University of Siegen, Germany
Faculty III, Wirtschaftswissenschaften, Wirtschaftsinformatik, Wirtschaftsrecht

German Aerospace Center, Germany
HSI, Human Factors for Software Engineering, Institute of Software technology

Master thesis

MSc. Human Computer Interaction

Prof. Dr. Markus Rohde

Prof. Dr. Shadan Sadeghian

Dr. Ing. Sabine Theis

Designing and Evaluating a Conversational ELN Interaction Concept under Simulated AI Conditions: A Wizard-of-Oz Study in Scientific Research Contexts

Name:	Ahire, Tanishq Tushar	**Matriculation number:**	1785236
E-mail:	tanishq.ahire@student.uni-siegen.de		
Address:	Holsterhauser Str. 34 45147, Essen	**Study programme:**	MSc. Human Computer Interac- tion
		Semester:	05
Phone:	+49 151 44935581	**Deadline:**	16.02.2026

Abstract

Electronic Laboratory Notebooks (ELN) have become central infrastructures for scientific documentation, offering structured data storage, traceability, and regulatory compliance. However, their prevailing interaction paradigm remains predominantly form-based and keyboard-centered, which may conflict with laboratory realities such as sterility constraints, multitasking, and distributed collaboration. This thesis investigates how a conversational interaction concept for ELN can be designed and evaluated under simulated AI conditions, with a primary focus on interaction structure rather than algorithmic performance.

Grounded in Human–Computer Interaction (HCI) theory and informed by Computer-Supported Collaborative Learning (CSCL), the work conceptualizes the ELN as a socio-technical artifact that mediates cognition, documentation practices, and collaborative continuity. Design foundations were derived from literature-informed considerations, insights from a DLR user requirements survey, and contextual observations from Kadi4Mat and POLiS laboratory environments. These inputs were synthesized into structural design drivers, including multimodal interaction, in-situ documentation support, human-in-the-loop control, and transparency in AI-mediated workflows.

A Wizard-of-Oz (WoZ) study was conducted to evaluate the conversational ELN prototype under simulated AI conditions. Voice-based documentation and chatbot-style retrieval were manually simulated to isolate interaction effects from technical AI limitations. A baseline condition using traditional documentation was compared with the prototype. Quantitative measures included the System Usability Scale (SUS), NASA-TLX workload assessment, and a Trust in Automation scale, complemented by qualitative interviews. Due to the exploratory design and limited sample size, quantitative results were analyzed descriptively.

Findings suggest that conversational interaction structures appear promising for supporting perceived workflow fluidity and contextual retrieval when technical performance is assumed optimal. At the same time, issues of predictability, transparency, and epistemic responsibility emerged as critical design considerations. The thesis contributes to HCI research by empirically examining conversational interaction paradigms within high-stakes scientific documentation contexts using a methodologically controlled simulation approach.

I. Table of contents

II. List of figures

III. List of tables

V. List of abbreviations

Abbreviation	**Explanation**
LN	Laboratory Notebook
ELN	Electronic Laboratory Notebook
RDM	Research Data Management
GMP	Good Manufacturing Practice
AI	Artificial Intelligence
NLP	Natural Language Processing

1 Introduction

The idea of having an intelligent assistant to support us in everyday activities—both at home and at work—has long been fascinating. This fascination is particularly evident in laboratory environments, where documenting experiments, guiding complex procedures, and reflecting on next steps place a significant cognitive and practical burden on researchers. Scientists and storytellers alike have imagined futures in which such challenges are eased through intelligent support systems. Science fiction has frequently depicted this vision in the form of futuristic laboratories equipped with voice assistants. One iconic example is J.A.R.V.I.S. from the Iron Man films, an artificial voice assistant that supports Tony Stark not only by managing large amounts of data, but also by providing intelligent, context-aware assistance during highly complex tasks.

In this thesis, such science-fiction narratives do not serve as a basis for scientific claims or evaluation. Instead, they act as a source of inspiration for envisioning interaction possibilities and informing the design of the proposed prototype. The research itself is grounded in established theories and methods from Human–Computer Interaction (HCI) and Computer-Supported Collaborative Learning (CSCL), as well as empirical insights from real laboratory practices. Against this background, the central question is not whether systems like J.A.R.V.I.S. can be fully realized, but how selected elements of intelligent assistance—such as voice interaction and conversational support—can be meaningfully designed and integrated into Electronic Laboratory Notebooks (ELN), which form the backbone of modern research documentation and collaboration.

A laboratory notebook (LN) is a fundamental or a primary tool used by researchers to document the research. In (Schnell, 2015) it is defined as researchers use a laboratory notebook to document their hypotheses, experimental procedures, measurement activities, observations, analyses, and interpretations in a structured and chronological manner. Unlike scientific publications, which present finalized and curated results, laboratory notebooks capture the ongoing history of scientific work, including detailed procedural steps, equipment configurations, calibration settings, environmental conditions, unexpected deviations, and contextual influences that may affect experimental outcomes. The notebook also serves as an organizational tool, a memory aid that guides the scientist in their workflows, it also acts as an intellectual property that comes from the research (Schnell, 2015). Since centuries these records were kept on paper, from Archimedes's parchment notes to Darwin's field journals. The traditional paper notebook not only documents experiment for reproducibility and patent evidence but also embodies a lab's intellectual legacy. However, paper is limited – it is static, and not easily searchable, it makes it cumbersome to share or duplicate. As the digital data generation exploded, the "mere idea of putting everything down in a bound paper notebook" began to seem absurd. By the early 2000s, Research and Development (R&D) organizations started actively replacing paper notebooks with ELN (Machina & Wild, 2013)

Illustration 1: [A traditional lab notebook]

Source: [("Lab Notebooks | JILA - Exploring the Frontiers of Physics", 2026c)]

An ELN is broadly defined as an information system to create, store, retrieve and share fully electronic lab records in compliance with legal and scientific requirements(Analytical Science Article DO Series, 2026). Early ELN began as simple systems that were generic and used systems like word document, excel and notepads to take digital notes, as they progressed, they adopted to software and then later to web-based platforms and as they progressed, they tried to mimic the familiar notebook format while using computing, networking, and cloud technologies. This evolution sought to ensure that the digital transition was as comfortable and intuitive as possible, preserving and ideally enhancing the scientist's creative and analytical processes. Over the past two decades, ELN have matured significantly and become widely adopted across academic and industrial research settings. Contemporary systems such as Kadi4Mat, eLabFTW, and Benchling offer structured digital documentation, cloud-based collaboration, version control, and compliance-oriented features tailored to modern laboratory environments. While Benchling has begun incorporating AI-supported functionalities into its ecosystem, platforms such as Kadi4Mat and eLabFTW primarily emphasize robust data management, interoperability, and structured research documentation within collaborative infrastructures.

Although these systems represent important progress toward more connected and digitally enabled laboratories, their core functionality largely centers on data storage, workflow coordination, and, in some cases, isolated AI features such as transcription or query-based assistance. A gap therefore remains in the systematic design and empirical evaluation of AI-driven voice and conversational systems as deeply integrated, context-aware collaborators within ELN—particularly when examined through the lenses of Human–Computer Interaction (HCI) and Computer-Supported Collaborative Learning (CSCL). This thesis builds upon existing ELN infrastructures but extends them conceptually and methodologically by investigating how voice assistants and conversational chatbots can be intentionally designed and evaluated to support collaborative learning, research continuity, and cognitive assistance within real laboratory workflows.

This idea of innovating laboratory environments with intelligent systems is not new. Vannevar Bush, in his famous essay (Bush) imagined a future where scientists used advanced devices to manage and document information effortlessly a vision that this thesis aims to achieve.

1.1 Benefits of ELN in Scientific Research

ELN solves a lot of the shortcomings that regular notebooks possess, this makes ELN quite useful in scientific and research environments. Researchers in laboratories can easily record, organize, and retrieve data for their projects and experiments. Data inputs, such text, photos, instrument readings, and so on, are stored in a way that makes them easy to read and search, indexed, and back up digitally. For instance, keyword or metadata search makes it easier to identify any experiment that used a specific material or process. With traditional LNs, one has to flip pages back and forth to find something specific. This helps scientists avoid doing the repetitive tasks and focus on the research and other important tasks. According to reports from Atrium Research in 2011 cited in (Machina & Wild, 2013), show an instant 20% accepted productivity gain over paper based (LNs) processes by enabling scientists to document experiments, find and reuse information, and collaborate more efficiently.

In an ELN data is time-stamped, versioned, and often immutable, which makes a secure safe audit trail as noted in (Guerrero et al., 2016). Digital records are clear and backed up, so there are no difficulties in identifying handwriting or causing any physical damage. This is not the case with paper pages that are hard to read or lost. Many ELN also have to comply with regulatory standards and requirements (such FDA 21 CFR Part 11) and support electronic signatures and limited access to fulfil good laboratory practice and legal standards. This makes sure that research records and experiments can be reproduced and are defensible.

Additionally, when we talk about the platform Kadi4Mat, it brings structure to data that is still evolving and not yet ready for publication (Warm Data). Instead of treating lab notes and datasets as separate entities, Kadi merges research data management (RDM) with ELN capabilities, allowing researchers to document context, workflows, and experimental parameters alongside this warm data, which strengthens traceability and FAIR (Findable, Accessible, Interoperable, Reusable) compliance ("Kadi4Mat", 2025). This ability to automate tasks using APIs, Python scripting, and CLI tools makes research more reproducible and less error-prone, especially in distributed lab environments where collaboration depends on shared understanding and structured documentation ("Kadi4Mat", 2025). The integration of AI and machine learning modules further supports data-driven decision-making, analysis, and model training within the same research lifecycle, meaning not having to rely on separate external tools as much. ("Kadi4Mat", 2025). Kadi hence allows a lot of flexibility within itself, and hence different laborites are able to tailor and adopt it to their workflows.

1.2 Thesis Outline

This thesis follows a structured progression from problem identification to design, evaluation, and reflection. It begins by establishing why conversational interaction in Electronic Laboratory Notebook (ELN) is worth investigating and how this topic connects to broader developments in Human–Computer Interaction (HCI) and collaborative scientific work.

Chapter 2 reviews the relevant literature and existing systems. It examines the current state of ELN, highlighting both their strengths and their limitations—particularly in terms of interaction design. The chapter also introduces key theoretical perspectives from HCI and Computer-Supported Collaborative Learning (CSCL), which frame the ELN not merely as a storage tool but as a cognitive and collaborative artifact. This analysis leads to a clearly defined problem statement and research objective.

Chapter 3 presents the design of the conversational ELN prototype. Drawing from literature insights, a DLR user requirements survey, and contextual observations from real laboratory environments (Kadi4Mat and POLiS), the chapter explains how recurring themes were translated into concrete design decisions. These decisions shaped a multimodal prototype that combines structured interfaces with voice-based documentation and chatbot-style interaction.

Chapter 4 describes the Wizard-of-Oz study conducted to evaluate the interaction concept under simulated AI conditions. It explains the study design, tasks, participant characteristics, and data collection methods. Quantitative and qualitative findings are then presented descriptively.

Chapter 5 discusses the results with respect to usability, workload, multimodality, trust, and collaborative continuity. Chapter 6 outlines limitations of the study and methodological constraints. Finally, Chapter 7 summarizes the main contributions and identifies directions for future research, particularly regarding real AI integration and long-term deployment in laboratory settings.

.

2 Related Work

This chapter summarizes the relevant related work and present state of the art to reach the research target, as previously described in Chapter 1.1. Initially concentrating on the evolution of Electronic Laboratory Notebooks (ELN) and their contemporary application in scientific research settings. The chapter then talks about the main benefits of ELN, including how they help with scientific documentation, working together, and sharing in-formation.

After that, the current limitations of ELN systems are addressed that make it difficult for the scientists in laboratories environments and may disrupt their workflows. The chapter next looks at how AI may help current ELN in response to these problems. Alongside this, important considerations from human–AI interaction research are discussed, with a focus on usability, transparency, and trust in AI-supported scientific workflows.

Finally, the chapter shares what has been reported from a user survey study at the German Aerospace Centre (DLR). While these observations may not conform with standard academic literature, they provide a valuable practice-oriented perspective on how researchers currently document their work and collaborate in real laboratory settings. This related work directly help ground the discussion in real-world use and directly inform the user-centered design decisions described in the following methodology chapter.

2.1 Current State of the Art in ELN's

Despite the increasingly digital nature of society there are some areas of research that remain firmly rooted in the past; in this case, in particular (Kanza et al., 2017) describes the laboratory notebook, as the last remaining paper component of a experimental practice. While many researchers still rely on general-purpose digital tools like Microsoft Word, Excel, or OneNote for scientific note-taking due to their accessibility and shareability, (Guerrero et al., 2019) notes that dedicated ELN are increasingly replacing traditional handwritten lab records across scientific disciplines. ELN offer organized, searchable, and interoperable data formats which satisfy the needs of current research environments that are data-heavy, collaborative, and interdisciplinary.

There is a growing shift toward viewing digital research tools not only as data repositories but as interactive intelligent systems. (Yagamurthy, 2023) shows that advances in artificial intelligence particularly in natural language processing, speech recognition, and predictive modeling enable more conversational and assistive forms of human–computer interaction. These developments lay important groundwork for reimagining ELN as active research assistants rather than passive record-keeping systems. With these AI-powered methods, it opens opportunities to support assistant in real time, operate hands-free in sterile areas, and automatically document data when we talk in context of ELN.

For instance, scientists at the German Aerospace Centre (DLR) Institute of Software Technology are working on the Kadi4Mat ELN, which is also used at the POLiS Cluster of Excellence, KIT. Where they are employing Kadi4Mat as the main RDM platform ("Kadi4Mat", 2025). Kadi4Mat is a research data application that aims to combine an electronic laboratory notebook ELN and a repository. The goal of this application is to facilitate structured data storage and exchange and to document data analysis to publish data. The application aims to support researchers throughout their research process as described by (Meinecke, Heidrich, Dworatzyk, & Theis).

Illustration 2: [Screenshot Dashboard for Kadi4Mat ELN**]**

The figure shows the standard Kadi4Mat dashboard used at the German Aerospace Center (DLR) and KIT. It illustrates project navigation, experiment records, and repository integration. This screenshot provides contextual grounding for the development of the AI-enhanced interaction concept proposed in this thesis. Source: Screenshot, https://kadi.iam.kit.edu

When we talk about other projects, they are taking further steps toward intelligent ELN. For instance, In the paper (van Eikeren, 2004) describes one of the earliest examples of an intelligent ELN for a pharmaceutical process development. In this project the system supported an intelligent planning, execution, and analysis through rule-based domain logic, that enabled automated checks for process safety, equipment compatibility, and workflow completeness. It allowed and enabled reuse of protocol modules, automatic scaling adjustments, and design-of-experiments (DOE)–based optimization. Which represented an early form of computational decision support. It also had the ability to connect and integrate with laboratory instruments and internal databases, which helped enabled data acquisition and structured record management directly from the instruments. These capabilities anticipate key features of modern AI-enabled ELN and lab environments.

(Guerrero et al., 2016) Evaluated a tablet-based and wearable-supported ELN workflow, in this paper the author demonstrated how mobile devices and Apple Watch integration with OneNote can extend ELN beyond traditional desktop use. Their work highlights how image capture, protocol access, timers, and mobile synchronization can support real-time laboratory work. It also suggests that wearable technologies can expand the scope of data collection and contextual logging. However, this system was not AI-driven and only established a multi-modal interaction foundation which is relevant for future intelligent ELN.

Similarly, (Gates, McLean, & Osborn, 2015) at the National Institute of Standards and Technology (NIST) proposed a "Smart Electronic Laboratory Notebook (SELN)" although it is not a fully AI driven system, this concept was designed to support portable, secure, and context-aware scientific work in laboratory environments. Here they emphasized on mobility, connectivity, and integration with digital services. They positioned the ELN as a portable personal hub for organizing, analyzing, and research data on the go, accessing anytime and anywhere. They also out-lined future capabilities like voice assistants, sensor integration, and a wearable-linked contextual monitoring, thereby establishing a foundation for more intelligent and adaptive ELN for the futures.

Beyond research prototypes in academia, commercial platforms such as Benchling, According to the ("2026 Biotech AI Report | Benchling", 2026a), Benchling already embed domain-specific AI into laboratory workflows for molecular biology. It provides integrated tools for DNA/RNA sequence design-alignment, primer-cloning workflow support, and reagent-sample tracking. Demonstrating how computational assistance is increasingly embedded directly within the researcher's primary working interface rather than limited to backend analytics. These capabilities suggest an emerging trajectory for ELN towards a more intelligent, predictive, and error-aware systems, with the potential to support real-time decision-making and documentation in domain-specific laboratory environments.

As laboratory research becomes increasingly data-driven, machine learning and deep learning techniques are being explored to support chemical analysis and reaction modeling. For example

(Menke, Homberg, & Koch, 2023) presents the interactive deep-learning workflows for cheminformatics tasks and demonstrating how AI can be applied to molecular data processing, reaction-related learning exercises, and computational chemistry pipelines. Their work highlights the growing role of AI in laboratory experiments and scientific workflows in general. These developments suggest a broader trajectory in which digital laboratory tools including future ELN could evolve beyond just simple record-keeping books and move towards a more active analytical and decision-support roles.

Collectively, these advancements signal a paradigm shift: from static digital notebooks to interactive, intelligent systems capable of supporting researchers through real-time assistance, automation, and smart data analysis. As AI capabilities continue to improve, ELN are positioned not only to enhance productivity and reproducibility, but to fundamentally reshape the human–machine relationship in scientific inquiry. The future of ELN lies in systems that not only document but also interpret, recommend, and evolve alongside the researcher’s work.

Table 1: [Comparative Overview of Current ELN and Their Limitations]

System	Strengths	AI Integration	Interaction Modalities	Limitations
Kadi4Mat	ELN + RDM integration, FAIR compliance, structured metadata	No AI integration (currently focused on RDM and documentation)	Desktop interface	No conversational interface, no AI-assisted documentation, limited hands-free interaction
Benchling	Integrated molecular biology workflows, sequence design, domain-specific computational tools	Domain-specific AI modules (e.g., sequence analysis support)	Desktop interface	AI is domain-focused and backend-oriented; limited conversational interaction; limited transparency of AI reasoning
eLabFTW	Open-source ELN, experiment tracking, compliance and audit trails	No built-in AI features	Desktop interface	Primarily documentation-oriented; no intelligent assistance or conversational capabilities
SciNote	Structured experiment templates, team collaboration features, compliance support	No integrated AI assistance	Desktop interface	Focused on structured documentation; limited automation; no conversational or voice interaction
Traditional ELN (general category)	Searchable documentation, version control, traceability	None	Desktop interface	Limited real-time workflow support; minimal multimodal interaction

The comparison above demonstrates that current ELN platforms—such as Kadi4Mat, Benchling, eLabFTW, and SciNote—primarily prioritize structured documentation, compliance, and data management. Their core value lies in traceability, metadata handling, and reproducibility support. However, across these systems, the dominant interaction paradigm remains form-based, menu-driven, and keyboard-centered.

