologies introduce new opportunities to improve efficiency and analysis capabilities.

3. Human-Computer Interaction for REM Tools

The effectiveness of REM tools depends not only on functionality, but also on usability and UX, which directly influence efficiency, accuracy, and user satisfaction in complex automotive environments [27].

HCI provides the theoretical foundation for understanding UX and usability, focusing on human-centred design and structured evaluation methods such as empirical testing, inspection techniques, and heuristic evaluation. All the HCI rules are included in the HCI standard, ISO 9241 [9].

Although generic heuristics offer a strong baseline, they lack contextual specificity for complex systems such as REM tools, which involve high cognitive load, traceability structures, versioning, and multidisciplinary workflows [28].

These limitations highlight the need for domain-specific heuristics, validated through structured methodologies, to better capture workflow-related usability issues.

Recent research and applications across digital platforms demonstrate the importance of tailored UX approaches and combined evaluation methods, while AI and LLMs introduce new opportunities for usability assessment, as complementary tools due to their limitations [11].

The study [29] demands an expansion of the usability research of REM tools, focusing on collaboration, traceability, and visualisation to better understand recurring issues. Another prior research addressing the existing gap explores the usability assessment of REM tools collaboration features [30]. The study presents the usability evaluation of collaboration features in DNG within an automotive department, using methods such as user testing, focus groups, and the SUS survey. Another intriguing prior research has addressed the topic of analysing and comparing Agile role-specific needs for REM tool interfaces and is a survey-based study [31].

Overall, effective usability evaluation of REM tools requires a combination of domain-specific heuristics, structured methodologies, and emerging AI support to address the complexity of modern automotive systems.

4. Proposal of REM-Specific Usability Heuristics for Automotive REM Tools

The use of generic heuristic evaluation methods is often insufficient for complex and specialised domains characterised by specific workflows, terminology, and constraints, as such heuristics provide only general guidance and fail to capture deeper usability issues. Although frameworks such as Nielsen's [32] or Gerhardt-Powals' [33] heuristics offer a valuable baseline, they require adaptation to effectively address domain-specific interaction complexities.

Existing literature across domains such as interactive television [34], medical systems [35] and virtual learning environments[36] demonstrates that tailored heuristics provide more relevant and precise usability insights, particularly when developed using structured methodologies such as Prometheus. Previous research conducted within this thesis confirms that generic heuristics are effective in identifying basic usability issues, while deeper problems related to domain-specific workflows remain insufficiently addressed, highlighting the need for customised evaluation frameworks.

In complex systems such as automotive REM tools, this limitation becomes more pronounced due to the presence of specialised requirements engineering activities, including traceability, validation, change management, and collaborative workflows [22].

These platforms involve multiple user roles with distinct responsibilities, increasing interaction complexity, and amplifying the risk of usability issues in different usage contexts.

Previous usability studies on automotive REM tools, such as DOORS and DNG, reveal recurring problems related to navigation, interface complexity, traceability mechanisms, versioning accessibility and collaboration features [30] [29]. Empirical evaluation of these tools also demonstrates that usability testing, although effective, is resource-intensive due to the need for real environments, multiple users, and extensive scenarios.

These challenges motivate the development of domain-specific heuristics as a more efficient and scalable approach for usability evaluation, capable of uncovering deeper structural issues while reducing the dependency on extensive user testing.

To support this objective, the Prometheus methodology [37] is adopted as a structured framework for heuristic development, providing a systematic multi-stage process including domain analysis, heuristic selection, specificity evaluation, normalization, prioritization, description, validation and refinement.

Within this research, the resulting REM-specific heuristics are applied in a comparative heuristic evaluation on an industrial tool (DNG), involving expert evaluators and benchmarked against standard heuristics, enabling the assessment of their effectiveness in capturing domain-relevant usability issues.

In addition, the proposed REM-specific and generic heuristics are evaluated using an LLM – Chat GPT-5 [38], to explore its ability to perform structured usability assessments and provide a secondary validation of the proposed framework.

The thesis contributes a novel usability evaluation framework tailored to automotive REM tools, allowing the identification of workflow-related and structural usability issues. Prometheus methodology [37] was applied in order to get to the REM-specific heuristic set. The applied stages are presented in Figure 2.

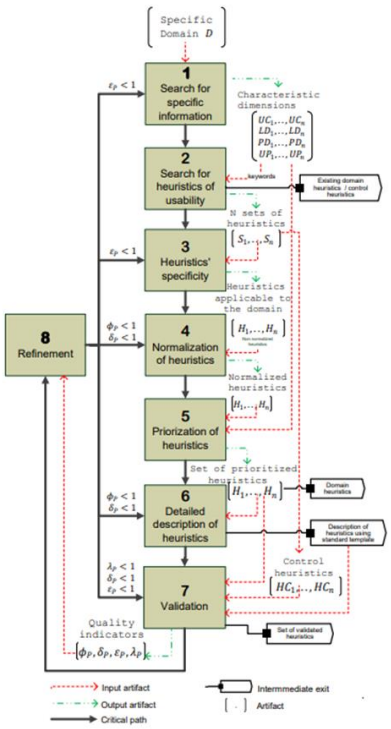


Figure 2. Prometheus stage flow diagram [37]

After applying all the stages, the REM-specific usability heuristics are presented as follows:

HREM1. Feedback and System Status

Description: The REM tool should provide clear and contextual feedback to the users and keep them informed on the system operations. This can include visible status changes, failure/success messages, progress indicators on tasks like artifacts linking, and requirements specification.

HREM2. Interface and Terminology Consistency

Description: The REM tool interface should be consistent in visual elements, messages, and terminology. The terms used must be specific to the REM and contextually adapted to the scenarios.

HREM3. Error Prevention

Description: The REM tool interface should prevent the users from making errors during usual activities such as artifacts linking, editing, specification, and deletion.

HREM4. Flexibility

Description: The REM tool interface should allow users to perform actions in multiple ways, depending on their expertise, preferences, or needs. The interface should allow shortcuts or auxiliary views to improve user flexibility.

HREM5. User control, reversibility

Description: When interacting with REM tools, users should feel in control of their actions, being able to easily revert their actions. These actions can refer to requirements changes, deletions, artifacts links updates.

HREM6. Familiarity with the domain

Description: The REM tool interface should support domain relevant language, RE workflows expected by the domain users.

HREM7. Error recovery / diagnosability

Description: REM tools should support recoverable actions in case of user errors. Using the REM tool, the users should be able to understand the issues, locate the source (e.g. incorrect linkage) and easily recover where possible.

HREM8. Memory load reduction

Description: The REM tools interface should minimise the users' need to memorise information between different screens or interactions. Terms, functions, widgets, artefact states should be visible, easy to remember from past interactions.

