

MASTERARBEIT | MASTER'S THESIS

Titel | Title

POTATO: Explainable AI for Business Automation
Evaluation of the POTATO Framework at Kontron
Transportation GmbH to Enhance Work Efficiency with
Customizable AI Solution – Use Case Scenario

verfasst von | submitted by
Anna Hierholcz BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 915

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Betriebswirtschaft

Betreut von | Supervisor:

ao. Univ.-Prof. i.R. Mag. Dr. Josef Windsperger

Erklärung zur Verfassung der Arbeit

Anna Hierholcz, B.Sc.

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig verfasst habe. Sämtliche verwendeten Quellen und Hilfsmittel sind vollständig angegeben. Insbesondere versichere ich, dass ich alle wörtlichen und sinngemäßen Übernahmen aus anderen Werken als solche kenntlich gemacht habe.

Wien, 31. Jänner 2025

Anna Hierholcz

Acknowledgements

I would like to thank my supervisor, ao. Univ.-Prof. i.R. Mag. Dr. Josef Windsperger, and Dr. Gábor Recski, for their guidance and support throughout this year as I worked on my thesis. I am also grateful to all the employees at Kontron Transportation GmbH who participated in the meetings and contributed to my use case scenario, especially Mehdi Masinovic for his valuable insights and assistance.

Abstract

The proposed work explores the business potential of a novel human-in-the-loop explainable AI method, the POTATO framework. Through a use case scenario, this master's thesis aims to demonstrate the resource-saving potential of a POTATO-based text processing solution, while also assessing and improving the platform's usability and customizability within the context of Kontron Transportation GmbH's onboarding ticket management process.

Efficiency gains were assessed using ADONIS simulations, which compared manual ticket preparation with the automated process. In the current manual process, employees solve tickets with the help of a ticket database where they manually search, categorize, and identify relevant tickets. This process constitutes the ticket preparation phase, involving the gathering of necessary information to support resolution. In the automated version, POTATO filters and categorizes these tickets automatically based on predefined characteristics, such as solved or unsolved tickets. This helps to shorten the time spent on manual searching. The results demonstrated significant improvements in both task completion time and cost savings. User experience was evaluated through UX testing with employees of Kontron Transportation GmbH, conducted within the framework of a Design Thinking approach. The testing of the software focuses on four key categories: understandability, navigation and usability, autonomy, and customizability. The testing showed that users had a strong understanding of the platform's functionalities, and customizability was positively rated, aligning well with Kontron Transportation GmbH's needs. However, navigation and autonomous software usage emerged as areas requiring improvement to enhance the overall user experience. While POTATO offers clear benefits, targeted refinements are needed to fully realize its business potential.

Abstract

Die vorliegende Masterarbeit untersucht das geschäftliche Potenzial einer neuartigen, menschenzentrierten und erklärbaren KI-Methode – des POTATO-Frameworks. Anhand eines konkreten Anwendungsfalls soll gezeigt werden, welches Einsparungspotenzial eine POTATO-basierte Textverarbeitungslösung bietet. Gleichzeitig werden die Benutzerfreundlichkeit und Anpassbarkeit der Plattform im Kontext des Ticket-Management-Prozesses von Kontron Transportation GmbH bewertet und optimiert.

Zur Analyse der Effizienzgewinne wurden ADONIS-Simulationen durchgeführt, die den manuellen Ticketvorbereitungsprozess von Kontron Transportation GmbH mit der automatisierten Variante verglichen. In der bisherigen manuellen Bearbeitung greifen Mitarbeitende auf eine Ticket-Datenbank zu, in der sie relevante Tickets manuell suchen, kategorisieren und identifizieren. Diese Phase der Ticketvorbereitung dient der Sammlung notwendiger Informationen zur Problemlösung. In der automatisierten Version übernimmt POTATO diese Aufgaben eigenständig, indem es Tickets anhand vordefinierter Merkmale – wie „gelöst“ oder „ungelöst“ – filtert und kategorisiert. Dadurch verkürzt sich die Zeit, die für manuelle Suchprozesse aufgewendet werden muss. Die Ergebnisse der Simulationen zeigen deutliche Verbesserungen sowohl in der Bearbeitungszeit als auch bei den Kostenersparnissen.

Die Nutzererfahrung wurde durch UX-Testing mit Mitarbeitenden von Kontron Transportation GmbH evaluiert, die im Rahmen eines Design-Thinking-Ansatzes durchgeführt wurden. Dabei lag der Fokus auf vier zentralen Aspekten: Verständlichkeit, Navigation und Benutzerfreundlichkeit, Autonomie sowie Anpassbarkeit. Die Tests ergaben, dass die Nutzer die Funktionen der Plattform gut verstanden und insbesondere die Anpassbarkeit positiv bewerteten. Allerdings wurden Schwierigkeiten in der Navigation und der eigenständigen Nutzung der Software festgestellt, was Verbesserungspotenzial für die Gesamtnutzererfahrung aufzeigt.