Kadi4Mat, see illustration 2 above, for example, provides strong integration between ELN functionality and research data management, particularly in relation to FAIR principles. Nevertheless, it currently does not include AI-assisted interaction, conversational querying, or real-time cognitive support features. Its strength lies in structured documentation and repository integration rather than dynamic interaction during ongoing laboratory activity.

Benchling represents a more advanced integration of computational tools, particularly in molecular biology contexts. While it incorporates domain-specific AI modules for sequence analysis and related workflows, these features are largely backend-oriented and domain-constrained. They do not fundamentally transform the interaction layer of ELN into a conversational or dialogue-based system. Moreover, the reasoning processes of such AI modules are not always transparent to users, raising questions regarding explainability and trust calibration.

Open-source platforms such as eLabFTW and structured workflow systems like SciNote emphasize documentation standardization, compliance, and team-based record keeping. However, they do not provide integrated AI assistance, predictive support, or multimodal interaction capabilities. Their interaction logic remains primarily transactional rather than conversational.

Across these systems, three overarching gaps become visible:

1. Interaction Gap: Current ELN are optimized for structured storage and retrieval rather than fluid, real-time interaction. Conversational querying, voice-based documentation, and interactive summarization are not systematically implemented or empirically evaluated.
2. Cognitive Support Gap: While computational analytics may exist in domain-specific contexts (e.g., Benchling), few systems explicitly conceptualize ELN as collaborative cognitive artifacts that scaffold documentation, reflection, and shared understanding during experiments.

3. Human–AI Interaction Gap: Even where AI components are present, limited attention is given to transparency, explainability, user control, and trust formation within everyday laboratory workflows. Empirical studies examining how researchers experience AI-like interaction within ELN remain scarce.

Consequently, although technical AI capabilities are increasingly embedded within research infrastructures, there is a lack of systematic investigation into how conversational ELN interaction paradigms influence documentation behavior, collaboration dynamics, and trust development. This indicates a clear research gap: empirical, interaction-focused evaluation of conversational ELN—independent of technical AI performance—remains underexplored.

2.2 Gaps of current ELN Systems

Despite their benefits, current ELN solutions have notable limitations that can hinder its use in certain workflows and environments.

LNs are often desktop based, which may be impractical in settings like for example biosafety level 3/4 labs, cleanrooms, or GMP production suites where computers are hard to sterilize or using a keyboard isn't always feasible to enter data. (Guerrero et al., 2016) highlight that such interaction styles are not always well suited for experimental contexts in which researchers are gloved, mobile, or actively handling materials. Their evaluation also shows that only a small number of ELN systems supported alternative input modalities, such as tablets, smart pens, or wearable devices, which could enable more flexible and hands-free documentation. This limited availability of multi modal interaction options constrains real-time note-taking during experiments and underscores the need for ELN interfaces that better accommodate mobile and touch-free interaction in laboratory environments (Guerrero et al., 2016).

Another limitation that users of ELN face is entering data, be it in our daily life entering data for other purposes, or it be the researchers in labs who face this tiresome task of entering data into ELN. (Barchard & Pace, 2011) also mentions that this can be labor-intensive, human data entry is error prone that ruin statistical results and conclusion. Many current ELN function as digital notebooks but require manual copying of data from machines or other software. This not only costs time but also introduces opportunities for errors. Ideally, an ELN should pull raw data directly from balances, sequencers, or other instruments. In practice, integrating a wide variety of lab instruments with an ELN is technically challenging and often not fully realized. (Machina & Wild, 2013) reports a survey of organizations which identified that 41.9%

of firms saw integration with other laboratory systems as the most challenging area of ELN adoption.

2.3 Benefits of AI-Enabled ELN

In this thesis, the term AI-enabled ELN refers to the concept that, in order to assist users with documentation, information retrieval, or contextual support, an AI component can support and augment these tasks within the ELN. Importantly, AI-enabled does not imply full autonomy; rather, it denotes decision-support and assistive features that operate under human oversight and preserve user control within the laboratory workflow.

One practical benefit of AI-enabled ELN lies in enabling hands-free interaction, which is particularly relevant in laboratory contexts where researchers work under sterility constraints or with occupied hands. AI-enabled ELN can externalize memory and documentation tasks, allowing cognition to be distributed across the human–AI system rather than remaining solely with the individual researcher, this can be done by having a voice-based input and output interaction. From a cognitive load perspective, (Leung, Shimabukuro, & Collins, 10132024) show that touchless and hands-free interaction can reduce cognitive load and support more immersive software use, particularly in sterile or safety-critical environments. In sterile or safety-critical laboratory environments, this reduction is particularly significant. Voice-driven documentation enables observations to be captured in situ, reducing the need for later recall and transcription. In HCI, conversational and voice interfaces (VUIs) are designed to align with principles of natural interaction, as they leverage familiar communicative practices rather than requiring adaptation to manual typing. When implemented with domain-appropriate language, structured prompts, and predictable behavior, such AI-supported voice interfaces can support uninterrupted workflows while maintaining user trust and control.

Beyond cognitive load reduction through voice interaction, AI-enabled ELN can also support researchers through what can be described as digital scaffolding. Digital scaffolding, in the context of AI-enabled ELN, refers to system-provided, context-sensitive guidance that supports users in completing complex documentation or experimental tasks without taking control away from them. Drawing on the educational concept of scaffolding—where temporary support structures help learners accomplish tasks they could not yet perform independently—digital scaffolding within ELN can take the form of structured prompts, automatic detection of missing experimental parameters, suggestions of previously validated protocols, reminders about required controls, or contextual links to related prior experiments. Such “smart suggestions” are not arbitrary recommendations, but are generated based on the current experimental entry, project metadata, or organizational knowledge stored in the system. In this way, the ELN does not merely store information but actively supports the researcher's reasoning process by highlighting relevant information at the moment of need.

From a learning-theoretical perspective, these structured prompts can be interpreted through the concept of the Zone of Proximal Development (ZPD), as originally put forward by Vygotsky, (ResearchGate, 2026), as they offer guidance that is neither trivial nor overwhelming but calibrated to the researcher's immediate task. In a limited, task-specific manner, the AI component provides contextual support that helps researchers perform documentation tasks more effectively while retaining full control over experimental decisions. From a CSCL perspective, such features contribute to collaborative knowledge construction by embedding accumulated organizational knowledge—from previous experiments, validated protocols, and institutional rules—for example, prompting a researcher heating a chemical solution to record the exact temperature and duration because previous experiments showed that small deviations affected the outcome—and amplifying it into the current workflow. Rather than requiring researchers to independently search for relevant information, the ELN surfaces it at moments of need, thereby supporting learning-through-action and shared understanding.

Complementing these scaffolding functions, AI-enabled ELN may also incorporate real-time validation mechanisms. Within the framework of a laboratory notebook, the role of the AI component is not to decide autonomously but to prompt reflection and correction. For instance, rule-based checks or pattern matching with prior entries may generate warnings about missing controls or incompatible parameters. These alerts function as cognitive prompts, encouraging users to reassess documentation choices while retaining responsibility for final judgment. This aligns with HCI principles of human-in-the-loop design("Humans in the Loop: The Design of Interactive AI Systems | Stanford HAI", 2026b) and supporting trust in AI-assisted systems(Hoff & Bashir, 2015).

Another key benefit of AI-enabled ELN lies in improving documentation quality and error prevention through real-time feedback during experimentation. Instead of functioning solely as a passive repository, the ELN can provide immediate validation cues while documentation is being created. From the perspective of Distributed Cognition (Hollan, Hutchins, & Kirsh, 2000), this establishes a feedback loop in which cognitive responsibility is shared between the researcher and the digital system: structured checks and contextual alerts support the researcher's reasoning at the moment of action. This immediate feedback resembles principles of cognitive apprenticeship, where guidance occurs during task execution rather than after completion. In practical terms, if the system flags missing parameters, inconsistent values, or unusual deviations while an experiment is being documented, researchers can correct these issues instantly rather than discovering them retrospectively after results have been compromised. From a CSCL standpoint, such real-time feedback enhances collective research quality by ensuring that deviations and contextual details are explicitly recorded. This strengthens reproducibility and shared understanding, as collaborators can later interpret results with a transparent account of what occurred during execution. Over time, this contributes to higher-quality shared knowledge artifacts within the ELN.

A primary motivation for ELN adoption is the reduction of administrative burden, therefore reducing manual entry directly meets that goal. Automation made possible by an AI helps this objective more seriously by taking care of simple, repetitive chores like transcribing, computations, and handling metadata. Cognitive Load Theory (Sweller) supports that automating regular paperwork lowers unnecessary cognitive load, which frees up cognitive resources for higher-order thinking, interpretation, and problem-solving. This change lets researchers focus on scientific research instead than following rules. Automation also improves collaborative knowledge building in CSCL by making records that are more full, standardized, and organized. AI systems consistently record steps, parameters, and contextual metadata, making the final documentation more credible for future collaborators. This improves research continuity, particularly in long-term projects or teams with rotating members, where incomplete or inconsistent records are a major source of knowledge loss.

2.4 Considering Human-AI Interaction

Much of the current discourse on "AI in work systems" emphasizes general principles—user agency, transparency, and trust calibration—yet ELN represent a distinct context: documentation is not a one-off decision task but a continuous, distributed practice that spans time, collaborators, and experimental variability. Consequently, applying human–AI interaction principles to ELN requires translating broad guidance into ELN-specific interaction requirements: when assistance is offered, how it is justified, how it is logged, and how it supports collaborative accountability.

Human-centered AI guidelines argue that AI systems should support human goals, preserve user agency, and communicate limitations—particularly in high-stakes domains where errors are costly (Amershi et al.);(Cai et al.). In the context of ELN, "high stakes" does not only mean safety; it also includes reproducibility, traceability, and scientific defensibility. Therefore, the main question is not whether an AI feature exists, but how its interaction design changes documentation behavior: Does it reduce omissions? Does it prompt more complete context capture? Does it support later interpretation by someone else?

This reveals a critical design tension: ELN are expected to produce stable records, while AI outputs are often probabilistic and may change depending on prompts, context, or model updates. Without careful design, AI assistance can introduce ambiguity into records that are expected to be authoritative. As a result, ELN-relevant AI assistance must be bounded and accountable: assistance should be clearly identifiable as AI-generated, editable by users, and traceable (what was suggested, accepted, or overridden).

Workflow fit is not "nice to have" in ELN: Studies of expert work emphasize that tools are effective when they integrate into existing workflows and collaboration patterns rather than forcing users to adapt to technical constraints (Zhang, Muller, & Wang, 2020). For ELN, this point becomes particularly important because documentation occurs under time pressure, interruptions, and hands-busy conditions. In other words, even highly capable AI assistance can fail in practice if it increases interaction overhead or interrupts the "critical moments" of experimentation.

A gap in existing work is that many AI discussions remain abstract, for example AI should support the user, whereas ELN settings require concrete interaction answers: Where in the entry workflow does the system intervene? What is the cost of a correction prompt at that moment? How are prompts coordinated when multiple researchers contribute to the same experiment record? These questions shift the evaluation focus from "AI capability" to "interaction consequences," which is particularly relevant for this thesis.

Transparency and explainability must be calibrated to scientific work: Work on algorithmic transparency suggests that explanations must be calibrated: too little transparency can undermine trust, while too much detail can overload users and reduce confidence (Kizilcec). In ELN, this calibration problem is amplified because researchers may need different explanation depths depending on their role (e.g., a student documenting steps vs. a senior researcher reviewing decisions). Socially transparent AI work further argues that explanations should be grounded in organizational context and accountability needs (Ehsan, Liao, Muller, Riedl, & Weisz, 2021). For ELN, this implies that explanations should not only describe what is suggested, but also why it matters for the record: e.g., linking suggestions to documentation standards, project conventions, or prior experiments.

A key limitation in current systems and literature is that explainability is often treated as a generic interface add-on. For ELN, transparency should be operationalized as record-relevant support: clear provenance of AI contributions, simple rationales tied to documentation completeness or comparability, and mechanisms to inspect what the system used as context (e.g., "based on your last entry," "based on the selected template," "based on project metadata").

Human-in-the-loop design and calibrated trust in ELN workflows: Human–AI interaction research consistently highlights that AI should enhance human expertise rather than replace it and that opportunities to inspect and challenge outputs support appropriate reliance (Cai et al.). In ELN, this requirement is not merely ethical; it is practical. Scientific documentation often involves uncertainty, interpretation, and judgment that cannot be delegated to an automated system without risking loss of context or accountability.

Accordingly, ELN-relevant AI assistance should be designed as decision support and documentation support, not automated authority. Concretely, this means users can accept, edit, reject, or defer suggestions; the system should support "repair" actions (undo/redo, change tracking); and AI contributions should be distinguishable from user-authored content in the record. These interaction properties are directly tied to trust calibration: when the system is transparent about its role and users retain meaningful control, trust can develop in a more appropriate, stable manner.

Critical gap and implication for this thesis: Across the reviewed work, a major gap is that many principles are well-established at a general level, but there is limited empirical evidence on how these principles play out in ELN-specific conversational interaction paradigms, especially regarding (a) real-time documentation, (b) collaboration and research continuity, and (c) trust formation around AI-like assistance. Much of the existing work either focuses on technical AI capabilities or discusses human-centered AI at a high level without examining ELN interaction consequences.

Therefore, this thesis focuses on the interaction paradigm—not AI model performance—and evaluates how researchers perceive and engage with AI-like conversational assistance embedded in an ELN workflow. This includes analyzing where assistance is helpful versus disruptive, how transparency and control shape user experience, and what trust- and expectation-related dynamics emerge when the system behaves as an "idealized assistant."

2.5 Empirical Foundation

This chapter summarizes empirical insights from a large-scale user survey conducted at the German Aerospace Center (DLR). While not part of the academic literature, these findings provide an applied perspective on real-world ELN usage and are used to contextualize and ground the design requirements derived in Chapter 3.

A cross-sectional survey study (N=71), conducted by DLR German Aerospace Center, under the eLAB Project, distributed a lime survey questionnaire of 31 questions. In this survey the goal was to understand the current LNs usage and adapt ELN to user's needs. indicates that although pen-and-paper notebooks remain the most available documentation medium, researchers strongly prefer digital workflows, primarily on laptops and secondarily on tablets, demonstrating a need for ELN designs that support laptop-based productivity while bridging paper-based capture through reliable digitization of handwritten notes and sketches. Consequently, the prototype developed in this thesis embeds conversational assistance within a familiar ELN interface, aligning with RQ1 (Interaction and Documentation)

Second, offline functionality was rated as highly important (59.2% scoring 4–5), particularly in fieldwork, low-connectivity lab environments, and travel contexts. This highlights the requirement that conversational or AI-like support must not introduce additional fragility into documentation workflows. Instead, interaction mechanisms must remain robust, minimally disruptive, and compatible with asynchronous synchronization models. This requirement directly informs the evaluation of perceived workflow integration and usability in the WoZ study (RQ1).

Third, although 69.1% of respondents collaborate regularly or frequently, collaborative ELN platforms remain underused. This indicates a structural gap between documentation systems and collaborative practices. Derived requirement: an AI-like conversational ELN should support not only individual notetaking but also shared understanding, contextual linking between entries, and retrieval of prior experiment knowledge. This empirical insight grounds RQ2 (Collaboration and Research Continuity) by framing the ELN as a potential collaborative cognitive artifact rather than a purely individual documentation tool.

Fourth, data integrity emerged as the most critical requirement (81.7% rating high importance), emphasizing traceability, version control, timestamped histories, and legal defensibility. This finding constrains the design of AI-like assistance: any generated summaries, prompts, or suggestions must preserve transparency and editability to maintain record authority. This directly connects to RQ3 (Trust and Human–AI Interaction), as trust in conversational assistance is closely linked to perceived reliability and accountability.

Finally, while data security was rated as important (M = 4.04, SD = 1.114), respondents emphasized that security measures must not impair usability. This highlights a recurring tension between control mechanisms and workflow efficiency. For the present thesis, this implies that conversational assistance should reduce interaction friction rather than introduce additional complexity—an aspect explicitly examined in the WoZ evaluation when assessing user perceptions of effort, usefulness, and trust.

In summary, the DLR survey results are not merely contextual background but serve as an empirical grounding for the prototype's design principles: seamless integration into existing workflows, support for collaboration and research continuity, preservation of data integrity, and usability-aware trust calibration. These empirically derived requirements directly shape the conversational interaction paradigm evaluated in this thesis and substantiate the formulation of RQ1–RQ3.

2.6 Problem Statement

Although current Electronic Laboratory Notebooks (ELN) provide structured documentation, compliance support, and research data management capabilities, the current related work analysis above reveals a persistent interaction gap. Existing ELN primarily rely on form-based, keyboard-centered interfaces that are optimized for structured storage rather than real-time, fluid documentation during experimental activity. Conversational interaction, voice-based documentation, and context-sensitive cognitive support are rarely implemented and even less frequently evaluated empirically.

At the same time, laboratory work is characterized by time pressure, collaboration across roles, sterility constraints, and the need for precise, traceable documentation. Researchers must document parameters, contextual factors, and procedural deviations while simultaneously conducting experiments. Current ELN systems offer limited support for in-situ documentation, real-time validation, and collaborative knowledge continuity. Where AI components are present, they are often backend-oriented, domain-specific, or insufficiently transparent, with little empirical investigation into how such interaction paradigms influence documentation behavior, collaboration dynamics, or trust formation.

Consequently, a central problem emerges there is a lack of systematic, interaction-focused evaluation of conversational ELN paradigms that examines how AI-like assistive interaction—independent of actual AI model performance—affects documentation practices, collaborative sense-making, and user perceptions under realistic laboratory conditions.

2.7 Research Objective

The main objective of this thesis is to design an interactive GUI prototype of a conversational ELN interface and evaluate this interaction paradigm with regard to documentation practices, collaboration, and user perceptions in laboratory environments. Rather than evaluating actual AI system performance, this work focuses on how an AI-like conversational interaction concept—embedded within an ELN—shapes user experience under controlled, idealized conditions. The prototype simulates assistive functionality in order to examine how conversational documentation and retrieval mechanisms are perceived and integrated into everyday laboratory workflows.