HREM9. Help and Documentation

Description: REM tools should offer easily accessible help and documentation. This includes general usage instructions and REM-specific guidance for more difficult actions such as traceability linkage or versioning.

HREM10. Collaboration and Communication support

Description: REM tools should support collaboration and communication mechanisms between stakeholders involved in RE topics. This includes commenting, user tagging, notifications, and visibility of changes made by other users.

HREM11. Aesthetic, minimalist design

Description: REM tools should offer clean interfaces, with information presented only for the tasks handed, with visuals and icons related to RE workflows.

HREM12. Review and Validation Support

Description: REM tools interfaces should provide clearly structured support for review and validation workflows. Interfaces should visually distinguish review phases, provide clear mechanisms for approval or rejection actions, and allow access to validation data (e.g. status, comments). These features should visually help in collaborative validation sessions.

HREM13. Versioning and Baseline accessibility

Description: The REM tools interface should provide accessible and user-friendly navigation to versioned artifacts or baselines. Users should easily compare and interpret different versions through clear visual structures and guided workflows.

HREM14. Artifact history tracking

Description: REM tools should allow users to easily track the history of changes for individual artifacts, with clean structuring of what changed, when, why, and by whom. Users should be able to access visual timelines, filters, and summaries.

HREM15. Navigation and Filtering support

Description: REM tools interfaces should provide easy navigation mechanisms and filtering options to help users operate on the relevant artifacts. Navigating in REM tools modules artifacts should follow the traceability logic of RE.

The proposed heuristics presented above, HREM1-HREM15, consist of a tailored framework for usability evaluation of REM tools used in the automotive industry. The heuristic set proposed to cover important usability concerns, focusing on key RE activities, features, and workflows, typical for the REM tools.

The HREM1-HREM11 heuristics are derived from the normalised heuristic set HN1-HN11 and are tailored for the REM tools domain, while HREM12-HREM15 are the new proposed heuristics, which were not covered by the normalised set.

Together, the proposed heuristic set covers usability challenges encountered while interacting with REM tools, reflecting the complexity, multi-role collaboration, and compliance-driven dimension of the automotive sector.

Following the Prometheus methodology, the proposed heuristic set will be validated within the next stage, including the next main chapter of the thesis. The effectiveness of REM-specific heuristics will be validated by comparing with a control heuristic set, Nielsen's heuristics. Prometheus supports the validation stage with quantitative indicators to support this comparison, which can be used further for refinement, if needed. In addition, qualitative insights will be obtained that support this evaluation.

5. Validation of the proposed REM-Specific Usability Heuristics

Stages 7 and 8 of the Prometheus methodology focus on the validation and iterative refinement of the proposed REM domain-specific heuristics. These stages are important in demonstrating the relevance, usefulness and fidelity of the customised set of heuristics when applying them in usability evaluations within the domain.

Stage 7 aims to compare the proposed REM-specific heuristics with a control set, Nielsen's, to evaluate the effectiveness in identifying usability issues. So, validation is done through the HE method, using Nielsen and REM-specific heuristic sets, on a domain representative platform, DNG.

The HE was organised using a dedicated test workspace prepared on DNG with a valid corporate licence and with the approval of an automotive company department, where the tool is used. The dedicated test workspace was especially created for the HEs, without conflicting with other projects modules or configurations used in the company.

Three requirement modules were created inside a "Test" local configuration created for the evaluation. The requirement modules are named: "Heuristic Evaluation test document #1", "Heuristic Evaluation test document #2", "Heuristic Evaluation test document #3". Each module contained several requirements, named using structured IDs such as 1.1, 2.1, 3.2. The test workspace captions are presented in Figure 4.

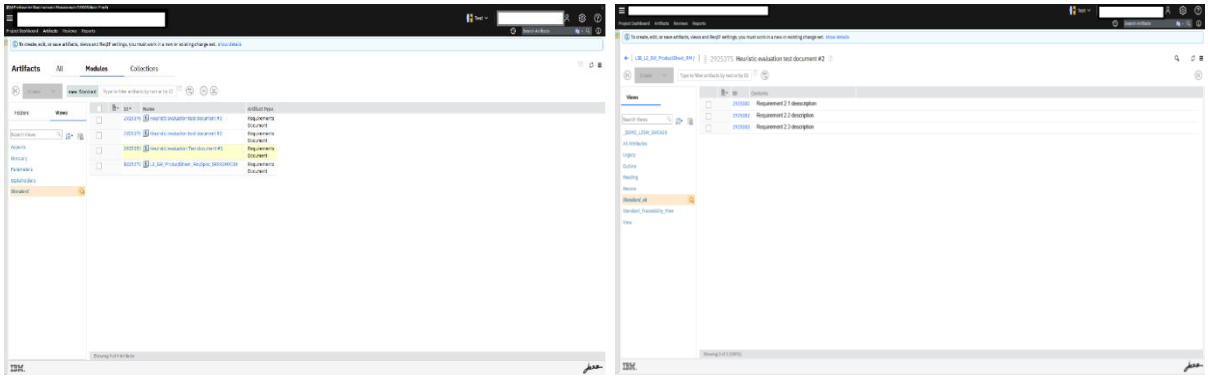
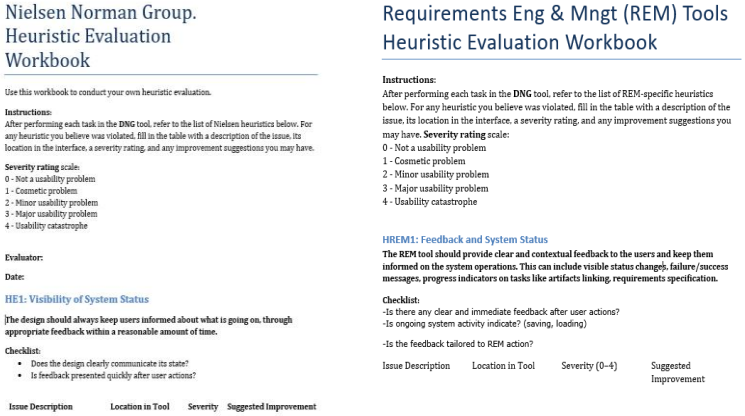


Figure 4. Test workspace in DNG

Ten evaluators were recruited for the HE. Five evaluators assessed the REM tool using the domain heuristics set, and five of them were the control group, evaluating the tool using the control heuristics set. Assessors were selected on the basis of their relevant experience in usability and interface evaluation. As a background, all the evaluators had a strong knowledge of the usability and HCI principles. To ensure smooth operation, the following materials were created and used in HE: evaluation workbooks and an introduction document for RE, snapshot presented in Figure 5.



Nielsen Norman Group.
Heuristic Evaluation
Workbook

Use this workbook to conduct your own heuristic evaluation.

Instructions:
After performing each task in the **DNG** tool, refer to the list of Nielsen heuristics below. For any heuristic you believe was violated, fill in the table with a description of the issue, its location in the interface, a severity rating, and any improvement suggestions you may have.