Obwohl POTATO klare Vorteile bietet, sind gezielte Optimierungen erforderlich, um das volle geschäftliche Potenzial einer POTATO-basierten Lösung auszuschöpfen.

Contents

Abstract	iv
Abstract	v
1 Introduction	1
1.1 About the Use Case Scenario Problem Statement and Potential Solution	1
1.2 Research Focus - Finding USP of POTATO	2
1.3 Motivation and Research Gap	3
1.4 Research Question	3
2 Background and Related Work	5
2.1 What is AI XAI & the challenges	5
2.1.1 About POTATO	6
2.2 What is Usability & User Experience	7
2.2.1 What Is Usability Testing & Goals of Usability Testing	7
2.2.2 Why is customizability important?	8
2.3 About Kontron Transportation GmbH	9
3 Research Method	10
3.1 Design Thinking	10
3.2 UX Testing Process	13
3.3 ADONIS - Business Process Simulation	14
4 Design Thinking	15
4.1 Emphasize and Define - Adonis	15
4.1.1 Mapping Existing Processes - Interview on Ticket Management Process at Kontron Transportation GmbH	16
4.1.2 Emphaty Map	18
4.1.3 User Persona	19
4.1.4 ADONIS Simulation - POTATO - Manual Version	21
4.1.5 User Problem Formulation	24
4.2 Ideate	26
4.2.1 MSCW Model	26
4.2.2 Automated Simulation	28
	vi

4.2.3	Comparison: manual vs automated simulation results	33
4.3	Prototype	35
4.4	Evaluation	39
4.4.1	Evaluation Approach	39
4.4.2	Background Questionnaire Results	41
4.4.3	Tasks Results	42
4.4.4	User Preference Data	48
4.4.5	Evaluation Summary	58
5	Conclusion	60
5.1	Limitation and Future Work	61
	List of Figures	61
	List of Tables	63
	Bibliography	64

Introduction

This chapter introduces the use case scenario and presents the background for the research purpose and motivation. Additionally, it defines the research focus, highlights POTATO's unique selling points, outlines the research question and method, and presents the thesis structure.

1.1 About the Use Case Scenario | Problem Statement and Potential Solution

The rule-based HITL XAI framework used in this thesis is called POTATO (exPlainable infOrmation exTrAcTion framewOrk), developed at TU Vienna. It is currently part of the university's startup incubator program. POTATO automates text processing with explainable AI, providing white-box solutions with human-in-the-loop learning (Kovács et al., 2022). In collaboration with Kontron Transportation GmbH, POTATO was tested on a specific business task.

To understand the research context, it is essential to mention the motivations behind involving POTATO and Kontron Transportation GmbH in the study and the implementation of the research. One of the daily tasks at Kontron Transportation GmbH involves a manual processing procedure, specifically related to ticket management. This manual process includes searching, categorizing, and identifying relevant tickets within a database to support ticket resolution. This process seems to be stretching employees' work hours and adding to their overall workload, which can hurt productivity. Automating part of the current manual ticket management process with POTATO could boost productivity and shorten the time spent on these tasks.

Facing this business challenge that POTATO could potentially address, Kontron Transportation GmbH partnered with us to explore this possibility. As POTATO is an emerging

software not yet available on the market, it also sought a pilot program to evaluate both its technical and business potential.

Within the Use Case Scenario of this thesis, two activities were undertaken: First, UX testing was conducted with employees of Kontron Transportation, where the current version of POTATO was tested as a prototype, addressing the company's ticket management process. The focus was on its practical application, with particular attention given to its customizability. The testing focused on four categories: understandability, usability and navigation, autonomy, and customizability.

Secondly, after analyzing their current manual process, the change in work efficiency between the manual and automated versions through a simulation was measured. As mentioned before, the manual process involves employees searching through a ticket database to find similar past cases and solutions, also categorizing tickets, and identifying relevant ones manually. Meaning, the company currently handles a daily flow of tickets that need to be solved. To assist with this, an existing database provides access to previously logged tickets, which may include similar issues, some of them are solved and others unsolved but still potentially useful for resolving the current ticket. In contrast, the automated process with POTATO classifies and filters tickets automatically based on predefined characteristics, like whether they are solved or unsolved. This automation can significantly reduce the time needed for ticket preparation and shorten the workload for employees. This thesis focuses its simulation solely on the ticket preparation phase, not the full ticket resolution process (*note: ticket preparation includes assigning priorities, reviewing ticket details, and searching the database for similar cases to support the resolution*) This phase was chosen because the automation efforts are specifically targeted at this part of the workflow. Moreover, the time needed to solve tickets is so widely spread that calculating a reliable average for the "solving" phase would not have been feasible.

Detailed descriptions of the use-case scenario and an introduction to Kontron Transportation GmbH are provided in Chapters 4 and 2, respectively.