Accordingly, this thesis investigates how conversational ELN interaction can support researchers in contexts where typing-based interaction is disruptive. Particular attention is given to hands-free voice interaction and chat-based dialogue interfaces as multimodal interaction concepts. These modalities are examined as complementary interface designs: voice interaction is explored as a means of supporting in-situ documentation, particularly in sterile or safety-critical contexts, while a text-based conversational interface supports structured retrieval of prior experiments and research continuity. The focus is therefore on the design and evaluation of multimodal interaction mechanisms rather than on implementing autonomous AI capabilities.

Furthermore, since laboratory environments are collaborative workplaces, this thesis conceptualizes the conversational ELN interface as a collaborative cognitive artifact from a Computer-

Supported Collaborative Learning (CSCL) perspective. The emphasis lies on how the interaction paradigm may influence shared understanding, documentation completeness, perceived cognitive support, and group awareness. Importantly, the study does not measure trust in a deployed AI system; instead, it examines how participants interpret and respond to an AI-like conversational interface under Wizard-of-Oz conditions. The interaction concept is developed using a user-centered design approach and evaluated through a Wizard-of-Oz study in which system responses are simulated to represent an idealized assistive interaction.

Based on the comparative analysis of current ELN, the identified interaction gaps, and the human–AI interaction considerations discussed above, this thesis derives its research focus from the following gap: there is limited empirical evaluation of conversational ELN interaction paradigms that investigates documentation behavior, collaboration dynamics, and user perceptions independently from actual AI model performance.

To address this gap, and in direct alignment with the objective of evaluating the conversational ELN interaction paradigm with regard to documentation practices, collaboration, and user perceptions under simulated conditions, the following research questions are formulated. Each research question corresponds to one core dimension derived from the problem analysis and research objective.

1. RQ1 – Interaction and Documentation
   Derived from the identified interaction gap in current ELN and the objective of examining documentation practices, RQ1 focuses on how the conversational interface affects everyday documentation behavior. How do researchers perceive and experience a conversational ELN interaction paradigm for documentation and information retrieval?
2. RQ2 – Collaboration and Research Continuity (CSCL)
   Building on the cognitive support gap and the CSCL perspective introduced above, RQ2 examines whether and how the interaction paradigm influences collaborative sense-making and continuity across experiments. How does a conversational ELN interaction paradigm influence perceived support for shared understanding and research continuity in collaborative laboratory settings?
3. RQ3 – Perception and Trust Calibration of AI-like Interaction
   Based on the human–AI interaction considerations and the identified transparency and trust gap, RQ3 investigates how users interpret and calibrate their trust toward an AI-

like conversational interface under idealized Wizard-of-Oz conditions. How do participants interpret and calibrate their trust toward an AI-like conversational interface when interacting with an ELN under simulated (WoZ) conditions?

3 Prototype Design

This chapter presents the design of the prototype artifact that forms the conceptual and technical foundation of the Wizard-of-Oz (WoZ) evaluation described in Chapter 4. It consolidates and reframes previously conducted user-centered design (UCD) activities as structured design grounding rather than as independent empirical studies. The purpose of this chapter is not to report evaluative findings, test hypotheses, or present statistical results. Instead, it provides a transparent and theoretically grounded account of how conceptual insights, contextual understanding, and practice-oriented evidence informed the development of the interaction concept and its implementation.

The prototype operationalizes a conversational interaction paradigm for an AI-assisted Electronic Laboratory Notebook (ELN). Laboratory documentation practices are embedded in complex socio-technical environments shaped by sterility constraints, time pressure, safety regulations, distributed teamwork, institutional standards, and heterogeneous digital competencies. Scientific documentation is not merely a technical activity; it is intertwined with sense-making, coordination, accountability, and long-term knowledge preservation. Consequently, the design focus extends beyond functional enhancement toward meaningful workflow integration, usability, transparency, traceability, and sustained support for collaborative continuity over time.

By making the design rationale explicit, this chapter establishes traceability between theoretical foundations, empirical design inputs, derived requirements, and the concrete interaction mechanisms implemented in the prototype. The chapter proceeds from design foundations (Chapter 3.1) to cross-source triangulation and requirement synthesis (Chapter 3.2), and finally to the operationalization of these requirements within the interaction concept and the WoZ-ready technical setup (Chapters 3.3 and 3.4).

3.1 Design Foundations

The interaction concept was grounded in three complementary design inputs: (1) literature-informed considerations from ELN, HCI, CSCL, and Human–AI Interaction research; (2) insights derived from an existing DLR user requirements survey; and (3) contextual insights gained through participatory observation at a Kadi4Mat community meeting and a guided visit to the POLiS laboratory. These inputs functioned as structured design foundations rather than as standalone empirical investigations.

Together, they provided conceptual, empirical, and contextual grounding for the prototype design. The following subchaptersdescribe the analytical procedures applied within each source and clarify how they contributed to the subsequent triangulation process.

3.1.1 Literature-Informed Design Considerations

The objective of the literature analysis was not to conduct a formal systematic review, but to identify recurring conceptual tensions and design challenges relevant to AI-assisted documentation systems in scientific environments. The analysis focused on publications addressing Electronic Laboratory Notebooks, Human–Computer Interaction, Computer-Supported Collaborative Learning, Distributed Cognition, Cognitive Load Theory, and Human–AI Interaction.

Rather than summarizing findings descriptively, the literature was examined through a thematic synthesis oriented toward requirement abstraction. The analysis aimed to extract cross-cutting tensions and design-relevant problem patterns recurring across sources. Themes were retained when they appeared across multiple independent publications or were strongly grounded in established theoretical frameworks.

Recurring conceptual patterns included constraints and opportunities for real-time, in-situ documentation; gaps in research continuity and limitations in reconstructing past experimental decisions; cognitive load implications of digital documentation practices; integration challenges between ELN and laboratory instruments; and requirements for transparency, accountability, and calibrated trust in AI-supported systems. These themes were abstracted into higher-level design tensions, such as flexibility versus traceable integrity and automation versus user control, which informed the triangulation process described in Chapter 3.2.

3.1.2 Insights from the DLR User Requirements Survey

A central empirical basis for the design grounding was an existing user requirements survey conducted at the German Aerospace Center (DLR) within the context of the eLab project (N = 71). The survey captured researchers' documentation practices, collaboration patterns, and expectations regarding ELN systems across multiple research domains.

The dataset comprised closed-ended Likert-scale items assessing importance ratings of ELN features (e.g., data integrity, security, offline functionality, and collaboration support), multiple-choice questions regarding currently used documentation tools and communication channels, and open-ended responses describing challenges, workarounds, and unmet needs.

The survey was conducted within the DLR eLab project, in which the author was involved as a project member. However, the survey design, data collection, and primary statistical analysis were completed prior to and independently of the present thesis focus. For this thesis, the dataset was treated as secondary data used for structured design grounding.

The descriptive statistical analysis was originally conducted by the DLR eLab project team and is documented in the official survey report. This thesis relied on reported descriptive statistics, including means, standard deviations, percentage distributions, and frequency counts. No inferential statistical testing was conducted within the scope of this thesis.

Key findings relevant for design grounding included the centrality of data integrity (81.7% rating 4–5; 53.5% assigning the maximum score), strong concern for data security ($M = 4.04/5$, $SD = 1.114$), provided that usability is not impaired; appreciation of offline functionality (59.2% rating 4–5) with a preference for automatic synchronization after reconnection (65.2%); high collaboration frequency (69.1%) alongside limited use of collaborative ELN features; and a strong preference for digital workflows despite continued reliance on paper.

These findings were interpreted as indicators of structural tensions, including integrity versus flexibility, security versus usability, and collaboration practice versus tool support. Open-text responses were examined deductively and aligned with literature-derived categories to ensure conceptual coherence and analytical consistency.

3.1.3 Contextual Insights from Kadi4Mat and POLiS

Contextual insights were gathered through participatory observation at the Kadi4Mat Community Meeting (2025), and a guided laboratory visit to the POLiS (Post Lithium Storage) Lab.

Due to security and confidentiality requirements, no audio recordings, photographs, or videos were permitted during the laboratory visit. An official photograph taken by DLR and POLiS representatives for organizational purposes was not used as research data. Structured field notes were documented immediately after each session to preserve contextual detail while respecting institutional constraints.

Observations focused on workflow sequencing, documentation timing, coordination patterns, sterility constraints, onboarding experiences, instrument interaction, and the practical realities of switching between experimental execution and documentation tasks.

Illustration 3: [POLiS Lab visit with DLR in Ulm, Germany.]

The photograph documents the author's visit to the POLiS (Post Lithium Storage) laboratory at the German Aerospace Center (DLR) in Ulm, Germany. Source: DLR/POliS lab official photograph, 2025("Post | Feed | LinkedIn", 2026d)

Field notes were analyzed using reflexive thematic analysis with an inductive orientation. Themes were retained when supported by multiple observations and when consistent with literature or survey findings. Recurring contextual themes included documentation frequently occurring after experiments rather than during them; ELN perceived as static archives rather than active workflow tools; parallel documentation practices outside formal ELN systems; tension between flexibility and institutional structure; concerns regarding AI correctness and responsibility; onboarding and learnability challenges; and cognitive switching burdens during experimental work.

These contextual patterns reinforced and nuanced themes identified in literature and survey findings, thereby providing situational grounding for requirement synthesis.

3.2 Cross-Source Thematic Convergence and Requirement Synthesis

The derivation of design requirements followed a structured triangulation logic. Rather than translating each source independently into features, themes were first identified within each design foundation (Chapter 3.1) and then systematically compared across sources. Only themes that demonstrated cross-source convergence were elevated to structural design drivers.

This process consisted of three analytical steps: (1) within-source theme extraction, (2) cross-source comparison to assess convergence strength, and (3) abstraction of convergent themes into structural design drivers. Themes appearing in at least two independent sources were considered analytically robust. Themes supported by all three sources were classified as high-convergence themes and prioritized as primary design drivers.

3.2.1 Convergence Overview

The table below summarizes the results of the cross-source comparison. It does not present statistical measurements but rather an analytical mapping of whether a theme was independently identified in literature, survey data, and/or contextual observation. The Table 2 below, operationalizes the triangulation logic and documents the analytical pathway from design foundations to structural themes.

Table 2: [Cross-Source Thematic Convergence Matrix]

Theme	Literature	Survey	Observation	Convergence
In-situ vs. post-hoc documentation	Yes	No	Yes	Medium
Collaboration–tool gap	Yes	Yes	Yes	High
Data integrity centrality	Yes	Yes	Yes	High
Security–usability tension	Yes	Yes	Yes	High
Cognitive switching burden	Yes	No	Yes	Medium
ELN as static archive	Yes	Yes	Yes	High
AI trust and accountability	Yes	Yes	Yes	High
Onboarding friction	Yes	No	Yes	Medium
Offline robustness	No	Yes	Yes	Medium

High-convergence themes formed the structural backbone of the interaction concept because they reflect systemic tensions consistently reported across conceptual, empirical, and contextual sources. Medium-convergence themes informed secondary refinements or implementation constraints but were not treated as central architectural drivers. The convergence matrix therefore functions as a traceability instrument that prevents arbitrary feature selection and demonstrates how design priorities were analytically justified.

The following chapter is structured around the triangulated structural themes that emerged from the cross-source analysis described above. These headlines do not represent predefined categories. Rather, themes were first identified inductively within each design foundation and subsequently abstracted deductively through cross-source comparison. Only themes demonstrating analytical convergence across at least two independent sources were elevated to structural design drivers. The chapter headings below therefore represent empirically and conceptually grounded themes that organize the design decisions of the prototype.

3.3 Structural Themes as Design Drivers and Operationalization

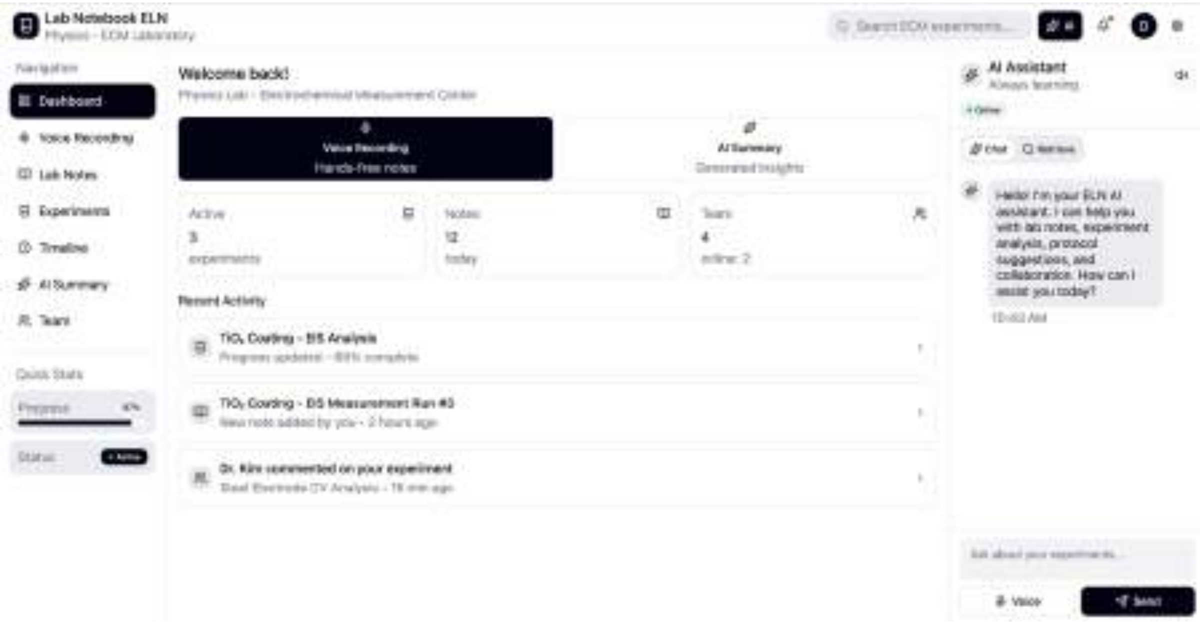
The triangulated themes were abstracted into structural design drivers. These drivers represent recurring tensions within scientific documentation practice and define the conceptual space within which design decisions were made. Rather than being treated as isolated usability issues, they were interpreted as systemic contradictions within socio-technical laboratory environments.

3.3.1 In-Situ Documentation vs. Post-Hoc Reconstruction

Across literature and observation, documentation was frequently delayed until after experimental execution. Time pressure, sterility constraints, and physical workload often made real-time note-taking impractical. Post-hoc reconstruction increased cognitive burden and the risk of contextual loss.

The design implication was to support immediate, low-interruption documentation mechanisms. This was operationalized through voice-based logging integrated directly into the active experiment view, allowing structured capture without workflow interruption.

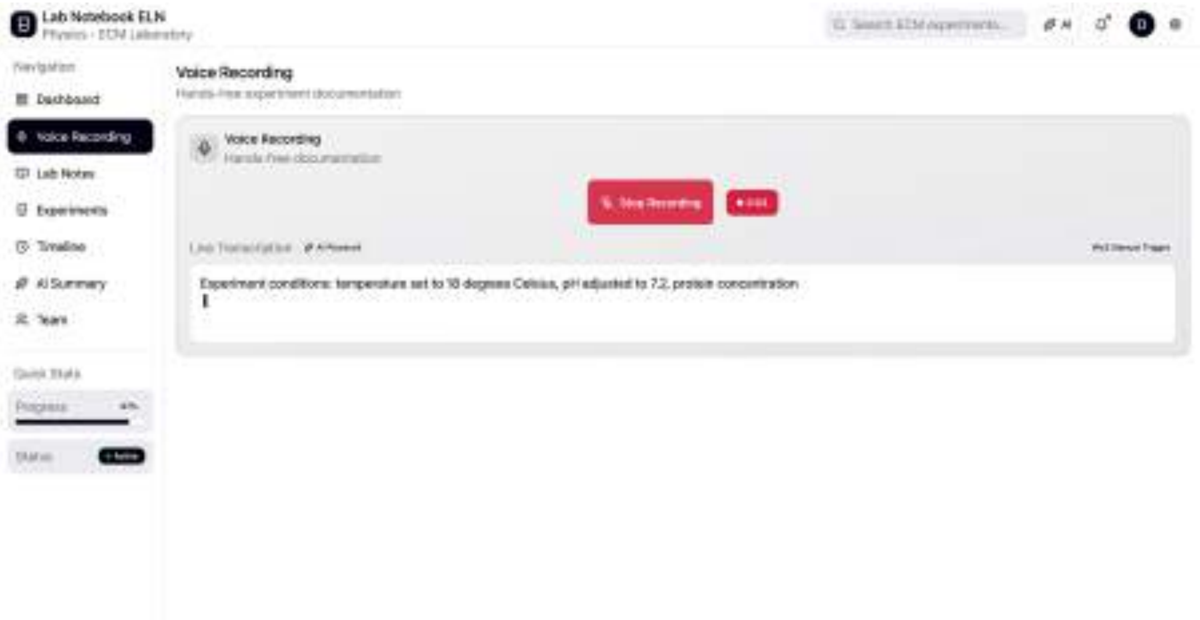
Illustration 4: [Dashboard screen for Prototype Concept]



The dashboard interface of the prototype, showing the left navigation panel and right the chatbot interface active.
Source: Own screenshot

This requirement is concretely instantiated in two primary interface components: the Dashboard Screen (see Illustration 4) and the Voice Recording Screen (see Illustration 5). The Dashboard serves as the central entry point of the ELN and prominently positions direct access to voice recording and AI summary functions in the main interaction area. This placement minimizes navigation depth and reduces friction during time-critical experimental moments. The Voice Recording Screen enables hands-free documentation and immediately binds transcribed speech to the active experiment context. By structurally linking spoken input to experiment metadata, timestamps, and structured note fields, the system reduces reliance on memory-based post-hoc reconstruction and directly addresses sterility constraints, occupied hands, and cognitive load.

Illustration 5: [Voice Recording screen from the Prototype Concept]



This screenshot presents the voice interaction screen used in the Wizard-of-Oz evaluation. The interface includes a visual listening indicator, real-time transcription field, and confirmation controls. The design aims to simulate hands-free documentation in sterile laboratory environments. Source: Own screenshot

3.3.2 Collaboration Practice vs. ELN as Static Archive

Survey findings revealed high collaboration frequency but limited use of collaborative ELN functionality. Observations and informal discussions indicated that researchers often rely on direct communication rather than consulting documentation. ELN were frequently perceived as static archives rather than active coordination tools.

The design implication was to transform the ELN into an interactive knowledge interface rather than a passive repository. This was operationalized through conversational retrieval, timeline visualization, experiment-linked summaries, and cross-entry querying through the chatbot interface.