Severity rating scale:
0 - Not a usability problem
1 - Cosmetic problem
2 - Minor usability problem
3 - Major usability problem
4 - Usability catastrophe

Evaluator:

Date:

HE1: Visibility of System Status

The design should always keep users informed about what is going on, through appropriate feedback within a reasonable amount of time.

Checklist:
- Does the design clearly communicate its state?
- Is feedback presented quickly after user actions?

Issue Description	Location in Tool	Severity	Suggested Improvement

Requirements Eng & Mngt (REM) Tools
Heuristic Evaluation Workbook

Instructions:
After performing each task in the **DNG** tool, refer to the list of REM-specific heuristics below. For any heuristic you believe was violated, fill in the table with a description of the issue, its location in the interface, a severity rating, and any improvement suggestions you may have. **Severity rating** scale:
0 - Not a usability problem
1 - Cosmetic problem
2 - Minor usability problem
3 - Major usability problem
4 - Usability catastrophe

HREM1: Feedback and System Status

The REM tool should provide clear and contextual feedback to the users and keep them informed on the system operations. This can include visible status changes, failure/success messages, progress indicators on tasks like artifacts linking, requirements specification.

Checklist:
-Is there any clear and immediate feedback after user actions?
-Is ongoing system activity indicate? (saving, loading)

-Is the feedback tailored to REM action?

Issue Description	Location in Tool	Severity (0–4)	Suggested Improvement

Figure 5. Snapshot of HE workbook: (left) control , (right) domain specific.

A set of six structured tasks was developed for the HE sessions, to fully trigger both control and domain-specific heuristics. These tasks were intended to simulate realistic user interactions within a REM tool, in this case DNG, aiming to ensure that main activities and workflows are performed. The tasks used during the HE are presented in Table 2.

Table 2 Task details for HE.

ID		Task details
1	Task name	**Create and specify a new requirement**
	Prerequisites	An open change set needed;
	Steps	Create a new requirement in „Heuristic evaluation document #1", named "Requirement 1.4". Set its Status to "In work" and add a comment.
2	Task name	**Link artifacts and explore traceability**
	Prerequisites	An open change set needed;
	Steps	You are in „Heuristic evaluation document #2": "Requirement 2.1" should derive from Requirement "1.2" (from "Heuristic evaluation document #1"). Create the linkage from "Requirement 2.1" to "Requirement 1.2", add a comment for the "Requirement 2.1" and then open the links explorer to see the linkage.
3	Task name	**Edit an artifact**
	Prerequisites	An open change set needed;
	Steps	Modify "Requirement 3.1" from „Heuristic evaluation document #3". Deliver the change set. Now, track this modification of the requirement artifact.
4	Task name	**Search and filter requirements**
	Prerequisites	An open change set needed;
	Steps	Filter the requirements from „Heuristic evaluation document #1", based on the Released status. Create a personal view with the filtered artifacts.
5	Task name	**Access and compare baselines**
	Prerequisites	Outside a change set;
	Steps	Create a baseline for „Heuristic evaluation document #2". Compare the created baseline with the last one.
6	Task name	**Review and check for validation**
	Prerequisites	Outside a change set;
	Steps	Create a review for "Requirement 1.3" from „Heuristic evaluation document #1". Complete the review. Check for the validation artifacts of this requirement.

For both evaluation sessions, the induction part, explaining the setup, and a free exploration of the tool lasted a median of 15 minutes. The Nielsen-based evaluation took

approximately 75 minutes on average, while the REM-specific-based evaluation took approximately 105 minutes, due to its larger number of heuristics and their more detailed descriptions.

After the HE sessions conducted using control heuristic set, 30 unique usability issues were identified after duplicate removal. After the HE sessions conducted using the proposed REM-specific heuristic set, 46 unique usability issues were identified after duplicate removal. The issues mapped all heuristics, ten in the control case and fifteen in the domain case, and were rated with degrees of severity by the evaluators. Each issue was assigned a severity level between 0-4 (5-level Likert scale), depending on the HE practices. The severity distribution for each HE session is presented in Figure 6, respectively, the distribution of severity per heuristic in Figure 7.

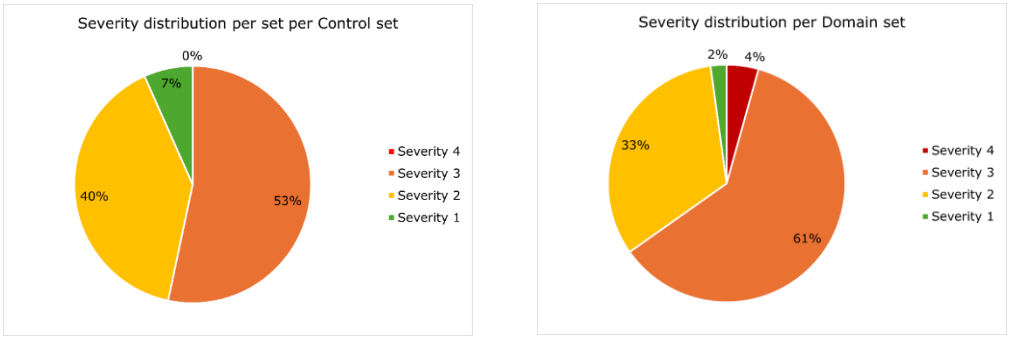


Figure 6. Severity distribution of usability issues per control and domain set.

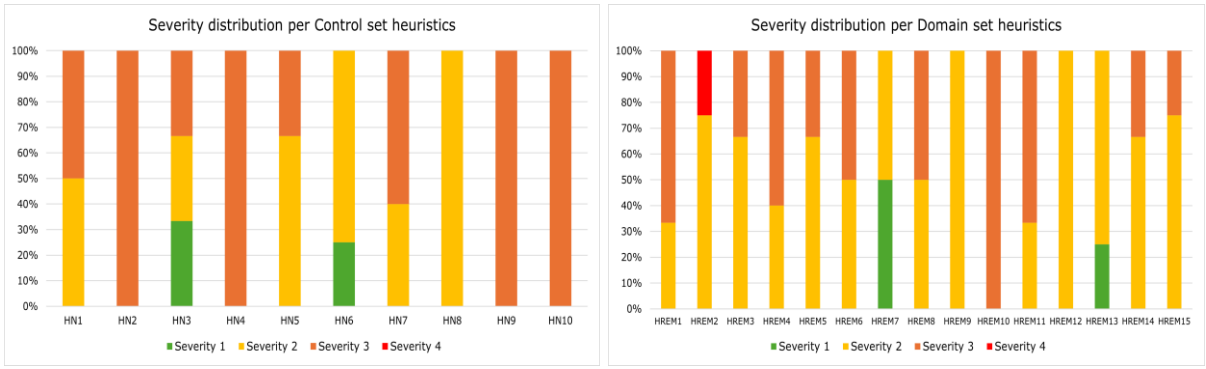


Figure 7. Severity distribution of usability issues per control and domain heuristic.