1.2 Research Focus - Finding USP of POTATO

The research focus encompasses two major themes. The first is the real-world application of POTATO, with an emphasis on customizability. While customizability was a key focus, other aspects were also evaluated for a well-rounded evaluation. The second is work efficiency, particularly in terms of the software's resource-saving potential. These two areas were identified as the unique selling points (USPs) of POTATO. UX testing was conducted to determine how effectively the software aligns with this specific business cases and to assess the user-friendliness of its interface and functionality. A detailed description of the UX testing process can be found in Chapter 3.2. Work efficiency was measured using the ADONIS Business Process Management tool, which simulated Kontron Transportation GmbH's business tasks in both analog and automated formats. The results of these simulations are analyzed in Chapter 4 (Karagiannis et al., 1996b).

1.3 Motivation and Research Gap

Technological advancements have granted extensive access to data, yet analyzing it presents both opportunities and challenges. One of these challenges concerns artificial intelligence (AI) and machine data processing. While many research fields and everyday applications of AI emphasize the benefits and opportunities, there are currently opaque functionalities that hinder the traceability of decisions. In critical sectors such as healthcare or finance, inaccurate predictions can have serious consequences (Streib et al., 2020). With the spread of AI in these critical areas, human limitations in understanding complex AI models become a problem. The main reason for this is that insights from the data and the tasks solved by machines are hidden in increasingly complex models. Users would need access to vast amounts of data to explain a deep network. Therefore, these models often act like "black boxes," making their decision-making processes challenging to comprehend. However, explainable AI (XAI) focuses on making the functioning of these AI systems more transparent, aiming to explain the decision-making processes of these systems and provide a better understanding of how an outcome was achieved (Dwivedi et al., 2023).

And here comes POTATO into play by addressing the complexity of these XAI systems. XAI often refers to black-box models, like neural networks, where it's unclear how decisions are made, and explanations are based on input and output analysis. POTATO, however, uses a transparent rule-based XAI system. Through its transparent and understandable analysis functions, it facilitates decision transparency. As mentioned before, POTATO automates text processing processes with explainable AI and introduces a framework that offers white-box solutions with human-in-the-loop learning. White-box solutions in AI are transparent, letting users see and understand how decisions are made. Human-in-the-loop (HITL) learning involves humans in the training process, improving the system's accuracy and relevance through their feedback (Kovács et al., 2022). Further details on XAI and how POTATO differs are explored in Chapters 2.1 and 2.1.1.

Beyond offering new solutions from a technical perspective, POTATO's other unique selling point is its customizability. This means that: The software interface can be tailored to a specific company's business case, focusing on their needs, priorities and tasks. Furthermore, since the primary function of POTATO from a technical standpoint is to automate current manually operated tasks, its usage may result in increased work efficiency (Kovács et al., 2022).

1.4 Research Question

After gaining insight into the increased significance of XAI and how the new human-in-the-loop learning paradigm of POTATO offers promising business potential in the XAI market, the aim of the master's thesis is to address the following three research questions:

R1: How does the utilization of POTATO for text processing tasks enhance work efficiency compared to manual text processing methods?

To measure efficiency gains, the ADONIS Business Process Management tool will be utilized. This tool makes it easy to simulate both manual and automated text processing workflows, streamlining data input and enabling in-depth analysis. The focus is on defining the steps of the process, allocating resources, and including time and cost data. ADONIS provides detailed analyses and visual outputs, which highlight important metrics like lead times, cycle times, and costs. By comparing these simulations, the results will clearly show how POTATO improves workflow efficiency, both in terms of percentages and cost savings (Karagiannis et al., 1996b).

R2: In the context of POTATO's current UX/UI design, which features can be enhanced and adapted to foster a more user-friendly experience for individuals interacting with the platform?

To answer Research Question 2, UX testing will be carried out with employees of Kontron Transportation GmbH. The research is based on the following literature: "Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests" by Jeffrey and Dana (2008); Peffers et al.'s (2007) "Design Science Research Methodology for Information Systems Research" and "Development of User Interface/User Experience using Design Thinking Approach for GMS Service Company," by Bambang Suratno and Jihan Shafira (2022).

R2.1: How well is POTATO customizable for the collaboration company's specific business task?

Given that one of POTATO's key strengths is its customizability and ability to adapt the interface to specific business needs, Research Question 2.1 focuses on this feature to provide a detailed analysis of how well POTATO performs in terms of customizability. The evaluation mainly takes a qualitative approach. However, since it is also of interest to research whether the implementation of POTATO can improve work efficiency, quantitative evaluation is included in the thesis as well, particularly when analyzing output data from ADONIS simulations.

Note: Qualitative evaluation will address aspects such as graph structure, process descriptions, and workflow efficiencies, while quantitative evaluation will examine process performance metrics, time durations, and costs.