Concretely, this structural transformation is reflected in the Experiments Screen (see Illustration 7), the Lab Notes Screen (see Illustration 6) the Timeline Screen (see Illustration 7), and the AI Chatbot Interface (see Illustration 9). The Experiments Screen structures documentation at the experiment level, aligning with how scientific work is conceptually organized. The Lab Notes Screen displays individual entries together with collaborators, timestamps, and tags, reinforcing accountability and shared authorship. The Timeline Screen visualizes chronological development and contribution traces, enabling retrospective reconstruction of decision pathways.

Illustration 6: [lab Notes screen from the Prototype Concept**]**

This is the screenshot of the Lab notes section in the ELN, it gives access to different notes within the ELN and thereby able to view edit and add notes.. Source: Own screenshot

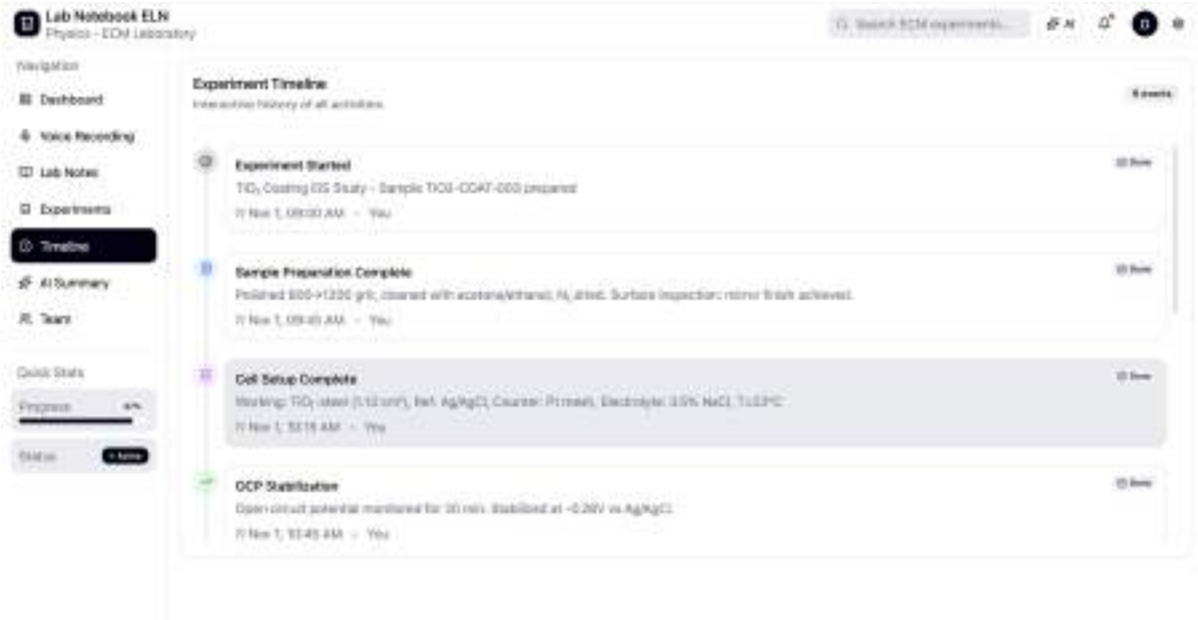
The chatbot further reduces manual search effort by allowing natural language queries across entries, thereby repositioning the ELN as an interactive coordination system rather than a static archive.

3.3.3 Integrity, Traceability, and Epistemic Responsibility

Data integrity emerged as the highest-priority survey requirement and was reinforced by literature emphasizing provenance and reproducibility. At the same time, observational insights revealed frustration with rigid or overly bureaucratic systems.

The design implication was to combine structured traceability with editable, user-controlled content. Operationalization included timestamped entries, visible versioning logic, AI-generated summaries that remain editable, and explicit attribution of AI involvement.

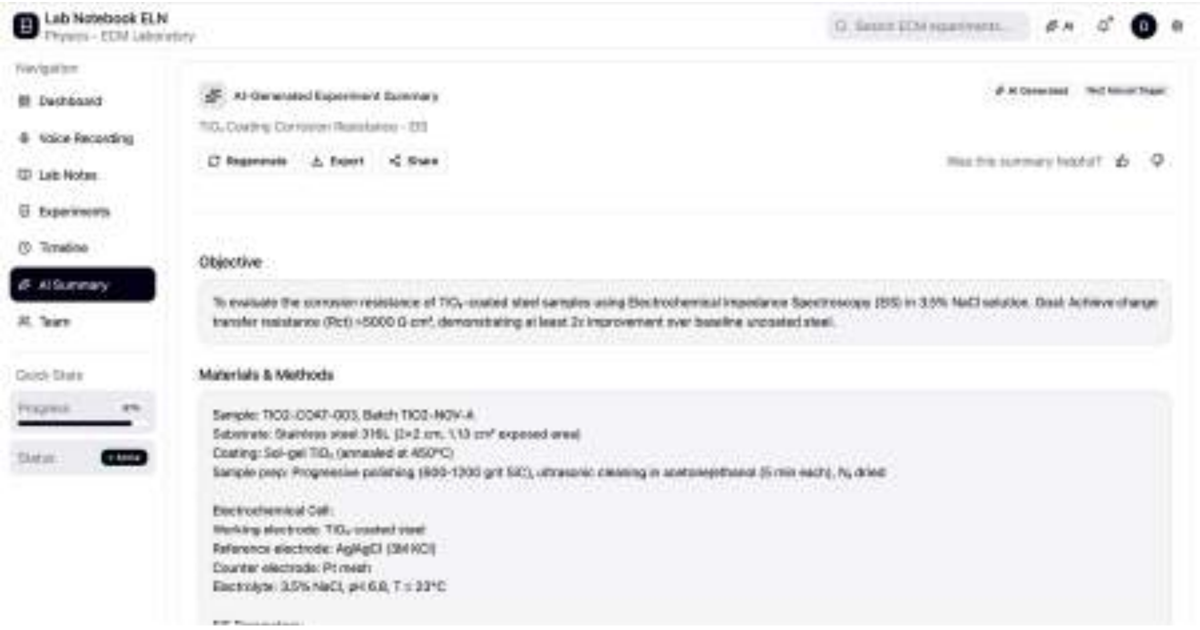
Illustration 7: [Timeline screen from the Prototype Concept**]**



This screenshot is the timeline screen, it menitons who did what and when, in the ELN entry if in a project with multiple team members. . Source: Own screenshot

These principles are made tangible within the AI Summary Page (see Illustration 8) and the Timeline Screen (see Illustration 7). The AI Summary Page presents automatically generated overviews that are clearly labeled as system-generated and remain fully editable by the user, thereby preserving epistemic responsibility. The Timeline Screen documents changes and contributions in a structured chronological format, ensuring traceability and transparency of authorship. Together, these interfaces operationalize integrity not as rigid procedural enforcement, but as visible and user-controllable documentation structure.

Illustration 8: [AI Summary screen from the Prototype Concept]



This screenshot presents AI summary screen, once navigated from the left panel, this screen gives a detailed summary of the current experiments. Source: Own screenshot

3.3.4 Security vs. Usability Friction

Survey data indicated strong concern for data security, but only when such measures did not impair workflow efficiency. The literature similarly identifies excessive authentication or interaction friction as detrimental to adoption.

The design implication was to position governance primarily at the infrastructural level rather than at the interaction layer. Within the prototype, this was reflected in the conceptual assumption of persistent authentication and minimal interruption logic.

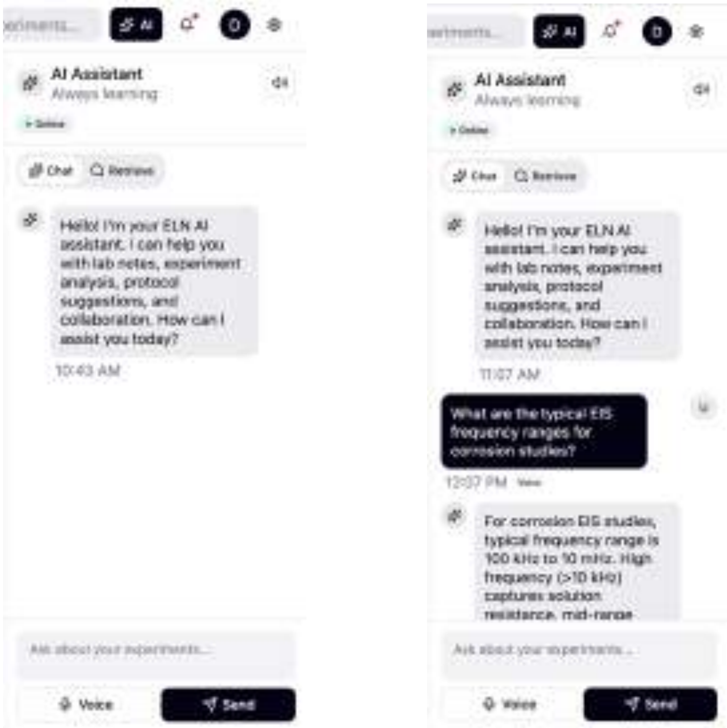
Although infrastructural security mechanisms were not technically implemented in the WoZ prototype, interaction-level friction was deliberately minimized across all screens. The Dashboard (see Illustration 4) and subsequent views avoid repetitive confirmation dialogs or authentication barriers, thereby aligning with survey findings emphasizing security without workflow disruption. Security is therefore conceptually positioned at the governance layer rather than embedded as intrusive interaction logic.

3.3.5 Automation vs. Human-in-the-Loop Control

Human–AI interaction research and survey feedback highlighted concerns about loss of control and responsibility in AI-supported systems. Observations further emphasized accountability in scientific documentation.

The design implication was that AI must augment rather than replace researcher judgment. This was operationalized through strictly user-invoked chatbot behavior, no autonomous modification of records, visible AI labeling, and editable outputs.

Illustration 9: [Chatbot screen from the Prototype Concept**]**



This screenshot presents the chatbot functionality in the prototype, on the left is the AI chatbot with the message from the simulated AI assistant, on the right, is a simulated response based on the query given to the chatbot.
Source: Own screenshot

The AI Chatbot Interface (see Illustration 9) is implemented as a collapsible side panel that remains inactive until explicitly triggered by the user. It does not autonomously initiate documentation changes or interrupt ongoing tasks. AI-generated outputs are visually distinguishable from user-authored content and can be edited or disregarded. This separation between the primary ELN workspace and conversational assistance preserves user authority while allowing optional augmentation.

3.3.6 Cognitive Switching and Multimodal Interaction

Laboratory work requires sustained focus and continuous physical engagement. Observations revealed frequent cognitive switching between experiment execution and documentation tasks.

The design implication was to reduce physical disruption while preserving verification capabilities. This was operationalized through a combination of voice-based input and persistent visual confirmation within the Figma-based prototype interface.

Multimodality is visible across the Voice Recording Screen (see Illustration 5) and all visual ELN views, including the Dashboard (see Illustration 4) and Lab Notes Screen (see Illustration 6) While voice enables hands-free interaction, all system responses are rendered visually to ensure verification and correction. Users may seamlessly switch between spoken commands and touch-based interaction depending on environmental conditions. This architectural layering reduces disruptive cognitive switching while maintaining transparency and confirmation capabilities.

3.3.7 Prototype Artifact Overview and Screen-Level Operationalization

The prototype serves as a concrete and experiential representation of the proposed AI-assisted ELN interaction concept. Rather than aiming to demonstrate technical feasibility or evaluate a fully implemented AI system, it makes the underlying design assumptions, interaction flows, and human–AI decision points tangible and open to discussion. The prototype therefore functions as a WoZ-ready interaction artifact that allows participants to experience how conversational and voice-enabled support could fit into everyday laboratory workflows, and it provides the basis for reflecting on usability, transparency, perceived usefulness, and trust during the evaluation (Chapter 4).

The prototype consists of a set of interconnected screens that operationalize the structural design drivers derived through triangulation (Chapter 3.2). Each screen is not treated as an isolated feature; instead, it implements one or more design requirements that respond to recurring tensions in laboratory documentation practice (e.g., in-situ documentation versus post-hoc reconstruction, collaboration gaps, integrity demands, and the need for human-in-the-loop control). To make this mapping explicit in the final manuscript, screenshots should be embedded immediately after the paragraph(s) in which the respective screen is introduced and interpreted. This placement ensures that figures function as evidential anchors for the argument rather than as detached illustrations.

Accordingly, the Dashboard Screen supports rapid entry into documentation and retrieval with minimal navigation overhead and should be shown after the Dashboard discussion in Chapter 3.3.1. The Voice Recording Screen operationalizes hands-free, low-interruption capture and should be shown after the corresponding paragraph in Chapter 3.3.1. The Experiments Screen and Lab Notes Screen instantiate experiment-level organization, metadata visibility, and shared knowledge artifacts and should be placed after their first mention in Chapter 3.3.2. The Timeline Screen and AI Summary Page enact traceability and editable synthesis and should be embedded where integrity, provenance, and accountability are discussed in Chapter 3.3.3. The Team Page supports group awareness and coordination and should be placed where this driver is introduced in Chapter 3.3.2 (or in a short paragraph on awareness if kept separate). Finally, the AI Chatbot Interface operationalizes conversational retrieval and user-invoked assistance under human-in-the-loop constraints and should appear directly after the chatbot description in Chapter 3.3.5. The following chapter describes how these design drivers were technically instantiated within the WoZ-ready prototype architecture.

4 Wizard of Oz Study

This chapter presents the Wizard-of-Oz (WoZ) evaluation, The Wizard-of-Oz approach is based on the simulation of an interactive system by a human operator who controls system's responses in real time (Steinfeld, Jenkins, & Scassellati). This method makes it possible to investigate user experience, expectations, and interaction behavior at an early design stage, without the technical constraints of building a fully functional AI system. Depending on the study design, participants may or may not be informed about the simulation; in this study, participants were informed that parts of the AI functionality were simulated in order to maintain transparency and avoid misleading any expectations.

which constitutes the sole formal empirical study of this thesis. The purpose of the study was to evaluate the conversational interaction paradigm of the proposed ELN prototype under simulated AI conditions. The study does not assess the technical performance of artificial intelligence systems. Instead, it investigates user experience, perceived usability, workload, and perceived system appropriateness in response to a human-mediated simulation of AI-supported interaction.

All findings must therefore be interpreted as perceptions of an interaction concept under controlled and idealized conditions.

Illustration 10: [A usability-test session using the Wizard of Oz method**]**

Source: (Rosala & Paul, 2024)

A Wizard-of-Oz (WoZ) methodology was employed to simulate AI-driven functionality without implementing an actual AI backend. In this setup, participants interacted with what appeared to be an AI-assisted ELN system, while a trained human operator (the "wizard") generated all system outputs in real time.

The wizard followed predefined scripts and response rules to ensure consistency across sessions. Participants were informed that certain system functions were simulated.

The WoZ method was selected because the objective of this study was not to evaluate AI model performance, but to examine the interaction paradigm of a conversational ELN under idealized AI conditions. Implementing fully functional speech recognition, natural language understanding, and summarization systems would have shifted the focus toward technical validation rather than interaction research.

The WoZ approach enabled the investigation of:

- Perceived usability of conversational interaction
- Perceived workload during documentation
- Perceived support for research continuity
- Perceived reliability and appropriateness of system behavior

Technical failures (e.g., recognition errors or latency issues) were intentionally minimized in order to evaluate the interaction concept independently of current technological limitations.

All AI-related outputs—including voice transcription, contextual summaries, and chatbot responses—were manually generated by the wizard (refer Appendix VIII. IV). No automated speech recognition, natural language processing, or machine learning models were deployed.

Consequently, the findings reflect perceptions of the interaction paradigm rather than validation of autonomous AI capabilities.

Participants were informed prior to participation that parts of the system were simulated. Written informed consent was obtained. Audio, video, and interaction data were collected exclusively for research purposes and anonymized during analysis.

4.1 Participants

Participants for the Wizard-of-Oz (WoZ) evaluation study were recruited from the Department of Physics at the University of Siegen, with a specific focus on students who had regular exposure to laboratory environments and established documentation practices. Although the user requirements were originally collected within the DLR context, participants for the evaluation were not recruited from DLR primarily due to logistical and organizational constraints. The Kadi4Mat team is distributed across different DLR locations between Oberpfaffenhofen and Cologne, which would have required significant travel, coordination of on-site access, transportation of technical equipment, and additional resource management within the limited timeframe of the thesis. Conducting the study at the University of Siegen allowed in-person sessions to be organized efficiently, reduced travel time, and ensured reliable access to laboratory infrastructure and participants. Furthermore, an advertisement calling for participation was shared among university students via social media channels. The recruitment advertisement is provided in Appendix VIII. II. Targeting students and researchers actively engaged in laboratory work ensured alignment between the study participants and the intended laboratory context of use, while also enabling an assessment of the concept's applicability beyond the original DLR environment.

It must be acknowledged that this recruitment strategy constitutes a form of convenience sampling, as participants were selected based on accessibility and availability rather than through random or stratified sampling procedures. Consequently, the sample cannot be considered statistically representative of the broader population of laboratory researchers. In addition, individuals who voluntarily chose to participate may be more open to digital tools and innovation, meaning that early adopters or technology-positive participants could be overrepresented. This introduces a potential bias toward favorable perceptions of AI-supported systems. The implications of this sampling strategy for generalizability and external validity are discussed in Chapter 6 (Limitations).

An online survey was distributed to potential participants upon confirmation of appointment via DFN terminplanner thereby later using Google Forms to collect background information relevant to the study. The survey captured demographic data as well as information on academic level, extent of laboratory experience, current documentation practices, and prior exposure to Electronic Laboratory Notebooks. The purpose of this survey was to confirm that all participants possessed sufficient familiarity with laboratory workflows and documentation routines.

Following confirmation of participation via Terminplanner, an online survey was distributed using Google Forms to collect background information relevant to the WoZ evaluation. The survey captured demographic data, academic level, extent of laboratory experience, current documentation practices, and prior exposure to Electronic Laboratory Notebooks. Its purpose was to ensure that all participants possessed sufficient familiarity with laboratory workflows and documentation routines to meaningfully engage with the simulated AI interaction concept.

Based on the survey responses, ten participants were selected to take part in the WoZ evaluation. The final sample consisted of eight master's students and two doctoral candidates, all affiliated with the Department of Physics. Among the master's students, two identified as female and six as male; both doctoral candidates identified as male. Participants' laboratory experience ranged from 1–2 years to over 10 years, reflecting different stages of academic training and research involvement. A detailed overview of participant characteristics is provided in Appendix VIII. XIII.