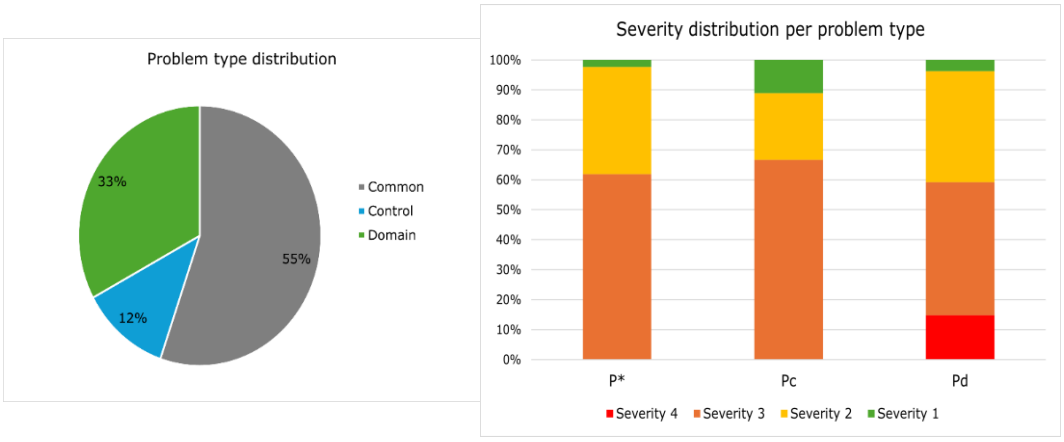


Figure 8. Distribution of problem types in HE and severity per problem type.

Out of all these identified issues, 42 problems were common (21 from each set), being detected using both sets, denoted as P^*. Also, 9 issues were found only by the control group, using Nielsen's set, denoted as P_C , while 25 issues were uniquely found by the domain-specific heuristics, denoted as P_D). Let $\mathbb{P}$, defined in Equation 5.3, be the

total set of problems found by applying both the control heuristics (P_C^*), defined in Equation 5.2, and the domain heuristics (P_D^*), defined in Equation 5.1.

$$P_D^* = P_D \cup P^* = 51 \tag{5.1}$$

$$P_C^* = P_C \cup P^* = 67 \tag{5.2}$$

$$\mathbb{P} = P_D \cup P_C \cup P^* = P_D^* \cup P_C^* = 71 \tag{5.3}$$

To complement the cross-comparison, the HEs results were examined using quality indicators as follows:

• Unique Problem Rate (Φ_P), counting the problems identified exclusively by the domain-specific set compared to the control. The metric is defined in Equation 5.4 and shows which group of heuristics found the most unique problems. If $\Phi_P > 1$, it means that the domain heuristics revealed more issues.

$$\Phi_P = P_D/P_C = P_D^*/P_C^* = 2.78 \tag{5.4}$$

• Problem Dispersion Rate (δ_P), measuring how evenly issues are distributed across heuristics; to avoid overloaded or redundant heuristics and it is defined in Equation 5.5. The distribution of issues in both groups of heuristics is measured considering the values noted δ_C and δ_D, representing the standard deviation of problem distribution for the control and domain heuristics. The standard deviation for a set of issues is defined in Equation 5.6, where n is the total number of heuristics, x_i the number of problems found using heuristic i, and μ_k mean number of problems per heuristic. If $\delta_P > 1$, issues distribution in the domain heuristics are better than in the control set.

$$\delta_P = \delta_C/\delta_D = 1.08 \tag{5.5}$$

$$\delta_k = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_k)^2} \tag{5.6}$$

• Problem Severity Rate (λ_P), assessing if the domain-specific heuristics lead to more severe issues on average. The metric is defined in Equation 5.7 where λ_D represents the average severity of the issues resulted using the domain set and λ_C represents the average severity of the problems resulted using the control set. In Equation 5.8 the average severity is calculated, where S_{k_i} severity score of problem i from set k and n is total number of problems considered (for that heuristic set). A value of $\lambda_P > 1$ shows that on average, the domain set led to more severe problems than the control set.

$$\lambda_P = \lambda_D/\lambda_C = 1.06 \tag{5.7}$$

$$\lambda_k = \frac{1}{n}\sum_{i=1}^{n} S_{k_i} \tag{5.8}$$

• Problem Specificity Rate (ε_P), measuring how domain-specific the detected problems are, via weighted scoring based on the FSI for each heuristic. The metric is computed using Equation 5.9, where ε_D represents the average specificity of the domain heuristics and ε_C represents the average specificity of the control set. In Equation 5.10 ε_C is detailed, representing the simple specificity average for each problem. In Equation 5.11 ε_D corresponding to the weighted sum of the number of problems in each heuristic, multiplied by the FSI assigned to it. If $\varepsilon_P > 1$ the problems found by the domain heuristics are on average more specific than the ones found by the control heuristics.

$$\varepsilon_P = \varepsilon_D/\varepsilon_C = 2.58 \tag{5.9}$$

$$\varepsilon_D = \frac{\sum_{H_i}(|P \in H_i^D| * FSI_i)}{|P_D^*|} = 0.77 \tag{5.10}$$

$$\varepsilon_P = \frac{\sum_{i=1}^{n} \varepsilon(P_i)}{|P_C^*|} = 2.58 \tag{5.11}$$

The quality indicator analysis provided by the Prometheus methodology in Stages 7 and 8 underscores the domain relevance of using REM-specific heuristics, compared to generic heuristics such as Nielsen's. Metrics such as the Unique Problem Ratio (Equation 5.4) with a value of $\Phi_P = 2.78$ and Problem Specificity Rate (Equation 5.9) with a value of $\varepsilon_P = 2.58$ proved to be the most valuable indicators. The obtained values illustrate that the domain set was able to detect more unique and domain-specific issues than the control set.

The Rate of Severity (Equation 5.7) with a value of $\lambda_P = 1.06$ and the Rate of Problem Dispersion (Equation 5.8) with a value of $\delta_P = 1.08$ suggest that only slight differences were observed between the sets, in terms of issues average severity and dispersion across the heuristics.

Analyzing the quality indicators, the proposed domain-specific heuristics confirm to be efficient not only to detect most of the general usability issues, but to detect high-impact, situational usability issues, strongly tied to RE workflows, providing a more targeted HE.

Although Nielsen's heuristics are still effective in detecting foundational usability issues, the proposed REM-specific set provides deeper and actionable insights into complex REM environments. A potential conclusion would be that a complementary approach while assessing REM tools, combining both frameworks, would ensure the detection of general issues while catching domain-specific challenges.