Background and Related Work

2.1 What is AI|XAI & the challenges

Artificial Intelligence (AI) has quickly become a key part of modern technology, influencing many areas of daily life and industry. AI focuses on building systems that can handle tasks like learning, reasoning, problem-solving, and decision-making—things typically done by humans. These systems are widely used in areas like healthcare, finance, and transportation, where they help drive progress and improve efficiency (Du-Harpur et al., 2020). As AI becomes more complex and influential in critical decisions, there's a growing need for transparency and understanding. These calls have become louder since 2021, and the demand for these frameworks has increased, especially in early 2023, after the release of ChatGPT (Alexander et al., 2023).

So in recent years, concerns have arisen due to instances where AI models produced biased or discriminatory results. These issues have caused problems for major companies and raised doubts about their effectiveness and actual reliability (Kale, 2023). To give some examples, Microsoft's Twitter bot named "Tay" was originally supposed to entertain people. However, within less than 24 hours, Tay started making racist and offensive comments. Another example occurred in 2015 with a self-driving Tesla that failed to recognize a street sweeper in autopilot mode and was involved in a fatal accident. Similar incidents happened in 2018 with a self-driving Uber that killed a woman in Arizona. These recent examples illustrate that despite technological advancements, AI models are far from being developed to the point where they can be without restriction (Kale, 2023).

AI systems often work as "black boxes," which creates serious challenge, meaning there is a lack of transparency in how decisions are made (Rudin Radin, 2019). This is where Explainable Artificial Intelligence (XAI) comes in. The goal of XAI is to provide insights into how AI models make decisions, enhancing trust, accountability, and fairness in AI applications. In its general form, an explanation in XAI is any information that helps the user understand why a model shows certain decision patterns and how individual

decisions are made (Miller, 2021). However, the term "explanation" does not have a single, clear definition and can mean different things. Therefore, various approaches have been undertaken to address this issue (Nayak, 2019).

The aim of this thesis is not to delve into the detailed technical aspects of AI and XAI, nor to present the various methods of XAI. However, it is important to briefly mention what exactly XAI does, beyond its goal of making AI decision-making traceable. The methods of XAI are diverse, because there is no clear consensus on what "explainable" means. Most existing methods attempt to explain models by influencing the input data and observing how this affects the outputs. In other words, what happens to the output (decision) when I change my input data (Nayak, 2019)? Reading this sentence raises the question of how much companies' trust can be based on the traceability of AI decisions, which primarily relies on the "play" of input and output data.

2.1.1 About POTATO

Having explored the challenges of AI and XAI, let's now examine how the main focus of this thesis, "POTATO," stands out and what unique value it provides.

When people talk about XAI, they usually think of a standard AI model—a large neural network, a black box—where we don't know why it does what it does, and we try to extract explanations using various tools and draw conclusions from them. This was introduced in the previous paragraph as the "play" with the input output data. In contrast, POTATO uses a rule-based system, which is not a black box, yet it still belongs to the category of XAI. The ExPlainable Information Extraction Framework, or POTATO, is an approach to information extraction that prioritizes human involvement in the learning process. POTATO includes human expertise in the process, combining automated methods with human insights to ensure meaningful explanations. Unlike most natural language processing (NLP) methods, which work directly with textual data, POTATO utilizes graph-based representations to interpret language. This framework supports various graph formats, like Abstract Meaning Representation (AMR, Banarescu et al. 2013), 4lang (Kornai, 2019) and Universal Dependencies (UD, Nivre et al., 2018). POTATO converts text into graph structures, enabling the creation of rule-based text classifiers that use graph patterns instead of raw text. Its core feature is suggesting rules for data classification using interpretable machine learning models like decision trees. These models rank rules by evaluating subgraphs and compare the number of positive and negative matches in training data. Users can create rule lists for each class label. These rules are applied to graph representations of sentences, classifying them based on whether a pattern matches as a subgraph. Essentially, POTATO's rule system is a collection of these graph patterns.

POTATO has shown its effectiveness in various real-world applications. These include relation extraction in English texts (Kovács et al., 2022), offensive content detection on social media (Gémes et al., 2021), and the digitization of building permit processes in Vienna's zoning plans (Recski et al., 2024). As an open-source framework, POTATO offers flexibility and adaptability across different tasks and languages.

In summary, POTATO stands out as a task- and language-independent framework. It incorporates human input into the rule-learning process and uses graph-based features to deliver explainable AI solutions. By focusing on explainability and user engagement, POTATO addresses the limitations of traditional black-box models, offering a more transparent and interpretable approach to information extraction (Kovács et al., 2022).

2.2 What is Usability & User Experience

Usability became a widely used term in the early 1980s, evolving from concepts like "user friendliness" and "ease of use." Despite its long history, defining usability has been challenging. Usability involves how effectively, efficiently, and satisfactorily users can achieve their goals in specific environments. It is not a fixed attribute but depends on the interaction between users, tasks, and environments. The goal of usability engineering has been to create products that users find effective, efficient, and satisfactory. This involves an iterative design process where designs are tested and refined based on user feedback to identify and fix usability problems (Lewis & Sauro, 2021).