To coordinate the individual evaluation sessions, a scheduling link was distributed to the selected participants using DFN Terminplaner. This allowed participants to independently select suitable time slots for the in-person WoZ sessions and supported efficient study organization.

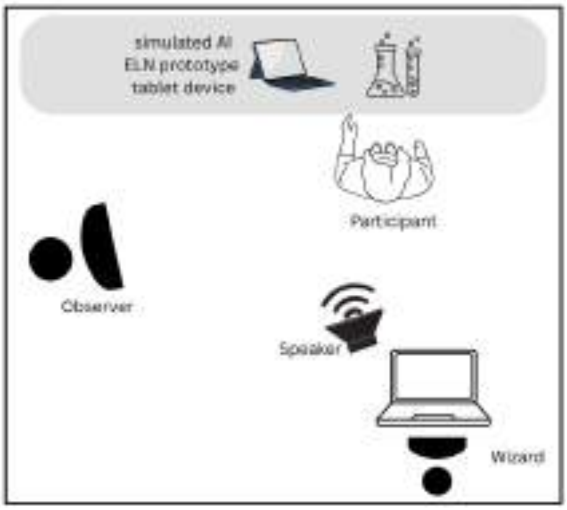
The majority of participants reported relying primarily on paper-based laboratory notebooks in combination with general-purpose digital tools such as Word, Excel, or OneNote for their documentation activities. Only one participant indicated regular prior use of a dedicated Electronic Laboratory Notebook, while several participants reported limited or indirect exposure to ELN systems. Appendix VIII. XIII.

Participation in the study was voluntary, and all participants received compensation for their time and effort, either in the form of a €10 allowance or a meal voucher for the university mensa, independent of study outcome. To ensure anonymity and compliance with data protection requirements, participants are referred to using pseudonymous identifiers (P01–P10) throughout this thesis.

4.2 Experimental Setup

The study was conducted in an authentic laboratory environment within the Department of Physics at the University of Siegen. The laboratory was selected for the same logistical and contextual reasons outlined in chapter 4.2. As previously discussed, conducting the study at the University of Siegen ensured practical feasibility within the timeframe of the thesis while maintaining contextual alignment with laboratory-based documentation practices. The environment therefore provided a realistic yet controlled evaluation context without requiring artificial staging or modification.

Illustration 11: [Top-view schematic of the Wizard-of-Oz experimental setup.]



Source: Own image

Illustation 11 shows, the prototype was displayed on an iPad positioned at a regular laboratory workstation. The Figma prototype, as described in Chapter 3, was executed within a web browser on the tablet. The wizard operated from a laptop located behind the participant within the laboratory space and remained physically out of the participant’s direct line of sight. A Bluetooth speaker was connected to the wizard’s laptop in order to project the simulated AI voice responses from behind the participant, thereby reinforcing the impression of a system-driven output rather than human intervention. During each session, the researcher additionally positioned themselves to the side of the setup to take observational field notes without interfering with the interaction. Existing infrastructure allowed integration of the prototype and supporting equipment without modifying the surrounding environment.

4.2.1 Baseline Condition (Traditional Laboratory Notebook)

Before the ELN prototype was introduced, participants first completed the same three tasks using a traditional laboratory notebook (paper-based documentation as typically used in their daily practice). This baseline condition was implemented to allow a descriptive comparison between established documentation practices and the conversational ELN interaction concept.

After completing the tasks with the traditional notebook, participants filled out the System Usability Scale (SUS) and the NASA Task Load Index (NASA-TLX). The adapted Trust in Automation (TiA) scale was not administered in the baseline condition, as traditional laboratory notebooks do not involve automated or autonomous system behavior. Consequently, perceived trust in automation was not conceptually applicable to this condition.

Following completion of the baseline measures, participants performed the identical set of tasks using the Simulated AI ELN prototype under Wizard-of-Oz conditions.

4.2.2 Tasks

Each session involved one participant completing three predefined tasks, Appendix VIII. V. All participants performed the same underlying experimental scenario to ensure strict comparability across sessions.

Before beginning the tasks, participants were familiarized with the study procedure using a short-written guideline. The guideline explicitly described the session as a simulated experience of an intelligent ELN assistant, without revealing the presence or operational role of the wizard. Participants were informed about the availability of voice and chat interaction modalities.

The three tasks were as follows:

1. **Recording an experimental observation using voice input.**
   Participants were prompted to activate the voice assistant. Upon activation, a standardized pre-recorded text-to-speech output was played: “Hi there, I am your AI ELN.” The participant then verbally dictated an experimental observation. After the input, a standardized confirmation message was played: “Experimental observation noted.” These voice outputs were generated using an online text-to-speech tool and triggered by the wizard.
2. **Retrieving a summary of a previous experiment via chatbot.**
   All participants requested a summary of the same predefined experimental scenario. The returned summary was identical across participants and was selected from a prepared response dataset.

3. **Retrieving contextual information about prior experimental work.**
   Participants asked for contextual details regarding prior experimental activity (e.g., who conducted the experiment, when it was performed, and which observations were recorded). The responses were standardized and identical for all participants.

By ensuring that all participants worked with the same experimental content and received the same summaries and contextual information, variability due to content differences was eliminated. This design decision allowed the evaluation to focus exclusively on interaction behavior and user perception rather than domain complexity.

4.2.3 Wizard Protocol

The wizard operated using a predefined dataset and structured response library. All experimental summaries, contextual information, and chatbot responses were prepared prior to the study and remained identical across participants.

The wizard's responsibilities included:

- Triggering standardized voice responses (text-to-speech outputs).
- Selecting pre-written summary responses for Task 2.
- Selecting pre-defined contextual information responses for Task 3.
- Structuring spoken input into formatted ELN entries for Task 1.

The wizard was not permitted to introduce new domain knowledge, modify experimental content, or improvise beyond minor linguistic adjustments that did not alter meaning. All content-related outputs were strictly standardized. (refer Appendix VIII. IV)

For Task 1, the wizard triggered two fixed voice outputs:

- Activation response: "Hi there, I am your AI ELN."
- Confirmation response after dictation: "Experimental observation noted."

For Tasks 2 and 3, the wizard selected the corresponding pre-scripted text responses from a prepared answer set. No dynamic content generation occurred.

Improvisation was restricted to operational adjustments (e.g., minor phrasing alignment), without changing informational substance. The wizard refrained from any direct verbal or non-verbal interaction with participants. All assistance was mediated exclusively through the system interface to preserve the illusion of autonomous system behavior. System responses were delivered within approximately 2–5 seconds to simulate realistic processing delays. No recognition errors, hallucinations, or system breakdowns were simulated. The study aimed to evaluate the interaction paradigm under idealized conditions rather than stress-test technical robustness.

4.2.4 Wizard-of-Oz Architecture (Interaction Simulation Layer)

The technical and logical architecture underlying the Wizard-of-Oz simulation is described in detail in Chapter 3.4. Within the context of this study, the prototype functioned as a frontend interaction layer without backend AI integration. The original high-fidelity interface was designed in Figma and subsequently migrated into a code-based prototyping environment (Figma Make) to enable functional conversational behavior, including real-time text input and dynamic rendering of responses.

No speech recognition engine, natural language processing pipeline, database query system, or machine learning model was implemented. All AI-related functionality was simulated through a human-in-the-loop structure.

User inputs (voice or text) were interpreted by the wizard and mapped to predefined intent categories. Corresponding response templates were then selected from a structured response set and rendered in the chatbot interface. Response classes included confirmation messages, retrieval outputs, summaries, contextual information, clarification prompts, and standardized fallback responses.

Several control constraints governed the simulation: the system was strictly user-invoked and did not initiate interaction autonomously; no adaptive learning or personalization occurred; no new experimental data were generated beyond predefined scripted content; unsupported queries triggered standardized fallback responses; and short artificial delays were inserted to simulate processing time.

These architectural constraints ensured that the evaluation focused on perception of the interaction paradigm rather than on operator improvisation or technical AI performance.

4.3 Data Collection and Analysis

Each session lasted approximately 30–45 minutes. The following data sources were collected:

- System Usability Scale (SUS) (Appendix VIII. VI)
- NASA Task Load Index (NASA-TLX, unweighted) (Appendix VIII. VII)

- Adapted Trust in Automation scale (Appendix VIII. VIII)
- Semi-structured post-study interviews (Appendix VIII.IX)
- Screen and audio recordings
- Observational field notes (Appendix VIII. XII)

Interviews were audio-recorded and transcribed verbatim. Questionnaires were administered immediately after task completion.

4.3.1 Quantitative Analysis

Descriptive statistics were calculated for all questionnaire measures, including mean (M), standard deviation (SD), minimum, maximum, and 95% confidence intervals (CI). Reporting follows APA guidelines.

Given the exploratory design and limited sample size, results are interpreted descriptively. Effect sizes are reported for contextual comparison; however, no inferential statistical claims are made.

4.3.2 Qualitative Analysis

Qualitative data were analyzed using thematic analysis following Terry et al. The procedure included:

- Familiarization with transcripts.
- Initial open coding.
- Development of candidate themes.
- Iterative refinement.
- Final definition and naming of themes.

Coding was conducted using MAXQDA. Segments were re-reviewed to ensure internal consistency of coding.

4.4 Results

This chapter reports the empirical findings of the WoZ evaluation. Only results are presented. All findings reflect user experience with a human-mediated simulation of AI functionality.

4.4.1 Quantitative Results

All quantitative measures reflect user experience under simulated AI conditions (N = 10).

System Usability Scale (SUS)

Perceived usability was assessed using the System Usability Scale (SUS)(Bangor, Kortum, & Miller, 2008) for both the traditional laboratory notebook and the AI enabled ELN.

Table 3: [System Usability Scale (SUS) Scores by Condition]

Condition	Mean (M)	SD
Traditional Notebook	54.50	15.17
AI enabled ELN	76.25	11.74

The traditional laboratory notebook achieved a mean SUS score of M = 54.50 (SD = 15.17), with scores ranging from 30 to 75. The distribution of scores indicates considerable variability in perceived usability of existing documentation practices within the same participant group.

In contrast, the AI enabled ELN achieved a mean SUS score of M = 76.25 (SD = 11.74), with individual scores ranging from 55 to 90. Scores clustered predominantly in the upper half of the scale, indicating generally positive usability perceptions under simulated AI-supported interaction conditions.

The difference between conditions corresponds to a large descriptive effect (Cohen's d = 1.60). Given the exploratory design and small sample size, this effect size is reported descriptively without inferential statistical testing.

NASA Task Load Index (NASA-TLX)

To assess perceived workload, participants completed the NASA-TLX (" NASA TLX NE08F1-4") for both conditions. Overall workload was calculated as the unweighted mean of the six dimensions per participant.

Table 4: [Overall NASA-TLX Workload Comparison]

Condition	Mean (M)	SD	Min	Max
Traditional Notebook	4.13	0.53	3.33	5.00
AI enabled ELN	3.05	0.63	2.33	4.33

Participants reported higher overall workload when using traditional documentation practices compared to the AI enabled ELN. The distribution of values in the traditional condition shows consistently moderate-to-high workload ratings across participants, whereas the AI enabled ELN condition demonstrates lower and more compact workload scores.

The difference between conditions corresponds to a large descriptive effect (Cohen's d = 1.86). As with the SUS results, this effect is interpreted descriptively.

Dimension-level means are presented below:

Table 5: [NASA-TLX Subscale Means by Condition]

Dimension	Traditional (M)	AI enabled (M)
Mental Demand	4.30	3.10
Physical Demand	3.90	1.90
Temporal Demand	3.90	3.10
Performance	4.20	5.00
Effort	4.80	3.10
Frustration	3.70	2.10

Across dimensions, physical demand and effort show notable descriptive differences between conditions. Self-reported performance ratings were higher in the Simulated AI condition.

Perceived System Appropriateness (Adapted Trust Scale)

Participants' perceptions of system trustworthiness were assessed using a Trust in Automation (TiA) questionnaire (Körber) following interaction with the AI enabled ELN prototype. Given the Wizard-of-Oz methodology, participants interacted with a simulated AI system under controlled conditions. Accordingly, the results reflect perceptions of the interaction concept and interface behavior rather than trust in a fully autonomous AI implementation.

Table 6: [Perceived System Trustworthiness – AI-enabled ELN]

Dimension	Mean (M)	Standard Deviation (SD)
Trust	3.70	1.16
Reliability	3.60	0.84
Predictability	3.00	0.82
Dependability	4.00	0.82
Safety	3.90	0.74

Note. Ratings were collected on a 5-point Likert scale (1 = low, 5 = high). SD represents sample standard deviation (n = 10).

Overall, participants reported moderate to high levels of perceived system trustworthiness. The highest ratings were observed for dependability (M = 4.00, SD = 0.82) and safety (M = 3.90, SD = 0.74), indicating that participants generally felt comfortable relying on the system and did not perceive it as unsafe during task execution.

Perceived reliability was rated positively (M = 3.60, SD = 0.84), suggesting that system responses were experienced as largely consistent. In contrast, predictability received the lowest mean rating (M = 3.00, SD = 0.82), indicating some uncertainty regarding how the system might behave across varying situations.

The broader trust dimension (M = 3.70, SD = 1.16) exhibited the highest variability across participants, reflecting individual differences in comfort with AI-supported interaction paradigms.

Importantly, these findings should be interpreted as evaluations of the perceived appropriateness, consistency, and dependability of the interaction paradigm under ideal performance conditions. As the system behavior was simulated through a Wizard-of-Oz setup, the results do not represent empirical validation of trust in algorithmic autonomy or technical robustness, but rather users' assessments of the interaction design and system behavior as experienced during the study.

4.4.2 Qualitative Findings

The qualitative analysis highlights key themes that emerged from interview transcripts and observational data.

Voice Interaction in Laboratory Context

Voice interaction emerged as a central input modality during the evaluation. Most participants relied primarily on voice input when documenting experimental steps or retrieving information. Even participants who initially described skepticism toward voice assistants engaged with the interface during the study.

One participant reflected:

> "I usually don't like talking to devices… But during the experiment, I actually ended up using the voice input much more than I have used in my other devices." (P07)

Participants reported that voice input allowed them to remain at their workstation and continue their tasks without interruption:

> "I could stay at my workstation and just continue what I was doing. If I wanted to note something or correct it, I could just ask again." (P02)

> "With voice input, I didn't have to break the workflow at all." (P05)

Concerns were primarily directed toward anticipated real-world robustness in noisy laboratory environments rather than toward the interaction itself:

> "In a real lab situation… I wouldn't expect the system to always get everything right." (P09)

> "It worked well here, but I imagine that in everyday lab work it may struggle sometimes." (P06)

Participants also mentioned that explicit activation phrases could interrupt workflow rhythm:

> "Having to say a specific activation phrase each time is fine, but in a busy moment it might feel like an extra step." (P04)

Multimodal Interaction

Participants alternated between voice and touch interaction depending on task demands. Voice was frequently used during documentation, whereas touch interaction was described as efficient for navigation and corrections.

> "Using only voice without any touch option would be frustrating for me." (P03)

> "If I can still control it with my fingers, that's usually faster. Especially when there's aalready a button for what I want to do." (P08)

The ability to switch between modalities was repeatedly described as beneficial and situationally appropriate.

Transparency and Perceived System Behavior

Participants emphasized the importance of visible and audible feedback when system actions were triggered.

> "When she said, 'I'm starting a new note now' or 'I've saved this entry,' I didn't have to worry about whether something was missed." (P10)

> "Even if I wasn't looking at the screen, I knew that the system had done something." (P09)

Uncertainty arose when system status was not clearly communicated:

> "At some point I was wondering, is the assistant still tracking the flow of my speech?" (P04)

Transparent feedback mechanisms were frequently described in relation to perceived reliability and coordination with the system.

Personification and Social Presence

Participants occasionally used gendered pronouns when referring to the assistant, which may be attributed to the use of a female voice in the simulation.

> "It feels like I'm not working alone." (P07)

> "Seeing the progress bar move forward somehow motivated me to finish my tasks." (P08)

Conversational feedback and activity indicators were associated with a perceived sense of collaboration or social presence during interaction.

Interpretation of these findings is provided in the subsequent discussion chapter.

.

5 Discussion

The previous chapter described the design of the conversational ELN prototype and the implementation of the WoZ study, including the experimental setup, measurement instruments, and data collection procedures. Building on that methodological foundation, the present chapter interprets the empirical findings within those established constraints. As the system operated under simulated conditions, the results reflect perceptions of an idealized interaction paradigm rather than validated AI performance.

5.1 Usability and Interaction Structure

The conversational ELN prototype achieved descriptively higher System Usability Scale (SUS) scores than traditional laboratory notebook practices. Although no inferential claims are made due to the exploratory design and limited sample size, the consistent descriptive difference indicates that participants experienced the structured conversational paradigm as more usable than their established documentation routines.

This perceived improvement appears to derive from interaction structure rather than technological intelligence. The prototype provided explicit confirmations, structured turn-taking, and immediate feedback. Traditional documentation practices, by contrast, rely on self-organized notetaking without interactive scaffolding. The variability observed in SUS scores for traditional notebooks suggests heterogeneous and individually developed documentation strategies. The conversational prototype reduced structural ambiguity by introducing standardized interaction flows.

From a distributed cognition perspective, this structuring effect can be interpreted as a redistribution of cognitive effort between user and artifact. The system does not merely store information but actively shapes how documentation is produced and retrieved. The usability advantage therefore appears rooted in interaction scaffolding and feedback clarity rather than in automation itself.

5.2 Perceived Workload and Documentation Flow

NASA-TLX results indicate descriptively lower perceived workload in the simulated AI condition, particularly in physical demand and effort. Participants reported that voice-based documentation enabled them to remain at their workstation during experimental procedures, thereby minimizing physical transitions between experimentation and documentation.

In laboratory environments characterized by gloves, instruments, and sterile constraints, reduced interaction friction may meaningfully influence perceived effort. However, the simulation operated under ideal technical conditions: no recognition errors, latency variations, or system misunderstandings were introduced. The observed workload reduction therefore reflects perception of an idealized conversational interface rather than realistic deployment performance.

Nevertheless, the findings suggest that conversational structuring reshapes how documentation effort is experienced. Even without objective efficiency measures, the interaction design altered subjective workload perception by reorganizing the distribution of cognitive and physical actions within the task environment.

5.3 Multimodality and User Agency

Qualitative findings demonstrate that participants did not rely exclusively on voice interaction. Instead, voice and touch inputs were used complementarily depending on situational demands. The ability to switch modalities was valued particularly in moments requiring immediate correction or visual confirmation.

These results indicate that conversational interaction should function as an augmentation layer rather than a replacement of graphical interfaces. The perceived usability gains appear to derive not solely from voice input but from multimodal flexibility. The availability of alternative interaction channels increased perceived control and reduced dependency on a single mode of input.