Given the criteria offered by the Prometheus methodology on the analysis of quality indicators, no refinements of the current REM-specific heuristics are necessary. Considering that minor adjustments would be needed in terms of severity and problem dispersion balance, these are not critical, and no refinement will be realised as part of the current thesis.

6. Exploratory Study: Heuristic Validation through LLMs

The exploratory study was designed to evaluate the capability of an LLM (ChatGPT-5) to perform heuristic evaluations of a REM tool using both generic (Nielsen) and the proposed REM-specific heuristics.

In this case, the evaluation scenario is described as follows: two separate evaluation sessions were conducted, each corresponding to a heuristic set, applied to the same six structured tasks previously used in human-based evaluations. The LLM was prompted using a structured approach that combines role prompting, chain-of-thought reasoning, and predefined output formatting, ensuring consistent and comparable results [26].Each task was independently assessed using prompts containing the evaluation role, objectives, reasoning steps, heuristic sets, task scenarios, workspace context, and relevant user-interface screenshots. The evaluation environment reproduced the human testing setup, using the same REM tool (DNG), identical tasks, and comparable interaction scenarios, although limited to static screenshots rather than real interaction.

Each Task was prompted individually, together with the standardised template used, which was formulated as follows:

„**Role**: Assume you are a usability expert evaluating an automotive Requirement Engineering and Management tool (IBM Doors Next Generation) used in the automotive industry, via Heuristic Evaluation.

Objective: The task is to identify usability issues according to the following heuristic set, rate their severity with values from 1-4 and suggest improvements.

Reasoning approach: Before providing your responses, examine the interface and think of the interaction step-by-step:

1.Observe every given screenshot to understand the entire context.

2.Identify the user interface elements, potential workflows, and terminology used.

3.The observations should be compared against the heuristic set given, in order to identify possible heuristic violations.

4.For every possible issue, think about the reason why it is a problem and how it affects usability.

5.Assign a severity rating for the identified issue.

6.Think of a potential improvement that can address the issue.

Heuristic Set: Nielsen / REM-specific.

Task description: Task 1 … Task 6.

Workspace context: A dedicated Test workspace was created in Requirements Engineering and Management tool for this Heuristic evaluation. The workspace contains three requirement modules, containing few requirement artefacts.

Screenshots: inserted relevant screenshots for the current task.

Output format: Heuristic violated | Issue Description | Severity (1-4) | Suggested improvement"

The study was conducted under controlled conditions, using single-run evaluations per task without regeneration, reflecting realistic LLM usage and preserving output consistency for analysis.

The LLM-based HE using Nielsen heuristics uncovered a total of 37 usability issues, and proposed REM-specific heuristics uncovered a total of 30 usability issues. The distribution of the reported problems is presented in Figure 9, and consists of: common with issues identified by human evaluators, new issues not logged by human evaluators, and potential or unverified issues.

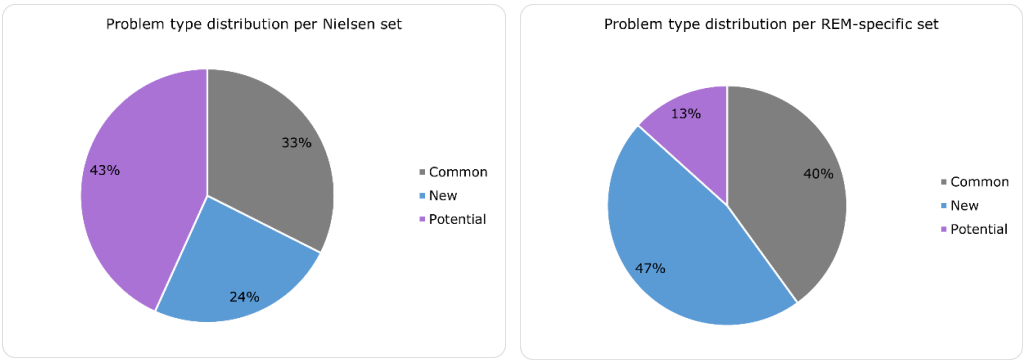


Figure 9. Problem type distribution identified LLM-based HE.

As observed in the distribution of the problem in Figure 9, the lower proportion, 13%, of unverified / potential issues compared to the Nielsen-based evaluation may suggest that domain-specific heuristics may offer a more contextual framework for LLM-based assessments, reducing the hallucination of hypothetical findings.

In terms of severity, issues identified using the Nielsen set were rated with severity 2 and 3, 41% of issues were perceived major and 59% were perceived minor, with no issues considered cosmetic or catastrophic. The severity distribution across the entire heuristic set is presented in Figure 10.

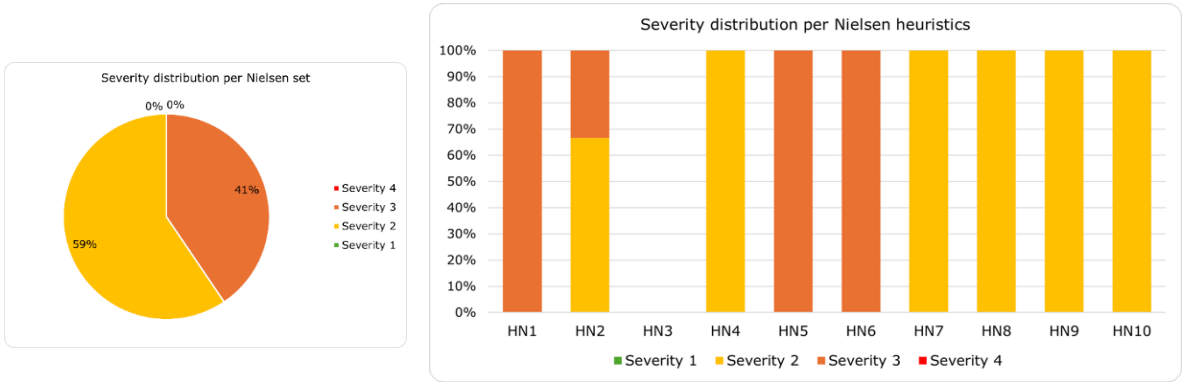


Figure 10. Severity distribution per LLM-based HE using Nielsen.