User Experience (UX) extends the concept of usability to include all aspects of a user's interaction with a product, focusing also on emotional outcomes. While usability deals with practical aspects like task completion, UX also considers emotional factors. Before 2000, those developing products and systems for human use mainly focused on classical usability. This meant checking how well designs allowed users to complete tasks quickly and successfully while keeping them satisfied. Around the year 2000, the focus began to shift towards UX, that considered emotional factors such as pleasure, beauty, and trust. This broader perspective on usability emphasizes not just functionality but also the emotional and behavioral impacts of using a product. In other words, UX aims to enhance user satisfaction by addressing both practical and emotional needs.

Over time, the importance of UX has grown. Organizations like the Usability Professionals Association (now the User Experience Professionals Association) have emphasized the need to go beyond usability and include emotional engagement (Lewis & Sauro, 2021).

2.2.1 What Is Usability Testing & Goals of Usability Testing

Usability testing involves observing how users interact with a product to identify any usability issues. These tests can vary from informal setups, where a tester sits next to the participant and notes any problems, to more formal environments, where participants work alone while being observed through one-way glass.

The main purpose of usability testing is to see how real users use a product and to find areas where they struggle. This helps designers to improve the product to make it more user-friendly. What makes usability testing different from other user-centered design methods is that it involves direct observation of users performing tasks in a controlled setting (Lewis & Sauro, 2021).

There are two main types of usability testing: formative and summative. Formative Usability Testing focuses on identifying and fixing usability problems during the development

process. This type of testing is more informal and involves interaction between observers and participants. The goal is to discover and prioritize issues by observing how often problems occur, their likelihood during normal use, and their impact on the user. On the other hand summative Usability Testing aims to evaluate the effectiveness, efficiency, and satisfaction of a product after its development. This type of testing is more formal and structured. It involves measuring how well the product meets predefined usability objectives or comparing it to other products. These goals aren't in direct conflict, but they focus on different things, which can change how tests are done. For example, focusing on measurement usually leads to more formal testing with less interaction between observers and participants. In contrast, focusing on finding problems often results in more informal testing with more interaction between observers and participants (Lewis & Sauro, 2021). Our user testing for POTATO primarily follows the formative testing approach, as the product is still under development, and a prototype has been created for Kontron Transportation GmbH. The focus was on testing and improving the current version. During the testing sessions, participants were allowed to ask questions and seek assistance. However, the completion of questionnaires and the collection of ratings were conducted in a structured, controlled, and predefined manner.

2.2.2 Why is customizability important?

Since one of the research questions focuses on how customizable POTATO is, and customization is highlighted as one of the software's unique selling points, this paragraph will explore why it is important and beneficial for an AI tool to be customizable. Additionally, to present a balanced perspective, the challenges associated with customization will also be mentioned.

General AI tools are often designed to be broadly applicable, which means they might not address the specific needs of different industries or applications. This can lead to decontextualization, where tools become less effective because they do not take into account the unique challenges and needs of specific domains or organizations.

There are several advantages to customizing AI tools. Firstly, it makes these tools more relevant for addressing specific project challenges. Customization also promotes stakeholder engagement by involving various team members in the process. This helps align goals and expectations, fostering a more inclusive approach to AI development. Additionally, customizable AI tools are also flexible, so practitioners can adjust them to fit changing project needs. Furthermore tailoring AI tools is important for identifying and reducing fairness-related harms. Generic frameworks might not be sufficient for tackling specific fairness and ethical issues, so they must be adapted to the context in which they are used. Customization helps teams develop a common understanding of AI fairness, which is important for a better collaboration across different fields. For instance, AI tools used in healthcare need to meet specific regulations like HIPAA, which requires customization to ensure compliance (Madaio et al., 2024).

However, customization can also be challenging. It adds complexity and increases the workload for AI developers, who need to decide which parts of the tools to keep, change, or remove. These decisions can be difficult and time-consuming. Not all AI developers

have the skills or experience needed for effective customization, which can lead to gaps in knowledge. Accountability is another issue, as it is often unclear who is responsible for ensuring that AI tools are customized effectively and ethically (Madaio et al., 2024). In summary, customizability technology is intended to provide users with content that in some way matches users' long-term predispositions and preferences. While customization can add complexity and challenges, from a business perspective, it also increases the appeal of the product by offering tailored solutions to specific needs. This flexibility makes the tool more attractive to a broader range of customers and increases the chances of selling the product to various domains or organizations with unique requirements.

2.3 About Kontron Transportation GmbH

In this chapter, Kontron Transportation GmbH is introduced.