Within human–computer interaction research, multimodality is associated with resilience and preservation of user agency in complex work environments. In laboratory contexts, maintaining control and procedural accountability is central, and the prototype appears to support this requirement.

5.4 Transparency, Predictability, and Trust Calibration

Transparency emerged as a central theme in the qualitative data. Participants emphasized the importance of audible confirmations and visible system updates during interaction. These mechanisms influenced perceived reliability and coordination with the system.

Predictability received comparatively lower ratings in the adapted trust measure, indicating that participants remained attentive to system boundaries even under idealized conditions. Because no autonomous AI functionality was present, the findings reflect trust in interaction consistency rather than trust in artificial intelligence.

Trust calibration in this study was shaped primarily by feedback clarity and boundary signaling. This aligns with research in human–AI interaction suggesting that perceived reliability and explainability are central determinants of calibrated trust.

Some participants used gendered pronouns and described a sense of "not working alone." Such responses indicate anthropomorphic attribution triggered by conversational framing. However, these reactions do not constitute evidence of perceived artificial agency. The system exhibited neither autonomy nor adaptive reasoning.

5.5 Collaboration and Research Continuity

Including a baseline condition with traditional laboratory notebooks makes it possible to reflect on how a structured conversational interaction differs from existing documentation practices. Traditional notebooks were perceived as more demanding in terms of workload and showed greater variation in usability, which can be attributed to highly individualized ways of documenting experimental work. In contrast, the conversational prototype offered consistent confirmation, retrieval, and summarization flows that reduced variability in how documentation was structured. From a CSCL perspective, such structuring can support shared knowledge construction by helping establish common ground and continuity across experiments. At the same time, collaborative benefits were captured only through participants' subjective impressions, and no objective conclusions can be drawn regarding improvements in coordination, documentation accuracy, or collaborative performance. This thesis contributes to understanding how conversational interaction paradigms can be designed for scientific documentation contexts. By intentionally separating the interaction layer from any backend AI implementation using a WoZ approach, the study allows a focused examination of how users experience usability, workload, transparency, multimodal interaction, and perceived collaborative support under idealized conditions. The findings highlight which interactional qualities shape user experience and position conversational ELN within broader discussions of distributed cognition and CSCL. Overall, the results suggest that a structured conversational ELN interaction is experienced as usable and supportive of laboratory documentation workflows when AI functionality is simulated, with perceived benefits stemming from interaction design, transparency, multimodality, and cognitive scaffolding rather than from validated AI performance.

6 Limitations

This chapter critically reflects on the methodological, conceptual, and contextual boundaries of the present thesis and clarifies the conditions under which the findings should be interpreted. The study was deliberately exploratory and centered on a Wizard-of-Oz (WoZ) evaluation of a simulated AI interaction concept rather than on a fully implemented AI system. Accordingly, the results cannot be statistically generalized, nor can they be interpreted as validation of deployable AI performance in real-world AI systems. Instead, they provide early-stage, design-oriented insights into how researchers perceive and experience a conversational, simulated AI interaction paradigm within an AI enabled ELN prototype.

6.1 Sample Characteristics and Transferability

All participants in the WoZ study were recruited from physics-related research environments. While this ensured contextual coherence with the POLiS laboratory visit and partially aligned with the DLR survey background, it limits cross-domain transferability. Documentation practices, regulatory constraints, collaboration norms, and data governance requirements vary substantially across disciplines such as chemistry, biology, medical research, and engineering. It therefore cannot be assumed that the interaction patterns, usability perceptions, and trust assessments observed in this study would translate directly to other scientific domains without adaptation and re-evaluation.

In addition, the majority of participants were male. Prior research in human–AI interaction and voice interface adoption indicates that gender-related differences may exist in trust calibration, communication preferences, and perceptions of conversational agents. Consequently, the present findings may not fully capture how female researchers or more gender-diverse samples would evaluate AI-simulated documentation and voice-based interaction in laboratory contexts. Future work should therefore aim for more balanced gender representation to examine potential differences in trust development, perceived authority of AI systems, and acceptance of voice interaction.

The participant group consisted exclusively of Master's and PhD students. Senior researchers, principal investigators (PIs), laboratory managers, and technical staff were not included. These stakeholder groups may hold different expectations regarding documentation responsibility, authorship attribution, institutional compliance, and long-term data governance. For senior researchers in particular, concerns related to legal validity, auditability, and accountability may outweigh considerations of interaction smoothness or convenience. The findings presented in this thesis therefore primarily reflect early- to mid-career researchers' perspectives rather than the full spectrum of laboratory roles.

6.2 Ecological Validity and Study Context

Although the study was conducted in a real laboratory environment, the evaluation tasks were predefined, time-bounded, and artificial. Participants interacted with a scripted scenario rather than integrating the system into their own ongoing research projects. As a result, ecological validity is inherently limited. The study does not capture how the system would perform under authentic time pressure, prolonged experimental campaigns, or evolving project demands.

A further limitation concerns the absence of long-term usage data. The evaluation measured immediate impressions of usability, workload, and trust, but did not assess habituation effects, sustained utility, or integration into daily routines over extended periods. Positive feedback regarding voice interaction, in particular, may partly reflect novelty effects rather than durable practical value. Without longitudinal observation, it remains unclear whether researchers would continue to rely on voice-based documentation once initial curiosity diminishes or when confronted with recurring real-world constraints such as background noise, shared lab spaces, or recognition inaccuracies.

Moreover, although the conceptual design draws strongly on Computer-Supported Collaborative Learning (CSCL) frameworks and emphasizes group awareness and shared understanding, the evaluation itself did not include a collaborative scenario. Participants interacted with the prototype individually. Consequently, collaborative constructs remain theoretically grounded but empirically untested within multi-user conditions. Claims regarding enhanced group awareness, shared knowledge construction, or improved coordination must therefore be interpreted as conceptual implications rather than validated collaborative outcomes.

6.3 Wizard-of-Oz Methodological Constraints

The WoZ methodology introduces inherent trade-offs. Because core AI functionality was simulated by a human operator within the AI enabled ELN prototype, participants interacted with an idealized version of the system that did not reflect real-world technical limitations such as speech recognition errors, latency variability, incomplete database queries, or model hallucinations. While this abstraction was methodologically necessary to isolate the interaction paradigm from backend performance, it limits conclusions about deployable system robustness and reliability.

The same trained wizard operator conducted all sessions to maximize procedural consistency. A structured response protocol, predefined intent categories, and standardized response templates were used to minimize variability across sessions. Nevertheless, subtle differences in interpretation speed, phrasing, or response timing may have influenced participants' perceptions of system intelligence and fluency. Even under standardized conditions, human mediation can affect interaction quality and perceived system competence.

Participants were informed that the AI was simulated. This transparency reduced deception and supported ethical clarity, but it may also have influenced trust judgments. Awareness of human mediation could have reduced performance-related expectations or softened reactions to limitations, thereby affecting scale responses. The precise influence of this awareness on trust calibration cannot be fully disentangled within the present design.

6.4 Measurement Constraints and Statistical Scope

The evaluation relied predominantly on self-reported measures, including the System Usability Scale (SUS), NASA Task Load Index (NASA-TLX), and the Trust in Automation (TiA) scale. These standardized instruments provide valuable insight into perceived usability, workload, and trust, yet they do not capture objective performance metrics such as task completion time, documentation accuracy, retrieval precision, or measurable productivity improvements. The absence of behavioral performance indicators limits conclusions regarding efficiency gains or task effectiveness.

Given the small sample size, only descriptive statistics were reported. Inferential testing was not conducted, and relationships between usability, workload, and trust remain exploratory rather than statistically validated. Any interpretation of associations among these constructs must therefore be treated as correlational and tentative rather than confirmatory.

The reliance on self-report data also introduces the possibility of social desirability effects. Participants may have provided favorable evaluations due to awareness of the research context, interest in simulated AI innovation, or perceived expectations from the researcher. Minor inconsistencies across certain scale items may additionally suggest conceptual ambiguity or measurement sensitivity that warrants further investigation in larger samples.

6.5 Alignment Between Design Grounding and Evaluation Sample

The design requirements were derived through triangulation of literature insights, the DLR eLab survey, and contextual observations from the Kadi4Mat community meeting and the POLiS laboratory visit. However, the evaluation sample did not fully match the populations represented in these sources. Although the WoZ participants were researchers, their institutional and disciplinary backgrounds differed from those of the DLR survey cohort. This mismatch limits direct validation of the original requirement sources. The evaluation therefore assesses whether the derived interaction paradigm is coherent and plausible for a related, but not identical, user group. Future studies should include participants from the original requirement-generating contexts to strengthen the link between design inputs and empirical validation.

6.6 Iterative Scope and Maturity of the Prototype

Finally, the thesis represents a single design–evaluation iteration within a broader user-centered design process. Iterative refinement, longitudinal deployment, cross-institutional replication, and multi-domain validation were beyond the scope of this work. The prototype should therefore be understood as a research artifact intended to explore and critically examine a conversational interaction paradigm, rather than as a finalized, production-ready ELN system.

Taken together, these limitations delineate the interpretive boundaries of the findings. The results should be understood as exploratory, perception-based insights derived from a simulated AI interaction concept evaluated in a controlled laboratory setting with a limited and relatively homogeneous sample. They do not constitute statistical generalization, longitudinal validation, or performance benchmarking. Being explicit about these constraints enhances the methodological transparency and credibility of the research while clarifying the scope within which its contributions can be interpreted. Future research should therefore incorporate longitudinal deployment, cross-domain participant samples, objective performance metrics, and evaluation with fully implemented AI systems under real-world laboratory conditions (beyond simulated AI interaction).

7 Conclusion

This thesis set out to investigate how a conversational interaction paradigm can be integrated into Electronic Laboratory Notebooks (ELN) and how such an interaction concept is perceived within scientific documentation practice. Motivated by identified limitations in current ELN interaction models—particularly the lack of interactive scaffolding, limited cognitive support, and minimal conversational affordances—the study focused on the experiential dimension of AI-like interaction rather than on technical AI performance.

To address this objective, a conversational ELN prototype was designed based on literature-informed considerations, prior contextual insights, and theoretical perspectives from human–computer interaction, distributed cognition, and computer-supported collaborative learning (CSCL). The prototype was evaluated using a Wizard-of-Oz methodology in a realistic laboratory setting. This approach enabled controlled investigation of user experience, workload perception, and trust calibration without relying on fully implemented AI components. A mixed-methods design combined quantitative instruments (System Usability Scale, NASA-TLX, adapted trust items) with qualitative interviews and observational data, allowing both descriptive measurement and in-depth exploration of interaction experiences.

Across measures, the results indicate that the conversational interaction paradigm was perceived as more usable than traditional notebook practices under simulated conditions. Structured turn-taking, explicit confirmations, and immediate feedback improved the clarity and coherence of documentation tasks. Voice-supported interaction reduced perceived physical demand and effort, particularly in situations where manual documentation interrupts experimental activity. Although these findings reflect idealized technical conditions, they demonstrate that interaction structure can meaningfully influence how documentation effort is experienced.

Qualitative findings highlight the importance of multimodality and user control. Participants combined voice and graphical interaction strategically, indicating that conversational ELN should function as augmentation systems that preserve agency and procedural accountability rather than as autonomous documentation tools. Transparency and feedback clarity emerged as central determinants of trust calibration. Trust was grounded in interaction reliability and predictability rather than assumptions of artificial intelligence capability. While conversational framing occasionally triggered anthropomorphic interpretations, such responses did not translate into perceived artificial agency.

In relation to collaborative work, participants perceived that standardized retrieval and summarization mechanisms may support research continuity and shared understanding across experiments. Interpreted through a CSCL lens, structured conversational scaffolding may contribute to common ground formation. However, collaborative enhancement was assessed subjectively and was not validated through objective performance indicators.

The primary contribution of this thesis lies in isolating and empirically examining the interaction layer of conversational ELN independent of backend AI implementation. By employing a Wizard-of-Oz architecture, the study demonstrates how interaction paradigms can be evaluated prior to full technological realization and clarifies which experiential design factors shape user perception in scientific documentation contexts.

Taken together, the findings suggest that conversational ELN hold potential as augmentation tools within laboratory environments when designed with transparency, multimodality, structured feedback, and human oversight in mind. However, the study remains bounded by simulated AI conditions, a limited sample size, and short-term exposure. Further longitudinal research with fully implemented AI systems is required to determine whether the experiential advantages identified here translate into sustained improvements in documentation quality, collaborative continuity, and workflow efficiency.

7.1 Future work

Based on the evaluation and discussion, this thesis makes several contributions to research on Simulated AI ELN and human–AI collaboration. First, it provides empirical evidence that conversational and voice-based interaction can support collaborative learning by improving transparency, shared context, and research continuity, positioning ELN as collaborative learning artifacts rather than purely administrative tools.

Second, the thesis extends distributed cognition perspectives by showing how Simulated AI ELN can function as cognitive partners that externalize memory and structure information, while maintaining human control. The findings underline that effective cognitive distribution depends on predictability and transparency, not automation alone.

Third, the thesis offers design-relevant insights into managing cognitive and physical workload through multimodal interactions. Voice-based input, combined with visual confirmation and user control, can reduce physical effort without introducing excessive cognitive demand.

Finally, the study here demonstrates the value of Wizard-of-Oz methodology for investigating human–AI interaction concepts in early-stage research tools. This approach enabled systematic exploration of usability, workload, trust, and collaboration prior to full technical implementation. Taken together, these contributions position AI-enabled ELN as socio-technical infrastructures that actively shape collaborative scientific practice and open new directions for research and design.

VII. Bibliography

Reference List

2026 Biotech AI Report | Benchling (2026a). Retrieved January 22, 2026, from https://www.benchling.com/biotech-ai-report-2026.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., et al. Guidelines for Human-AI Interaction. In (pp. 1–13).

Analytical Science Article DO Series (2026). *Not Electronic – Not a Notebook - 2008 - Wiley Analytical Science.* Retrieved January 02, 2026, from https://analyticalscience.wiley.com/content/article-do/not-electronic-not-notebook.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction, 24*(6), 574–594.

Barchard, K. A., & Pace, L. A. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Computers in Human Behavior, 27*(5), 1834–1839.

Bush, V. As We May Think, from https://web.mit.edu/sts.035/www/PDFs/think.pdf.

Cai, C. J., Reif, E., Hegde, N., Hipp, J., Kim, B., Smilkov, D., et al. Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making. In (pp. 1–14).

Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021). *Expanding Explainability: Towards Social Transparency in AI systems,* from https://arxiv.org/pdf/2101.04719.

Gates, R. S., McLean, M. J., & Osborn, W. A. (2015). Smart Electronic Laboratory Notebooks for the NIST Research Environment. *Journal of research of the National Institute of Standards and Technology, 120*, 293–303.

Guerrero, S., Dujardin, G., Cabrera-Andrade, A., Paz-Y-Miño, C., Indacochea, A., Inglés-Ferrándiz, M., et al. (2016). Analysis and Implementation of an Electronic Laboratory Notebook in a Biomedical Research Institute. *PloS one, 11*(8), e0160428, from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0160428.

Guerrero, S., López-Cortés, A., García-Cárdenas, J. M., Saa, P., Indacochea, A., Armendáriz-Castillo, I., et al. (2019). A quick guide for using Microsoft OneNote as an electronic laboratory notebook. *PLoS computational biology, 15*(5), e1006918, from https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006918.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: integrating empirical evidence on factors that influence trust. *Human factors, 57*(3), 407–434.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition. *ACM Transactions on Computer-Human Interaction, 7*(2), 174–196.

Humans in the Loop: The Design of Interactive AI Systems | Stanford HAI (2026b). Retrieved February 12, 2026, from https://hai.stanford.edu/news/humans-loop-design-interactive-ai-systems.

Kadi4Mat (2025). Retrieved January 03, 2026, from https://kadi.iam.kit.edu/.

Kanza, S., Willoughby, C., Gibbins, N., Whitby, R., Frey, J. G., Erjavec, J., et al. (2017). Electronic lab notebooks: can they replace paper? *Journal of cheminformatics, 9*(1), 31.

Kizilcec, R. F. How Much Information? In (pp. 2390–2395).

Körber, M. Theoretical Considerations and Development of a Questionnaire to Measure Trust in Automation. In (pp. 13–30).

Lab Notebooks | JILA - Exploring the Frontiers of Physics (2026c). Retrieved January 29, 2026, from https://jila.colorado.edu/lewandowski/research/lab-notebooks-0.

Leung, B., Shimabukuro, M., & Collins, C. (10132024). NeuroSight: Combining Eye-Tracking and Brain-Computer Interfaces for Context-Aware Hand-Free Camera Interaction. In L. Yao, M. Goel, A. Ion, & P. Lopes (Eds.), *The 37th Annual ACM Symposium on User Interface Software and Technology* (pp. 1–3). New York, NY, USA: ACM.

Machina, H. K., & Wild, D. J. (2013). Electronic laboratory notebooks progress and challenges in implementation. *Journal of laboratory automation, 18*(4), 264–268.

Meinecke, A., Heidrich, D., Dworatzyk, K., & Theis, S. A Comparative Heuristic Evaluation of Kadi4Mat Through Human Evaluators and GPT-4. In (pp. 91–108).

Menke, J., Homberg, S., & Koch, O. (2023). Introduction to artificial intelligence and deep learning using interactive electronic programming notebooks. *Archiv der Pharmazie, 356*(7), e2200628.

NE08F1-4, from https://humansystems.arc.nasa.gov/groups/tlx/downloads/TLXScale.pdf.

Post | Feed | LinkedIn (2026d). Retrieved February 16, 2026, from https://www.linkedin.com/feed/update/urn:li:activity:7308419948563861505/.

ResearchGate (2026). *Vygotsky's Zone of Proximal Development.* Retrieved February 12, 2026, from https://www.researchgate.net/publication/383563118_Vygotsky's_Zone_of_Proximal_Development#full-text.

Rosala, M., & Paul, S. (2024, April 19). The Wizard of Oz Method in UX. Retrieved January 05, 2026, from https://www.nngroup.com/articles/wizard-of-oz/.

Schnell, S. (2015). Ten Simple Rules for a Computational Biologist's Laboratory Notebook. *PLoS computational biology, 11*(9), e1004385, from https://pmc.ncbi.nlm.nih.gov/articles/PMC4565690/.

Steinfeld, A., Jenkins, O. C., & Scassellati, B. The oz of wizard. In (pp. 101–108).

Sweller, J. Cognitive Load Theory. In (pp. 37–76).