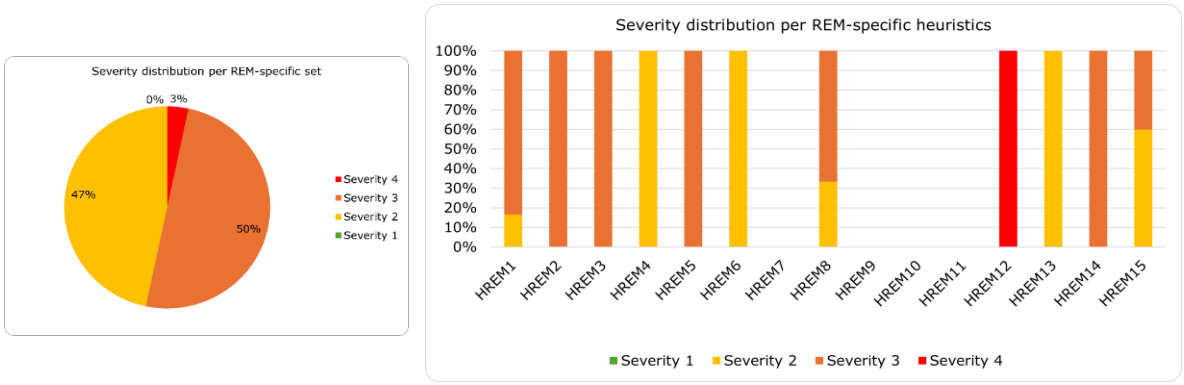


Figure 11. Severity distribution per LLM-based HE using REM-specific.

In terms of severity, issues identified using the REM-specific set were rated with severities from 2 to 4. Of 30 issues, 50% were rated as major, 47% minor and 3% catastrophic. The severity distribution for the entire set of REM-specific heuristics is presented in Figure 11.

This exploratory study revealed that LLMs such as ChatGPT-5 can identify a significant number of issues using both heuristic sets, with a substantial overlap with human evaluator issues.

When prompted with generic heuristics, the model detected surface-level issues related to feedback mechanisms, terminology, and UI clarity. Meanwhile, when prompted with the REM-specific heuristics, the model was able to identify more contextual and workflow-related issues, indicating that contextual guidance improves LLM reliability and relevance of the issues.

LLM-based heuristic evaluations demonstrate limitations due to the lack of real interactions, leading to model hallucinations and hypothetical findings.

In general, LLMs could be a powerful auxiliary instrument to be used in heuristic evaluations, capable of generating additional insights and improvement suggestions. On another note, due to their limitations, LLMs cannot replace human expertise, whose interaction-based evaluation is essential.

7. Conclusions and Contributions

The research for my thesis was done as part of the Communications Department, Faculty of Electronics, Telecommunications and Information Technologies of the Polytechnic University of Timisoara. The research contributes to the doctoral domain of Information Technology, intersecting the domains of SW Engineering, with a special focus on RE and HCI [17], [39]. The thesis addresses the usability evaluation of REM dedicated SW platforms and the complex SW engineering environments needed to support standardised RE activities.

The thesis extends generic HCI principles to the case of REM SW platforms. More specifically, my thesis can be summarised as a HE framework proposal tailored for REM Tools used in the automotive domain. The proposed REM-specific heuristic set was validated while demonstrating a more targeted and contextual foundation of usability principles for HEs of the SW specialised platforms.

In recent years, the topic of domain-specific heuristics has gained increasing attention in the HCI domain. Although Ben Shneiderman's Eight Golden Rules were first published in 1986 [40], Nielsen's heuristics in 1994 [32], the topic of developing the heuristic framework remained interesting during recent years. During the last decade, several methodologies were intended to facilitate a systematic development of domain heuristics. Several well-known and widely applied methodologies are the Prometheus methodology, proposed by Jimenez et al in 2017 [37], and the methodology proposed by Rusu et al. in 2011 [41]. These methodologies have been proven to be successfully applied to develop domain heuristics for Virtual Learning Environments [36], Virtual Reality applications [42], Grid Computing [43] or Interactive Digital TV [34] applications.

In recent years a special interest was observed in exploring LLMs-based HEs, for example, LLMs are used in studies as heuristic evaluators, demonstrating the ability to act as cognitive assistants, capable of evaluating usability, based on given heuristics [26]. However, LLMs are shown to be an auxiliary instrument to support qualitative research, improve early-stage evaluations, and improve human competence.

The novel domain heuristics that I proposed were validated through two HE sessions, one using a generic set of heuristics (Nielsen's) and one using the proposed set specific to REM. Ten evaluators were involved and their analysis was conducted based on their experience with DNG, the chosen REM tool. Analysing the quality metrics computed obtained for the results, the domain heuristics proved to be more efficient in identifying both general usability issues (usually overlapping with the ones detected while using generic heuristics) and contextual usability problems tied to specialised RE workflows and activities, essential to be performed in automotive RE.

In addition, I decided to explore the use of AI tools in the field of HCI. After validating the proposed domain heuristics, I conducted an exploratory analysis, using LLMs as potential heuristic evaluators [44]. Two HEs were conducted using ChatGPT-5 as the selected LLM, where the model was prompted with a standard template, containing heuristic sets, the same tasks used in the human-based evaluations and static screenshots. LLMs-based HEs demonstrate both potential and real limitations of LLM performance. In this case, LLMs proved to be useful augmentative instruments when contextually prompted, but were unable to replace human-based evaluations. Based on my research findings, I attempted to answer the research questions raised at the beginning of the thesis.

1. ***How do the HCI considerations reflect in the assessment of REM tools*?**

The intersection of HCI and REM tools is presented in Chapter 3. HCI considerations are relevant for REM tools, but have not been explored sufficiently. Specific for REM tools, HCI considerations refer to usability, navigation, feedback, error prevention, cognitive load, but have a stronger view on collaboration support and traceability visualisation. All of these concepts influence the AE during certain development phases. With the shift in engineering due to AI, the consideration of HCI in REM tools is becoming more critical.

2. ***Are generic heuristics effective in identifying domain-specific usability issues related to specific workflows of automotive REM tools?***

The results of the HE on DNG, an automotive REM tool, using generic and domain heuristics, are presented and interpreted in Chapter 5. Generic heuristics are effective in identifying more general-purpose issues than contextual usability issues.

3. ***Compared to generic heuristics, how effective are domain-specific heuristics, tailored to REM tools, in identifying usability issues?***

A cross-comparison of the results from both sets is presented in Chapter 5. Domain-specific heuristics detect most of the issues detected by the generic set and, in addition, capture deeper, more functional, and workflow-integrated aspects overlooked by generic heuristics. The methodology in which I obtained the domain specific heuristics is presented in Chapter 4.

4. ***How accurately can a LLM perform HEs of REM tools, in terms of issue detection, severity and overall coverage?***

The LLM-based evaluation results are presented in Chapter 6. LLMs demonstrate potential as heuristic evaluators, being able to provide accurate issues and actionable suggested improvements, but it also provides a significant number of hallucinated outputs due to missing interactions with the tool, only from a static perspective, using screenshots. Although LLMs can be a useful instrument in HE, human expertise is required to obtain verified and more accurate results.