To understand the company and the business processes relevant to the use case scenario, an interview was conducted with an employee working as a Product Architect in the company. Additionally, the company's website served as a supplementary source of information. First, a general introduction to the company is provided, followed by a detailed description of its ticket management processes and operations in Section 4.1.1. Kontron Transportation, headquartered in Vienna, is a leading provider of end-to-end communication solutions for mission-critical industries and carrier-grade networks. With more than 690 employees and 11 sites across Europe, the company supports customers in over 25 countries. Its primary focus is the railway industry, providing safe communication systems for operators throughout Europe and beyond.

End-to-end communication solutions mean that Kontron Transportation manages every step of the process, including planning, development, production, deployment, maintenance, and operation. This integrated approach ensures that their networks are seamless, reliable, and customized to meet the specific needs of their clients, particularly in safety-critical environments such as railways. Kontron specializes in GSM-R, the Global System for Mobile Communications-Railways, a standardized communication system specifically designed to ensure efficient coordination between trains and control centers. Additionally, the company is expanding into Public Safety LTE, a modern communication technology for emergency services, and private mobile radio systems based on TETRA (Terrestrial Trunked Radio) and DMR (Digital Mobile Radio). Both TETRA and DMR are widely used for secure communication in industries requiring high reliability and safety. With more than 90 customers and revenues exceeding €150 million, Kontron Transportation is a trusted partner for critical communication needs. The company continuously invests in innovation (ETCS | Kontron Transportation, o. D.).

Research Method

This chapter outlines the research methods used in the study, focusing on the Design Thinking approach as the guiding framework for problem-solving and user-centered design (Suratno Shafira, 2022). A significant part of the research involved usability testing, where participants completed predefined tasks with POTATO, related to their ticket management process. After completing each task, participants filled out questionnaires using a Likert scale. Observations were also conducted during the testing to document behaviors, problem-solving strategies. Additionally, the ADONIS business process management tool will be presented, which was used to simulate and compare the company's manual and automated ticket management process.

3.1 Design Thinking

Design Thinking is a method that started in professional design practices in the late 1960s and became widely used in commercial design by the 1990s. It focuses on understanding user needs and improving designs based on user feedback through an iterative process. A very similar approach to Design Thinking is User-Centered Design. User-centered design (UCD) and Design Thinking share similarities in that both methods aim to produce initial designs, followed by iterative processes for design improvement. So both methods involve users from the beginning of the design process and use their feedback to enhance the product. However, Design Thinking places a greater emphasis on understanding users' emotional and experiential needs (Lewis Sauro, 2021).

This research applies the Design Thinking approach to guide the project, involving structured steps that create a systematic UX testing framework. Before detailing the steps of the Design Thinking process, it is important to mention that various literature sources sometimes name or define these steps slightly differently. However, broadly speaking, Design Thinking generally consists of the same five steps. First, a general

overview will be provided. Following that, Peffers et al. (2007b) methodology will be integrated into the basic model, based on the work of Suratno and Shafira (2022b).

The five steps of Design Thinking, according to Suratno and Shafira (2022b), are the following:

Empathize

The Empathize stage is the first step in the Design Thinking process. It involves gathering insights into users' needs through methods such as interviews and questionnaires. For example, in the study, an interview with key personnel was conducted, which resulted in an empathy map to highlight users' needs and challenges.

Define

In the Define stage, the information collected during the Empathize phase is analyzed to identify core problems. Tools like User Persona or User Journey Maps, which outline users' interactions with the current system, are used. This helps in narrowing down the focus to specific issues that need addressing.

Ideate

During the Ideate stage, a wide range of ideas and solutions are brainstormed to overcome and find potential solutions for the previously identified challenges and issues. This creative phase enables to find innovative solutions to the identified problems. Methods such as Use Case Diagrams and Flowcharts can be employed.

Prototype

The Prototype stage involves creating a preliminary version of the solution. If possible, these prototypes can be developed through new software development. However, tools like Figma can also be used to design a prototype.

Test During the testing phase, the prototype is evaluated by real users to collect feedback and assess usability. This phase provides measurable data on the design's effectiveness and user satisfaction, and the feedback is used to guide further refinements and improvements.

After providing an overview of the five steps of Design Thinking, the model by Peffers et al. (2007) will be introduced to compare it with the model by Suratno and Shafira (2022b). For a better understanding, Table 1 provides a closer look at the similarities between these approaches in how they define the steps for Design Thinking:

Design Thinking Steps (Suratno and Shafira, 2022b)	Design Science Research Methodology Steps (Peppers et al., 2007)	Similarities
Empathize	Problem Identification and Motivation	Both methodologies start with understanding the problem.
Define	Problem Identification and Motivation, Define Objectives of a Solution	Both emphasize clearly defining the problem and setting objectives based on user perspectives.
Ideate	Design and Development	Both involve generating potential solutions to address identified issues.
Prototype	Demonstration	Both include creating a tangible version of the solution for testing
Test	Evaluation	Both methodologies involve evaluating the effectiveness of the solution. A trial on the prototype
	Communication	Communication step is specific to DSRM and involves sharing findings.