Terry, G., Hayfield, N., Clarke, V., & Braun, V. Thematic Analysis. In (pp. 17–36).

van Eikeren, P. (2004). Intelligent Electronic Laboratory Notebooks for Accelerated Organic Process R&D. *Organic Process Research & Development, 8*(6), 1015–1023.

Yagamurthy, D. N. (2023). Advancements in Natural Language Processing (NLP) and Its Applications in Voice Assistants and Chatbots. *Journal of Artificial Intelligence & Cloud Computing*, 1–6.

Zhang, A. X., Muller, M., & Wang, D. (2020). *How do Data Science Workers Collaborate? Roles, Workflows, and Tools,* from https://arxiv.org/pdf/2001.06684.

VIII. Appendix

VIII.I eLab DLR Survey Questionnaire

Electronic Laboratory Notebooks - Survey

Welcome to our survey on the use of electronic laboratory notebooks (ELNs). As part of a project to introduce a universally available electronic laboratory notebook, we would like to understand how laboratory notebooks are currently used at DLR in order to best adapt the electronic laboratory notebook to existing needs.

With this survey, we would like to find out how you currently use laboratory notebooks, be it in analog or digital form. Your experiences and insights will help us to adapt the electronic laboratory notebook to your requirements and are therefore very important at this stage of the project.

Thank you for taking the time to share your thoughts and contribute to the further development of digital laboratory notebooks in the DLR scientific community.

There are 31 questions in this survey.

Background Information

Understanding your scientific background helps us tailor electronic laboratory notebook improvements and features to your diverse scientific fields and practices. Therefore, we want to ask you some questions regarding your background.

Please select your age group:

Choose one of the following answers
Please choose **only one** of the following:

- ◯ < 18 years
- ◯ 18-25 years
- ◯ 26-35 years
- ◯ 36-45 years
- ◯ 46-55 years
- ◯ 56-65 years
- ◯ > 65 years
- ◯ Other [ ]

What is your gender?

Choose one of the following answers
Please choose **only one** of the following:

- ◯ Female
- ◯ Male
- ◯ Non-binary
- ◯ Other [ ]

What is your highest level of education?

Choose one of the following answers
Please choose **only one** of the following:

- ◯ High School Diploma
- ◯ Bachelor
- ◯ Master
- ◯ PhD
- ◯ Other

In which discipline is your highest academic qualification? Which is your field of research?

Choose one of the following answers
Please choose **only one** of the following:

- ◯ Not applicable, I work in administration
- ◯ Aeronautical & Manufacturing Engineering
- ◯ Agriculture & Forestry
- ◯ Astronomy
- ◯ Biological Sciences
- ◯ Chemical Engineering
- ◯ Chemistry
- ◯ Civil Engineering
- ◯ Computer Science
- ◯ Communication & Media Studies
- ◯ Economics
- ◯ Electrical Electronic Engineering
- ◯ Geology
- ◯ Geography & Environmental Science
- ◯ Information Management
- ◯ Mathematics
- ◯ Mechanical Engineering
- ◯ Medicine
- ◯ Pharmacy
- ◯ Physics
- ◯ Politics
- ◯ Psychology
- ◯ Sociology
- ◯ Other

How many years of research related work do you have?

Choose one of the following answers
Please choose **only one** of the following:

- Less than 1 year
- 1-3 years
- 4-6 years
- 7-10 years
- 11- 15 years
- more than 15 years

Are you responsible for a laboratory?

Choose one of the following answers
Please choose **only one** of the following:

- Yes
- No
- Other

General use of Electronic Laboratory Notebooks (ELNs)

In this section, we aim to gather insights into whether you primarily use analog (paper-based) or digital (electronic) laboratory notebooks. Additionally, we want to know if electronic devices are available in your lab, which types of devices are accessible, and your preferences when it comes to using either format.

What type of laboratory notebook are you currently using?

*

Choose one of the following answers
Please choose **only one** of the following:

- Analog (paper-based)
- Digital - please list the software of any electronic laboratory notebook you use in the comment box
- Both - please list the software of any electronic laboratory notebook you use in the comment box
- Neither

Make a comment on your choice here:

Please rate how frequently you use a laboratory notebook in the following setting

*

Please choose the appropriate response for each item:

	Never	Occasionally (less than once a month)	Regularly (once a month to once a week)	Frequently (several times a week)	Constantly (daily)
In the laboratory	○	○	○	○	○
In the office	○	○	○	○	○
During field experiments	○	○	○	○	○
In team meetings or collaborative sessions	○	○	○	○	○
While working remotely (e.g., home, hotel)	○	○	○	○	○

For each device listed below, rate their current availability for research purposes:

*

Please choose the appropriate response for each item:

	1 - not available	2	3	4	5 - always available
Pen and Paper	○	○	○	○	○
Desktop Computers	○	○	○	○	○
Laptops	○	○	○	○	○
Tablets	○	○	○	○	○
Smartphones	○	○	○	○	○

Rate your preference for each type of device to use a laboratory notebook with:

*

Please choose the appropriate response for each item:

	1 - not preferred	2	3	4	5 - highly preferred
Pen and Paper	○	○	○	○	○
Desktop Computers	○	○	○	○	○
Laptops	○	○	○	○	○
Tablets	○	○	○	○	○
Smartphones	○	○	○	○	○

Are there any additional devices you want to use to interact with an electronic laboratory notebook? Please list them below:

Please write your answer here:

Offline Functionality

In this section, we aim to explore the importance of offline functionality when using an electronic laboratory notebook (ELN). We want to understand your needs regarding working without an internet connection and how critical access to your laboratory notebook data is to maintaining productivity.

Offline functionality refers to the ability of software to function without an internet connection, allowing users to access, input, and manage data locally. In the case of electronic laboratory notebooks (ELNs), this includes working within local networks and syncing with central systems once online connectivity is restored.

How important is it for your work to have access to your laboratory notebook data in situations without internet connection?

*

Please choose the appropriate response for each item:

	1 - not at all important	2	3	4	5 - extremely important
	○	○	○	○	○

Offline functionality refers to the ability of software to function without an internet connection, allowing users to access, input, and manage data locally. In the case of electronic laboratory notebooks (ELNs), this includes working within local networks and syncing with central systems once online connectivity is restored.

How often do you use laboratory notebooks in situations without internet access?

*

Please choose the appropriate response for each item:

	Never	Occasionally (less than once a month)	Regularly (once a month to once a week)	Frequently (several times a week)	Constantly (daily)	Irregular, but intense during periods
	○	○	○	○	○	○

Offline functionality refers to the ability of software to function without an internet connection, allowing users to access, input, and manage data locally. In the case of electronic laboratory notebooks (ELNs), this includes working within local networks and syncing with central systems once online connectivity is restored.

Rate the importance of being able to perform the following actions offline:

*

Please choose the appropriate response for each item:

	1 - not at all important	2	3	4	5 - extremely important
Viewing entries	○	○	○	○	○
Creating new entries	○	○	○	○	○
Editing existing entries	○	○	○	○	○
Deleting entries	○	○	○	○	○
Adding attachments (e.g., images, files)	○	○	○	○	○
Viewing attachments	○	○	○	○	○
Searching within the notebook	○	○	○	○	○

Offline functionality refers to the ability of software to function without an internet connection, allowing users to access, input, and manage data locally. In the case of electronic laboratory notebooks (ELNs), this includes working within local networks and syncing with central systems once online connectivity is restored.

Describe situations where access to your laboratory notebook is critical, but internet connection is not available or available to a limited extent:

Please write your answer here:

Offline functionality refers to the ability of software to function without an internet connection, allowing users to access, input, and manage data locally. In the case of electronic laboratory notebooks (ELNs), this includes working within local networks and syncing with central systems once online connectivity is restored.

After being offline, how should your data preferrably synchronize after reconnecting to the internet?

Choose one of the following answers

Please choose **only one** of the following:

- ◯ Manual selection: You choose specific entries to update
- ◯ Automatic synchronization: All changes update automatically
- ◯ Combination of both: Specify which data to auto-update and which to manually select
- ◯ Other

Offline functionality refers to the ability of software to function without an internet connection, allowing users to access, input, and manage data locally. In the case of electronic laboratory notebooks (ELNs), this includes working within local networks and syncing with central systems once online connectivity is restored.

Collaboration & Network

In this section, we aim to explore the aspects of collaboration that are most important to you when using an electronic laboratory notebook (ELN). We are interested in understanding how you share data, work with team members, and ensure smooth communication within collaborative projects.

Collaboration is the process of individuals or teams working together to achieve a shared goal. In ELNs, this involves sharing data, communicating efficiently, and tracking contributions, allowing seamless teamwork across different locations and ensuring transparency.

How frequently do you collaborate with other parties on research?

*

Please choose the appropriate response for each item:

	Never	Occasionally (less than once a month)	Regularly (once a month to once a week)	Frequently (several times a week)	Constantly (daily)
	◯	◯	◯	◯	◯

Collaboration is the process of individuals or teams working together to achieve a shared goal. In ELNs, this involves sharing data, communicating efficiently, and tracking contributions, allowing seamless teamwork across different locations and ensuring transparency.

Rate how frequently you use the following methods to collaborate with colleagues

*

Please choose the appropriate response for each item:

	Never	Occasionally (less than once a month)	Regularly (once a month to once a week)	Frequently (several times a week)	Constantly (daily)
In-person meetings	○	○	○	○	○
Email	○	○	○	○	○
Shared digital documents (e.g. Google Docs, Teamsite, Onenote)	○	○	○	○	○
Online messaging platforms (e.g. Mattermost, Slack)	○	○	○	○	○
Video conferencing (e.g. Skype, Zoom)	○	○	○	○	○
Electronic laboratory notebook with collaboration features	○	○	○	○	○

Collaboration is the process of individuals or teams working together to achieve a shared goal. In ELNs, this involves sharing data, communicating efficiently, and tracking contributions, allowing seamless teamwork across different locations and ensuring transparency.

How important is the ability to collaborate with colleagues within an Electronic La

*

Please choose the appropriate response for each item:

	1 - not at all important	2	3	4	5 - extremely important
	○	○	○	○	○

Collaboration is the process of individuals or teams working together to achieve a shared goal. In ELNs, this involves sharing data, communicating efficiently, and tracking contributions, allowing seamless teamwork across different locations and ensuring transparency.

Rate the importance of these collaboration features:

*

Please choose the appropriate response for each item:

	1 - not at all important	2	3	4	5 - extremely important
Real-time co-editing	○	○	○	○	○
Access Controls and Permissions	○	○	○	○	○
Traceability of Authorship	○	○	○	○	○
Commenting and Annotation Tools	○	○	○	○	○
Integrated Messaging or Chat	○	○	○	○	○
Task Assignment and Management	○	○	○	○	○

Collaboration is the process of individuals or teams working together to achieve a shared goal. In ELNs, this involves sharing data, communicating efficiently, and tracking contributions, allowing seamless teamwork across different locations and ensuring transparency.

Describe how you typically would collaborate with other scientists.

Please write your answer here:

Collaboration is the process of individuals or teams working together to achieve a shared goal. In ELNs, this involves sharing data, communicating efficiently, and tracking contributions, allowing seamless teamwork across different locations and ensuring transparency.

Data Integrity

This section seeks to identify what aspects of data integrity matter most to you when working with electronic laboratory notebooks (ELNs). We're interested in your views on preventing data changes, ensuring data accuracy, and maintaining a traceable record of all modifications.

Data Integrity: Data integrity refers to maintaining and assuring the accuracy, consistency, and reliability of data over its lifecycle. This includes ensuring that data is not altered or tampered with, and that it remains accurate and consistent across all systems and processes where it is stored, processed, and utilized.

How important is data integrity in your use of an Electronic Laboratory Notebook (ELN)?

*

Please choose the appropriate response for each item:

	1 - not at all important	2	3	4	5 - extremely important
	○	○	○	○	○

Data Integrity: Data integrity refers to maintaining and assuring the accuracy, consistency, and reliability of data over its lifecycle. This includes ensuring that data is not altered or tampered with, and that it remains accurate and consistent across all systems and processes where it is stored, processed, and utilized.

Rate the following data integrity features based on their importance for your work: *

Please choose the appropriate response for each item:

	1 - not at all important	2	3	4	5 - extremely important
Version Control: Ability to track changes and maintain previous versions of data entries, ensuring that historical data is preserved and can be audited.	○	○	○	○	○
Time Stamping: Automatic time stamps on all entries and modifications to provide a clear timeline of data creation and changes.	○	○	○	○	○
Error Checking: Mechanisms to detect and correct errors in data entries to maintain accuracy.	○	○	○	○	○
Data Validation: Ensuring that the entered data align with the collected data.	○	○	○	○	○
Immutable Records: Protection against unauthorized alterations, ensuring that once data is entered, it cannot be modified without proper authorization.	○	○	○	○	○
Consistency Checks: Regular checks to ensure data consistency across different parts of the system.	○	○	○	○	○
Electronic Signatures: Digital signatures to authenticate the identity of users and validate the integrity of data entries.	○	○	○	○	○
Redundancy: Use of redundant systems to ensure that data remains available and intact even if part of the system fails.	○	○	○	○	○

Data Integrity: Data integrity refers to maintaining and assuring the accuracy, consistency, and reliability of data over its lifecycle. This includes ensuring that data is not altered or tampered with, and that it remains accurate and consistent across all systems and processes where it is stored, processed, and utilized.

Describe any concerns you have regarding data integrity when using Electronic L

Please write your answer here:

Data Integrity: Data integrity refers to maintaining and assuring the accuracy, consistency, and reliability of data over its lifecycle. This includes ensuring that data is not altered or tampered with, and that it remains accurate and consistent across all systems and processes where it is stored, processed, and utilized.

Data Security

In this section, we aim to explore which aspects of data security are most important to you when using an electronic laboratory notebook (ELN). We want to understand your concerns regarding unauthorized access, data loss, and the secure transfer of information.

Data Security: Data security involves protecting digital data from unauthorized access, corruption, or theft throughout its lifecycle. This includes implementing measures like encryption, access controls, firewalls, and security protocols to ensure that sensitive information is safeguarded against cyber threats and breaches.

How important is data security in your use of an Electronic Laboratory Notebook *

Please choose the appropriate response for each item:

	1 - not at all important	2	3	4	5 - extremely important
	○	○	○	○	○

Data Security: Data security involves protecting digital data from unauthorized access, corruption, or theft throughout its lifecycle. This includes implementing measures like encryption, access controls, firewalls, and security protocols to ensure that sensitive information is safeguarded against cyber threats and breaches.

Rate the following data security features based on their importance for your work with an electronic laboratory notebook (ELN): *

Please choose the appropriate response for each item:

	1 - not at all important	2	3	4	5 - extremely important
Encryption: End-to-end encryption to protect data during transmission and storage.	○	○	○	○	○
Access Controls: Role-based access controls (RBAC) to ensure that only authorized personnel can access or modify data.	○	○	○	○	○
Authentication: Multi-factor authentication (MFA) to add an extra layer of security for user logins.	○	○	○	○	○
Audit Trails: Detailed logs of all user activities within the ELN to identify any unauthorized access or changes.	○	○	○	○	○
Data Backup: Regular and automated backups to prevent data loss due to accidental deletion or system failures.	○	○	○	○	○
Compliance: Adherence to relevant regulations and standards (e.g., GDPR, HIPAA) to ensure data privacy and security.	○	○	○	○	○
Firewall and Intrusion Detection: Network security measures to prevent unauthorized access and detect potential threats.	○	○	○	○	○
Secure Data Sharing: Controlled and encrypted methods for sharing data with collaborators.	○	○	○	○	○

Data Security: Data security involves protecting digital data from unauthorized access, corruption, or theft throughout its lifecycle. This includes implementing measures like encryption, access controls, firewalls, and security protocols to ensure that sensitive information is safeguarded against cyber threats and breaches.

Describe any concerns you have regarding data security when using Electronic Laboratory Notebooks (ELNs):

Please write your answer here:

Data Security: Data security involves protecting digital data from unauthorized access, corruption, or theft throughout its lifecycle. This includes implementing measures like encryption, access controls, firewalls, and security protocols to ensure that sensitive information is safeguarded against cyber threats and breaches.

Feedback on the Use of Electronic Laboratory Notebooks (ELNs)

In this section, we would like to understand your preferences regarding specific features of electronic laboratory notebooks (ELNs). We want to know which features you find most useful, what might discourage you from using an ELN, and what additional functionalities you would like to see in the future.

What features or functionalities
of Electronic Laboratory Notebooks (ELNs) encourage you to use it?
Please describe any specific features or functionalities.

Please write your answer here:

What features or functionalities
of Electronic Laboratory Notebooks (ELNs) discourage you from using it? Please d

Please write your answer here:

What features or functionalities
would you wish an Electronic Laboratory Notebook (ELN) to have?
Please describe any specific features or functionalities.

Please write your answer here:

Is there anything else you would like to add or suggest that has not been covered in the previous questions regarding the use and features of electronic laboratory notebooks (ELNs)?

Please write your answer here:

Thank you for taking part in our survey!

Further information on our DLR project can be found on our GitLab page, where you have the opportunity to tell us your individual wishes and ideas for electronic laboratory books at DLR.

You can also join our DLR-Mattermost Channel, where we want to connect everyone who is interested in electronic laboratory notebooks at DLR.

In addition, we are planning a community workshop in March for anyone interested in electronic laboratory notebooks, especially Kadi4Mat, be it as (potential) users, or people involved in the development or on the administrative side. We are pleased to invite you to our upcoming event. This is a great opportunity to network, share insights and grow together. Please join us to get the most out of our community:
Register for our event here!

25.02.2025 – 12:42
Submit your survey.
Thank you for completing this survey.

VIII.II Calling for Participants Advertisement

Participants Needed for a Thesis Study!

Looking for 10 participants

Study Topic

AI-Assisted Laboratory Documentation
(Voice Assistant + Chatbot for Electronic Lab Notebooks)

Who Can Participate?

I am specifically looking for:
✓ Students, researchers, or PhD candidates
✓ Who work in **laboratory environments**
✓ And have experience using **laboratory notebooks** (paper or electronic)

What You Will Do

A short **Wizard-of-Oz session** where you interact with a simulated AI assistant for documenting lab work.
You will use a prototype ELN interface and perform simple documentation tasks.
No preparation needed.