In addition to the theoretical validation of the proposed REM-specific heuristics, the findings introduced by the thesis may have significant implications for practice in automotive industry. REM tools are part of the critical digital infrastructure used for safety critical systems development, complying with automotive standards such as

ASPICE, ISO 26262, ISO 29148. REM tools are designed to provide mandatory structured workflows in the AE domain, as specified in ISO 29148. For this kind of engineering environment, usability deficiencies not only negatively impact the perception of diverse users, but introduce potential sources of engineering risks. Inefficient filtering mechanisms, poor navigation workflows, or other inconsistent representations can induce missing traceability links, late impact analysis, or misinterpreted requirements. In this context, good usability in REM tools influences process maturity levels, compliance, and engineering reliability.

Apart from the industrial relevance, the thesis contributes to the doctoral domain of Electronics, Telecommunication and Information Technologies, by positioning HCI specific fields of UI and UX into the context of SW engineering systems used in the automotive domain. REM tools are SW based informational systems that support the development of embedded and safety-critical systems, where SW, HW, ME and system-level requirements intersect. The research highlights the importance of aligning HCI principles with engineering workflows, with compliance guidelines for evaluating software development processes (ASPICE), by extending domain-specific heuristics within the REM context. In addition, the exploratory study of AI-based usability evaluation bridges the RE domain and usability research with advancing directions in AI, applied to engineering environments.

In conclusion, this thesis connects the fields of SW engineering and HCI, intersecting usability evaluation with RE workflows, and studying how AI can impact the evaluation process in complex engineering environments. The research findings indicate that usability in REM tools should be mandatory as an engineering quality factor, affecting domain workflows, compliance assurance, and collaborative efficiency within safety relevant development environments. As the automotive domain evolves and an increase in interconnected and SW architectures is observed, the need for domain-specific evaluation frameworks becomes more prominent.

The theoretical contributions of this thesis are as follows:

1. A theoretical framework of RE domain within the Automotive industry.
2. A functional classification of REM tools, from the RE workflows perspective, resulting in a synthesis of the core functionalities of REM tools.
3. Conceptualisation of the importance of the HCI domain for SW-based systems with respect to REM tools.
4. Introduction of the domain-specific heuristic concept in the context of automotive REM tools.
5. Pioneering research on the use of LLMs, within the domains of RE usability aspect.

The thesis brings several practical contributions:

1. The proposal of REM-Specific Usability Heuristics
2. Application of Prometheus methodology to develop the REM-specific usability heuristics.
3. The design and application of a comparative HE framework for automotive REM tools, using generic and domain-specific heuristics.
4. The HE and interpretation of quantitative and qualitative evaluation metrics, to validate the proposed domain-specific heuristics.
5. Exploratory research on LLM-based heuristic evaluation of REM tools.

Bibliography

[1] M. Vogel *et al.*, “Metrics in automotive software development: A systematic literature review,” *J. Softw. Evol. Process*, vol. 33, no. 2, p. e2296, 2021, doi: 10.1002/smr.2296.
[2] E. Tüzün, B. Tekinerdogan, Y. Macit, and K. İnce, “Adopting integrated application lifecycle management within a large-scale software company: An action research approach,” *J. Syst. Softw.*, vol. 149, pp. 63–82, Mar. 2019, doi: 10.1016/j.jss.2018.11.021.
[3] *ISO/IEC/IEEE 29148:2018 - Systems and software engineering — Life cycle processes — Requirements engineering*, 2, 2018. Accessed: Aug. 23, 2025. [Online]. Available: https://www.iso.org/standard/72089.html
[4] “ISO/IEC/IEEE 15288:2023(en), Systems and software engineering — System life cycle processes.” Accessed: May 05, 2026. [Online]. Available: https://www.iso.org/obp/ui/en/#iso:std:iso-iec-ieee:15288:ed-2:v1:en
[5] “ISO/IEC/IEEE 12207:2026(en), Systems and software engineering — Software life cycle processes.” Accessed: May 05, 2026. [Online]. Available: https://www.iso.org/obp/ui/en/#iso:std:iso-iec-ieee:12207:ed-2:v1:en
[6] ® 2025 Siemens Product Lifecycle Management Software Inc, “Requirements Management, Requirements Gathering, Requirements Management tools - Polarion REQUIREMENTS.” Accessed: Sep. 03, 2025. [Online]. Available: https://polarion.plm.automation.siemens.com/products/polarion-requirements
[7] “Jama Connect.” Accessed: May 13, 2025. [Online]. Available: https://www.jamasoftware.com/platform/jama-connect/
[8] S. SAS, “IBM Engineering Requirements Management DOORS Next.” Accessed: May 20, 2025. [Online]. Available: https://www.sodiuswillert.com/en/ibm-elm/ibm-engineering-requirements-management-doors-next
[9] “ISO 9241-11:2018(en), Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts.” Accessed: Aug. 21, 2025. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en
[10] J. Nielsen, “Usability 101: Introduction to Usability,” Nielsen Norman Group. Accessed: Oct. 08, 2025. [Online]. Available: https://www.nngroup.com/articles/usability-101-introduction-to-usability/
[11] M. Aubin Le Quéré *et al.*, “LLMs as Research Tools: Applications and Evaluations in HCI Data Work,” in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, in CHI EA ’24. New York, NY, USA: Association for Computing Machinery, May 2024, pp. 1–7. doi: 10.1145/3613905.3636301.
[12] C. Jimenez, H. Allende Cid, and I. Figueroa, “PROMETHEUS: Procedural Methodology For Developing Heuristics Of Usability,” *IEEE Lat. Am. Trans.*, vol. 15, no. 3, pp. 541–549, Mar. 2017, doi: 10.1109/TLA.2017.7867606.
[13] J. C. Bebel, B. L. Raskob, A. C. Parker, and D. J. Bebel, “Managing complexity in an autonomous vehicle,” *IEEE Aerosp. Electron. Syst. Mag.*, vol. 23, no. 3, pp. 3--13, 2008.
[14] “V-Model - an overview | ScienceDirect Topics.” Accessed: Aug. 21, 2025. [Online]. Available: https://www.sciencedirect.com/topics/engineering/v-model
[15] S. Pargaonkar, “A Comprehensive Research Analysis of Software Development Life Cycle (SDLC) Agile & Waterfall Model Advantages, Disadvantages, and Application Suitability in Software Quality Engineering,” *Int. J. Sci. Res. Publ.*, vol. 13, no. 8, pp. 120--124, Aug. 2023.
[16] K. Berger, “Adoption of agile methods in automotive software development.” Accessed: Jan. 19, 2025. [Online]. Available: https://essay.utwente.nl/81865/
[17] F. Flores, M. Mora, F. Alvarez, R. O’Connor, and J. Macias-Luevano, “Requirements Engineering: A Review of Processes and Techniques,” *Handb. Res. Mod. Syst. Anal. Des. Technol. Appl.*, pp. 90--104, 2009, doi: 10.4018/978-1-59904-887-1.ch006.
[18] O. Gotel *et al.*, “Traceability Fundamentals,” in *Software and Systems Traceability*, 2012, pp. 3–22. doi: 10.1007/978-1-4471-2239-5_1.
[19] *ISO 26262-8:2018(en), Road vehicles — Functional safety — Part 8: Supporting processes*, 4. Accessed: May 15, 2025. [Online]. Available: https://www.iso.org/standard/68386.html
[20] M. Knop, “An ASPICE Overview.” Accessed: May 04, 2025. [Online]. Available: https://www.suse.com/c/an-aspice-overview/
[21] *ISO/SAE 21434:2021 - Road vehicles — Cybersecurity engineering*, 2021. Accessed: May 19, 2025. [Online]. Available: https://www.iso.org/standard/70918.html
[22] J. M. Carrillo de Gea, J. Nicolás, J. L. Fernández Alemán, A. Toval, C. Ebert, and A. Vizcaíno, “Requirements engineering tools: Capabilities, survey and assessment,” *Inf. Softw. Technol.*, vol. 54, no. 10, pp. 1142–1157, Oct. 2012, doi: 10.1016/j.infsof.2012.04.005.
[23] S. SAS, “IBM Engineering Requirements Management DOORS - SodiusWillert.” Accessed: May 20, 2025. [Online]. Available: https://www.sodiuswillert.com/en/ibm-elm/ibm-doors
[24] Eccam, “HW/SW Requirements Management Tool on Git | ReqView.” Accessed: May 13, 2025. [Online]. Available: https://www.reqview.com/
[25] M. A. K. Raiaan *et al.*, “A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges,” *IEEE Access*, vol. 12, pp. 26839–26874, 2024, doi: 10.1109/ACCESS.2024.3365742.
[26] P. Sahoo, A. K. Singh, S. Saha, V. Jain, S. Mondal, and A. Chadha, “A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications,” Mar. 16, 2025, *arXiv*: arXiv:2402.07927. doi: 10.48550/arXiv.2402.07927.
[27] M. Körber, A. Eichinger, K. Bengler, and C. Olaverri Monreal, *User experience evaluation in an automotive context*. 2013, p. 18. doi: 10.1109/IVS.2013.6629440.
[28] H. Gaspard-Boulinc and S. Conversy, “Usability Requirements for Requirement Engineering Tools,” Oct. 2014, doi: 10.1145/2670444.2670458.