Table 1: Comparison of Design Thinking and DSRM
(Own representation based on Suratno and Shafira, 2022b; Peppers et al., 2007)

In the Design Science Research Methodology (DSRM) by Peppers et al. (2007), the first stage is **Problem Identification and Motivation**. This involves identifying and defining the research problem and justifying its importance. The goal here is to ensure that the problem is significant and worth addressing. This stage aligns with both the Empathize and Define stages in the Design Thinking model by Suratno and Shafira (2022b), where the focus is on understanding and clearly defining the problem from the user's perspective. Next, the **Define Objectives of a Solution** stage is about establishing the objectives that the solution must achieve. These objectives are intended to provide criteria for what constitutes a successful solution. This stage corresponds to the Define stage in Suratno and Shafira's Design Thinking model, where clear objectives are set for solving the identified problem. The **Design and Development** stage involves creating the artifact, which could be a model, method, or system, to address the problem.

This stage is similar to the Ideate and Prototype stages in Suratno and Shafira's Design Thinking model, where ideas are generated and turned into tangible prototypes. Following this, the **Demonstration** stage is where the use of the artifact is demonstrated in a relevant context to show how it solves the problem. This involves applying the solution in a practical scenario, akin to testing prototypes in the Test stage of Suratno and Shafira's Design Thinking model. The **Evaluation** stage involves assessing how well the artifact solves the problem. This stage focuses on evaluating the effectiveness of the solution, similar to the Test stage in both Design Thinking models, where the usability and effectiveness of the solution are evaluated based on user feedback. Finally, the **Communication** stage involves communicating the problem, the developed artifact to relevant audiences, including researchers, or other stakeholders. While this stage doesn't have a direct match with Suratno and Shafira's Design Thinking model, it covers the iterative sharing and feedback cycle.

At their core, both Design Science Research Methodology (DSRM) and Design Thinking share a similar structure and purpose. Even though they use different terms and have slightly different focuses they follow a similar path. Both guide you through identifying a problem, developing solutions, testing them, and refining them based on feedback to get the best outcome.

3.2 UX Testing Process

As previously introduced, the company currently has a ticket management process, certain steps of which could be automated to improve efficiency. At present, employees search within a database for similar tickets that can serve as a basis for solving their new tickets. This part of the process is referred to as **ticket preparation**, and it was simulated using the ADONIS tool in both manual and automated versions. Further details on this can be found in the following chapters. The UX testing, which focused on the actual implementation of POTATO, centered on the **training phase** for the software. Without proper training, the software, POTATO, would not be usable for their use case. Therefore this training phase was tested with the company's employees. For this, predefined tasks related to their ticket management process were created for participants to complete, followed by a questionnaire to gather feedback on their experience.

Users were required to complete three predefined tasks, and after completing each task, they filled out the same questionnaire to measure changes in their evaluations over time. Google Forms was chosen as the survey platform due to its accessibility, user-friendly interface, and data management capabilities. The objective of the questionnaire, was to gather insights in the following four categories of POTATO: Understandability, Usability and Navigation, Autonomy, and Customizability.

Understandability measured how clear and understandable the software functions were. Navigation and Usability focused on the intuitive use of the software, specifically how user-friendly and easy it was to use POTATO. The Autonomy category was closely related,

but here the focus was on how confident users felt when performing tasks independently within the software. Finally, Customizability assessed how well the prototype covered and reflected the specific business tasks. Each category included an average of 4-5 statements, which participants rated on a Likert Scale from Strongly Disagree (1) to Strongly Agree (5). Five employees from Kontron Transportation GmbH participated in the user testing. They were observed individually as they completed the testing, with each candidate spending an average of 40 minutes in the testing room.

Participants in the testing were employees at various stages of their careers, and they varied in gender and age. In addition to the survey, we strengthened our research methodology in two ways: first, users completed a pre-questionnaire to gather general information about them. Second, while they performed the tasks, we observed them, documenting both the time taken to complete tasks and their behavior, problem-solving strategies, and any other useful occurrences during the testing.

A detailed description of the Testing, its implementation, and results can be found in Chapter 4.4.

3.3 ADONIS - Business Process Simulation

ADONIS, a business process management tool, utilized for process modeling, analysis, and simulation. ADONIS facilitates the visual mapping and optimization of workflows within the organization and also helps to identify inefficiencies (Karagiannis et al., 1996b). Various data were collected from Kontron Transportation GmbH, including information on the company's goals and objectives, workflow steps, and user roles and responsibilities.

The simulation process unfolded in two main stages: Initially, the current analog business processes of the organization were simulated using ADONIS. This method helped understand workflow dynamics and pinpoint inefficiencies. Specifically, the process involved ticket management, where employees manually searched, categorized, and identified relevant tickets in a database to support issue handling. Subsequently, these processes were replicated using POTATO, focusing on automating the classification and filtering of tickets. The simulation aimed to highlight POTATO's advantages over traditional methods by reducing manual effort.