Duration

45–60 minutes

Compensation

10 euros OR **a free Mensa meal**

Location

A suitable **lab** at the University
(Will be confirmed individually after sign-up)

Contact

Ahire, Tanishq
Master's Student – Human-Computer Interaction
University of Siegen
tanishq.ahire@student.uni-siegen.de

VIII.III Figma Prototype

https://uni-siegen-masterthesis-1785236.figma.site/

VIII.IV Evaluation Study WoZ Guideline

Wizard-of-Oz Study – Participant Guideline Packet

1. Welcome and Purpose of the Study
Thank you for participating in this study.
This research investigates a conversational interaction concept for an Electronic Laboratory Notebook (ELN). The study explores how researchers experience voice and chat-based interaction for documenting experiments and retrieving research information.
Please note:

- The system you will interact with simulates AI-like behavior.
- The study focuses on interaction experience, not on evaluating actual AI performance.
- There are no right or wrong actions during the tasks.
- We are evaluating the system — not you.

Your participation contributes to research in Human–Computer Interaction (HCI) and Computer-Supported Collaborative Learning (CSCL).

2. What You Will Do
The session consists of two parts:
Part A – Baseline Condition (Traditional Documentation)
You will complete documentation tasks using a traditional laboratory notebook format.
This allows comparison between conventional documentation and the conversational prototype.
Part B – Conversational ELN Prototype (Simulated AI Interaction)
You will use a tablet-based prototype that includes:

- Voice-based documentation
- Chat-based information retrieval
- AI-generated summaries (simulated)

You may:

- Speak naturally when using voice input
- Ask the chatbot questions in text form
- Interact freely with the interface

The system may respond with structured confirmations, clarifications, or summaries.

3. Important Information About the System
This prototype uses a Wizard-of-Oz setup, meaning:

- Some intelligent responses are simulated.
- The system is designed to behave as if it were AI-assisted.
- The goal is to evaluate the interaction concept under ideal technical conditions.

You do not need to adjust your behavior — interact naturally.

4. Tasks Overview
You will be asked to:

1. Document experimental steps.
2. Record observations and parameter values.
3. Retrieve information from prior entries.
4. Request a summary of an experiment.
5. Ask clarification questions via chatbot.

The tasks are designed to resemble realistic laboratory documentation scenarios.
If anything is unclear, you may ask the moderator.

5. Think-Aloud Protocol
During the interaction, you are encouraged to briefly verbalize:

- What you expect the system to do
- If something feels confusing
- If something feels helpful
- If something feels unnecessary

There is no need to explain everything — short reflections are sufficient.

6. After the Tasks
You will complete:

- System Usability Scale (SUS)
- NASA-TLX (Workload)
- Perceived System Trustworthiness questionnaire

- Short post-interview discussion

Your honest feedback is highly valuable.

7. Ethical Considerations

- Your participation is voluntary.
- You may withdraw at any time without giving a reason.
- Your responses will be anonymized.
- No personal data will be published.
- The session may be audio-recorded for research purposes only.

All data will be stored securely and used exclusively for academic research.

8. Practical Information
Duration: approx. 45–60 minutes
Devices used: Tablet prototype
Environment: Controlled laboratory simulation setup
If you experience discomfort at any point, please inform the moderator.

9. Before We Begin
Please confirm that:
☐ You have read and understood the study information
☐ You signed the consent form
☐ You have no open questions

VIII.V WoZ Task Sheet

Wizard-of-Oz Study – Task Sheet (Updated)

Task 1: Voice Assistant Documentation
Use the voice assistant to record a short observation from your experiment.
Task 2: Retrieve Past Experiment Summary
Ask the AI chatbot to summarize the last experiment involving growth medium preparation.
Task 3: Collaborative Knowledge Retrieval
Ask the AI chatbot about a colleague's previous experimental activity (e.g., who conducted the experiment, when it was performed, or what was documented).

Instructions
• Perform each task in order.
• Think aloud as you interact with the system.
• There are no right or wrong answers—focus on your natural workflow.

VIII.VI SUS Scale

System Usability Scale (SUS)

	Strongly Disagree 1	2	3	4	Strongly Agree 5
I think that I would like to use this product frequently.					
I found the product unnecessarily complex.					
I thought this product was easy to use.					
I think that I would need the support of a technical person to be able to use this product.					
I found the various functions in this product were well integrated.					
I thought there was too much inconsistency in this product.					
I would imagine that most people would learn to use this product very quickly.					
I found this product very awkward to use.					
I felt very confident using this product.					
I needed to learn a lot of things before I could get going with this product.					

VIII.VII NASA TLX Scale

Figure 8.6

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

VIII.VIII TiA Trust in AI Scale

		Strongly disagree	Rather disagree	Neither disagree nor agree	Rather agree	Strongly agree	No response
1	The system is capable of interpreting situations correctly.	(1)	(2)	(3)	(4)	(5)	○
2	The system state was always clear to me.	(1)	(2)	(3)	(4)	(5)	○
3	I already know similar systems.	(1)	(2)	(3)	(4)	(5)	○
4	The developers are trustworthy.	(1)	(2)	(3)	(4)	(5)	○
5	One should be careful with unfamiliar automated systems.	(1)	(2)	(3)	(4)	(5)	○
6	The system works reliably.	(1)	(2)	(3)	(4)	(5)	○
7	The system reacts unpredictably.	(1)	(2)	(3)	(4)	(5)	○
8	The developers take my well-being seriously.	(1)	(2)	(3)	(4)	(5)	○
9	I trust the system.	(1)	(2)	(3)	(4)	(5)	○
10	A system malfunction is likely.	(1)	(2)	(3)	(4)	(5)	○
11	I was able to understand why things happened.	(1)	(2)	(3)	(4)	(5)	○
12	I rather trust a system than I mistrust it.	(1)	(2)	(3)	(4)	(5)	○
13	The system is capable of taking over complicated tasks.	(1)	(2)	(3)	(4)	(5)	○
14	I can rely on the system.	(1)	(2)	(3)	(4)	(5)	○
15	The system might make sporadic errors.	(1)	(2)	(3)	(4)	(5)	○
16	It is difficult to identify what the system will do next.	(1)	(2)	(3)	(4)	(5)	○
17	I have already used similar systems.	(1)	(2)	(3)	(4)	(5)	○
18	Automated systems generally work well.	(1)	(2)	(3)	(4)	(5)	○
19	I am confident about the system's capabilities.	(1)	(2)	(3)	(4)	(5)	○

Questionnaire „Trust in Automation" (TiA) | Moritz Körber, Technical University of Munich

VIII.IX Pre-Post Questionnaire

Pre & Post Interview Sheet

Pre-Interview Questions

1. Have you used an Electronic Laboratory Notebook before?
2. How do you typically document experiments in your workflow?
3. What do you expect an AI assistant in an ELN could help you with?

Post-Interview Questions

1. How easy or difficult was it to use the prototype?
2. Which interactions felt intuitive? Which were confusing?
3. How helpful did the AI responses feel?
4. Did the system behave in a way you could trust?
5. Would you use a tool like this in your actual lab workflow?
6. Anything you wish the system had done differently?

VIII.X Consent form

Participant Consent Form

Study Title: AI-Assisted Electronic Laboratory Notebook (ELN) – Wizard-of-Oz User Study

Researcher: Tanishq Ahire
Institution: University of Siegen / DLR

Purpose of the Study
You are invited to participate in a research study evaluating an early prototype of an AI-assisted Electronic Laboratory Notebook. Your participation will help us understand usability, workload, trust, and user perception of AI-supported lab documentation workflows.

What Participation Involves
• Completing short tasks using a simulated prototype (Wizard-of-Oz setup)
• Filling out questionnaires (SUS, NASA-TLX, Trust in Automation)
• A short pre- and post-interview

Voluntary Participation
Your participation is completely voluntary. You may withdraw at any time without penalty.

Data Handling & Privacy
All collected data will be anonymized and used solely for academic research purposes. Your identity will not be linked to the results.

Risks & Benefits
There are no known risks associated with this study. You may not receive direct benefits, but your input will contribute to improving AI-assisted research tools.

Confidentiality
All responses will be kept confidential. No personal identifying information will appear in publications or reports.

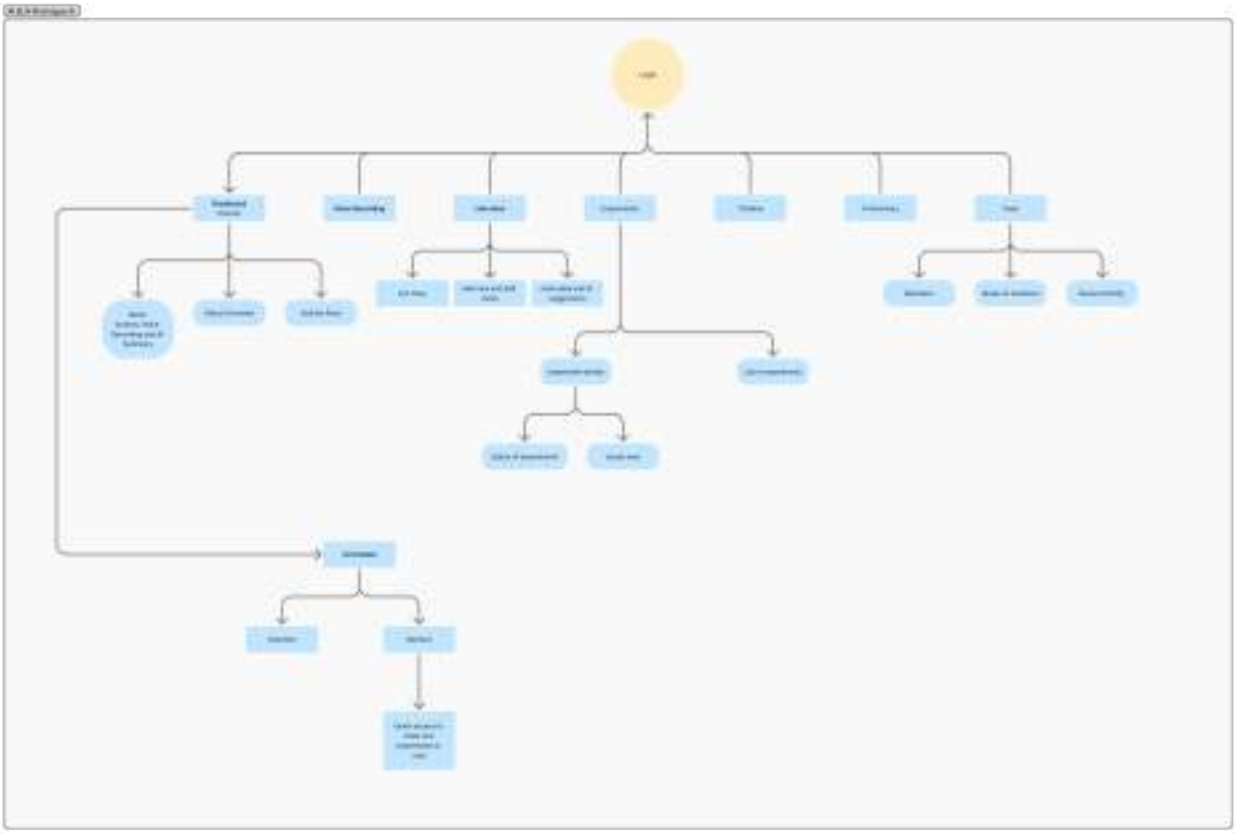
Consent Statement
By signing below, I confirm that:
• I have read and understood the information provided.
• I voluntarily agree to participate in this study.
• I understand I may withdraw at any time.

Participant Name: ____

Signature: ____

Date: __________________________

VIII.XI Information Architecture of Prototype



VIII.XII Personal Notes from Kadi4Mat Community Meeting and POliS Lab Visit

Lightning Talks – Personal Notes

- Lightning talks showed how differently Kadi4Mat is used across labs — for some it's part of daily documentation, for others more of a project archive.
- A recurring point was that documentation often happens after experiments, not during them.
- Several speakers mentioned difficulties when trying to understand older experiments, especially if they were conducted by someone else.
- Collaboration features exist, but many researchers said they don't always have a clear overview of what others are working on.

Daniel's presentation (stood out to me):

- Daniel talked about performing experiments with lasers and particles during parabolic zero-gravity flights.
- Each flight parabola lasts only a few minutes, and during that time the entire focus is on executing the experiment.
- He mentioned that taking notes during the flight was practically impossible — both because of the extreme time pressure and because of experiencing weightlessness.
- As a result, experimental details had to be reconstructed afterwards, relying on memory and post-flight discussions.
- This example really stayed with me, as it showed an extreme case where traditional note-taking simply does not work during critical experimental moments.

Workshop Thematic Stations – Personal Notes

- Moving between stations made it clear that many labs develop their own workarounds around ELNs.
- Several users mentioned keeping parallel notes in personal documents or messaging tools.
- There was a visible tension between wanting flexibility and needing structured documentation.
- AI support was generally seen as promising, but concerns were raised about:
    - correctness of AI-generated content,
    - losing control over documentation,
    - responsibility for what gets recorded.
- I noticed that users focused more on speed and ease of use, while developers emphasized structure and traceability.

Onboarding and training stood out as a recurring issue:

- Multiple participants mentioned that onboarding new researchers to the ELN takes time and effort.
- The user interface was often described as unfamiliar and disconnected from tools people are used to.

- Because the structure and navigation were not intuitive, new users struggled to understand:
  - where to start documenting,
  - where to find existing experiments,
  - how different pages or sections were connected.
- This lack of familiarity meant that training relied heavily on verbal explanations or informal guidance from colleagues.
- It became clear to me that the ELN was not perceived as user-centered in its interaction design.
- This reinforced the importance of designing an ELN interface that feels familiar, using:
  - clear buttons,
  - page-based structures,
  - and navigation patterns similar to common mobile applications.
- A more familiar interaction model would help users quickly understand when and where to go to find information, reducing training effort and lowering the barrier for adoption.

Informal Discussions – Personal Notes

- In informal conversations, many researchers said they often ask colleagues directly instead of checking ELN entries.
- New team members struggle to understand why certain experimental decisions were made.
- ELNs were frequently described as static archives rather than active tools during daily work.
- Several people liked the idea of asking questions to an ELN instead of manually searching through entries.
- Voice interaction was seen as useful in principle, but clearly dependent on the situation and environment.

POLiS Lab Visit – Personal Notes

- The lab environment involved constant interaction with instruments, materials, and safety equipment.
- Direct interaction with computers during experiments was often impractical.
- Documentation was usually postponed until after experimental steps were completed.
- I observed a lot of verbal coordination between team members during experiments.
- Switching between hands-on work and documentation seemed cognitively demanding.
- Being in the lab helped me better understand why documentation often lacks context that is obvious to the people performing the experiment.

VIII.XIII Participants

Table 7: [Participants Recruited]

Participant ID	Gender	Age	Role	Lab Experience (years)	Prior Documentation Practice	Prior ELN Experience
P01	23	Male	MSc	1–2	Paper + Word	No
P02	23	Male	MSc	2–3	Paper	No
P03	27	Female	MSc	2–3	Word + Excel	Limited
P04	26	Male	MSc	3–4	Word + OneNote	Limited
P05	28	Male	MSc	3–4	Paper + Excel	No
P06	25	Female	MSc	3–5	Word/OneNote	Limited
P07	24	Male	MSc	4–5	Paper + Word	No
P08	27	Male	MSc	4–6	Word + Excel	Limited
P09	32	Male	PhD	6–8	Paper + Word	No
P10	29	Female	PhD	10	ELN + scripts	Yes

VIII.XIV Pre-defined AI responses template for the Wizard

Wizard-of-Oz Response Library

Confidential – For Wizard Use Only

General Response Principles

- Responses must be structured, concise, and neutral.
- No exaggerated intelligence or improvisation.
- No proactive suggestions unless predefined.
- Maintain 1–3 second response delay.
- Avoid fillers such as 'umm' or human-like hesitation.

Voice Documentation Responses

- Confirmation: 'Entry recorded. Step added to Experiment Log.'
- Parameter Save: 'Parameters saved. Step successfully recorded.'
- Observation: 'Observation added to current experiment entry.'
- Structured Reformatting: System reformats informal input and asks for confirmation.
- Missing Parameter Prompt: Suggest adding temperature, duration, or equipment (trigger once per task block).
- Error Prevention: Prompt if conflicting values detected.

Chatbot Information Retrieval

- Retrieve experiment: Provide title, date, and main parameters.
- Summary: 3–4 structured sentences, neutral tone.
- Parameter query: Provide exact recorded value.
- If no data: 'No anomaly was recorded in the selected experiment.'

AI Summary Generation

- Generate short structured summary of current session.
- Do not overstate confidence or predictive ability.

Transparency Responses

- Explain suggestions as based on documentation standards.
- Keep explanations short and procedural.

Human-in-the-Loop Responses

- 'You may edit or override any generated entry.'
- Do not force user decisions.

Failure / Limitation Responses

- 'That functionality is not available in the current system.'
- 'I did not fully understand. Could you please repeat the step?'

Strict Do-Not-Do List

- Do NOT appear overly intelligent.
- Do NOT provide scientific recommendations.
- Do NOT invent reasoning.
- Do NOT use emotional tone or humor.

Response Style Summary

- Tone: Structured, neutral, predictable.
- Length: 1–4 sentences maximum.
- Goal: Simulate ideal conversational ELN interaction, not advanced AI.

VIII.XIII Declaration and Documentation of AI Tool Usage

In accordance with the University of Siegen's guidelines for dealing with AI-based language models, I declare that I have used the following AI tools during the research and writing of this thesis (where applicable)

Tool Name	Provider	Version / Platform	Access Type
ChatGPT	OpenAI	GPT-4 / GPT-4.1 (web interface)	User account

AI tool was used exclusively as a supportive writing aid for linguistic refinement, rephrasing, and structural clarification of already developed text, such as improving academic tone, coherence between paragraphs, and clarity of formulations. The AI tool was not used to generate original scientific content, conduct data analysis, derive results, interpret findings, or make methodological, theoretical, or design decisions. All intellectual contributions, evaluations, interpretations, and final formulations remain the sole responsibility of the author. All AI-assisted outputs were critically reviewed, verified for correctness and consistency, and substantially revised prior to integration into the thesis

Essen, 16.02.2026 TANISHQ TUSHAR AHIRE

____

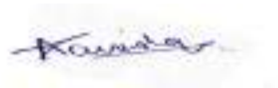
Place, date, signature

IX. Affidavit

Affidavit of Assurance

I declare on oath, that I have composed the present thesis independently. I only have used the sources and means specified in this thesis. Especially from the internet, I only have used the denoted references. I declare that I have identified any literal or analogous copying from other work and the use of AI-based language tools as such. I have taken note of the section in the examination regulations concerning attempts to cheat.

I confirm that the electronic version of the thesis which I deliver is identical to the printed version with respect to the content. I agree that an electronic version of the thesis will be stored for purposes of inspection of plagiarism



Essen, 16.02.2026 TANISHQ TUSHAR AHIRE

Place, date, signature