[29] O. Rotaru, S. Vert, and R. Vasiu, “Switching from DOORS to DOORS Next Generation. Identifying the Main Usability Issues. A Case Study,” in *Proceedings of Ninth International Congress on Information and Communication Technology*, X.-S. Yang, S. Sherratt, N. Dey, and A. Joshi, Eds., Singapore: Springer Nature, 2024, pp. 261–276. doi: 10.1007/978-981-97-3289-0_22.
[30] O. Rotaru, S. Vert, and R. Vasiu, “Usability Evaluation of Requirement Collaboration Features in Requirements Management Tools,” presented at the Proceedings of the 27th International Conference on Enterprise Information Systems (ICEIS 2025), Rome, Italy: SCITEPRESS – Science and Technology Publications., Aug. 2025, pp. 201–208. doi: 10.5220/0013281800003929.
[31] O. Rotaru, C. Orhei, and R. Vasiu, “Analyzing Agile role-specific needs for REM tools : A Survey-Based Study,” in *International Congress on Information and Communication Technology*, London: Springer Nature Singapore, p. to be published.
[32] J. Nielsen, “10 Usability Heuristics for User Interface Design,” Nielsen Norman Group. Accessed: May 28, 2025. [Online]. Available: https://www.nngroup.com/articles/ten-usability-heuristics/
[33] J. Gerhardt-Powals, “Cognitive engineering principles for enhancing human-computer performance,” *Int. J. Human–Computer Interact.*, vol. 8, no. 2, pp. 189–211, Apr. 1996, doi: 10.1080/10447319609526147.
[34] A. Solano, C. Rusu, C. A. Collazos, S. Roncagliolo, J. L. Arciniegas, and V. Rusu, “Usability heuristics for interactive digital television,” in *The Third International Conference on Advances in Future Internet*, 2011.
[35] J. Zhang, T. R. Johnson, V. L. Patel, D. L. Paige, and T. Kubose, “Using usability heuristics to evaluate patient safety of medical devices,” *J. Biomed. Inform.*, vol. 36, no. 1, pp. 23–30, Feb. 2003, doi: 10.1016/S1532-0464(03)00060-1.
[36] I. Figueroa, C. Jiménez, H. Allende-Cid, and P. Leger, “Developing usability heuristics with PROMETHEUS: A case study in virtual learning environments,” *Comput. Stand. Interfaces*, vol. 65, Apr. 2019, doi: 10.1016/j.csi.2019.03.003.
[37] C. Jimenez, H. Allende-Cid, and I. Figueroa, “PROMETHEUS: PROcedural METhodology for developing HEuristics of USability,” Feb. 27, 2018, *arXiv*: arXiv:1802.10121. doi: 10.48550/arXiv.1802.10121.
[38] OpenAI, “GPT-5 System Card,” OpenAI, Aug. 2025. Accessed: Aug. 24, 2025. [Online]. Available: https://cdn.openai.com/gpt-5-system-card.pdf
[39] A. R. Mathew, A. Al Hajj, and A. Al Abri, “Human-Computer Interaction (HCI): An overview,” in *2011 IEEE International Conference on Computer Science and Automation Engineering*, Jun. 2011, pp. 99–100. doi: 10.1109/CSAE.2011.5953178.
[40] IxDF - Interaction Design Foundation, “What is User Centered Design (UCD)?,” IxDF - Interaction Design Foundation. Accessed: May 27, 2025. [Online]. Available: https://www.interaction-design.org/literature/topics/user-centered-design
[41] C. Rusu, S. Roncagliolo, V. Rusu, and C. Collazos, “A Methodology to establish usability heuristics,” Jan. 2011.
[42] A. Mohammad and L. Pedersen, “Analyzing the Use of Heuristics in a Virtual Reality Learning Context: A Literature Review,” *Informatics*, vol. 9, no. 3, p. 51, Sep. 2022, doi: 10.3390/informatics9030051.
[43] D. Quiñones, C. Rusu, S. Roncagliolo, V. Rusu, and C. A. Collazos, “Developing Usability Heuristics for Grid Computing Applications: Lessons Learned,” in *Information Technology: New Generations*, S. Latifi, Ed., Cham: Springer International Publishing, 2016, pp. 485–495. doi: 10.1007/978-3-319-32467-8_43.
[44] G. Guerino, L. Rodrigues, B. Capeleti, R. Ferreira, A. Freire, and L. Zaina, *Can GPT-4o Evaluate Usability Like Human Experts? A Comparative Study on Issue Identification in Heuristic Evaluation*. 2025. doi: 10.48550/arXiv.2506.16345.