The comparative analysis focused on evaluating efficiency gains in terms of time and cost savings. This methodological approach provided valuable insights into process optimization and also emphasized the practical advantages of solutions like POTATO.

Design Thinking

Chapter 4 introduces the Design Thinking methodology as a framework for analyzing ticket management process at Kontron Transportation GmbH. The Methodology of the Design Thinking Chapter draws on two key pieces of literature as resources, and follows the themes presented in these research papers. The referenced works are: Peffers et al. (2007b) and Suratno and Shafira (2022b).

The chapter starts by identifying challenges in the existing manual workflow through interviews and an empathy map, followed by the creation of user personas. It then moves into the simulation of the current manual ticket management process using ADONIS to pinpoint inefficiencies. From there, potential automation solutions are explored, leading to the implementation of an automated version with POTATO. Finally, the chapter evaluates the impact of automation by comparing the manual and automated processes and presenting UX testing results to assess key aspects such as usability, navigation, autonomy, and customizability.

4.1 Emphasize and Define - Adonis

In the first phase, the goal was to map out the current internal processes at Kontron Transportation GmbH, with a specific focus on identifying challenges and difficulties that could be addressed through automation, achieved via an interview. The interview was conducted with one of Kontron's Product Architect, who is currently working in the IT automation field. Based on this interview, an empathy map was created, along with a simulation of Kontron's business processes using the ADONIS business process management tool. Not just a task sequence but also a questionnaire were developed for the testing session. The empathy map aimed not only to evaluate and represent the objective flow of the business process but also to better uncover and address user needs. The problem identification stage includes the aforementioned ADONIS simulation, which serves as a detailed evaluation and representation of the user journey. Also in the Define

stage, a user persona was created to represent a typical user, highlighting their problems and expectations. This chapter will first present the interview and the empathy map, followed by the presentation of the user persona. Lastly, it will showcase the simulation of the current manual processes at Kontron Transportation GmbH.

4.1.1 Mapping Existing Processes - Interview on Ticket Management Process at Kontron Transportation GmbH

An interview was conducted with Mehdi Masinovic, Product Architect at Kontron Transportation, who leads the team. He and five of his team members participated in the UX testing. In the interview, the ticket management process at Kontron Transportation GmbH was discussed, with a focus on how tickets are handled, assigned, and solved. Tickets at Kontron Transportation GmbH represent any issue, feature, or incident and they are categorized into defects, enhancements, or features. These tickets can be created mostly by customers, but also internally.

Ticket Assignment and Prioritization

The assignment of the tickets at Kontron Transportation GmbH varies by team. Some teams have dedicated project managers responsible for assigning tickets, while in other teams, any team member can open and assign tickets. Tickets are allocated between different groups, depending on who encounters the issue. Ticket prioritization is based on a combination of urgency (non-gating to blocking) and impact (minor to critical), which leads to a priority ranking from P1 to P3.

Searching for Existing Solutions

When employees at Kontron Transportation GmbH are assigned a ticket, they first search the internal ticket database to find similar issues that may have been previously documented. This database, built on an SQL system, allows for keyword searches across ticket descriptions and groups. Depending on the product and its time in production, the number of relevant tickets returned can range from a few dozen to several hundred. The time taken to locate a similar ticket varies. For newer products, this process typically takes about 30 minutes, while for older products, it can take several hours due to the larger number of tickets available, some of which already contain solutions. However, not all tickets in the database provide a solution, and even those marked as "solved" do not always contain useful resolution details (M. Mehdi, personal communication, Juni 19, 2024).

Ticket Complexity and Resolution Time

The time it takes to read a ticket depends on its level of detail. More elaborate tickets can take 15 to 30 minutes to fully understand, while simpler tickets can be reviewed in under five minutes. Once a similar ticket is found, the time required to resolve the

issue depends on its complexity. Simple bugs may be resolved within days, while more complex issues can take months, or even years. On average, the department creates five new tickets per day, with about 25 employees working on them. Multiple employees may collaborate on a single ticket (M. Mehdin, personal communication, Juni 19, 2024).

Potential Benefits of the POTATO Library

The introduction of the POTATO Library is expected to improve the ticket search process by filtering tickets more effectively. Currently, employees must manually determine whether a found ticket contains a solution, which can be time-consuming. POTATO would help streamline this process by reducing the number of irrelevant tickets and focusing on those that contain applicable solutions. The system could categorize tickets based on stages such as "problem reported," "problem described," "solution suggested," or "solution provided," enhancing efficiency and allowing employees to find solutions more quickly (M. Mehdin, personal communication, Juni 19, 2024).

At this point, it is important to mention a technical aspect that will be crucial for understanding the automated simulation later. POTATO is not a search engine but a rule-based classification system for text documents, as previously mentioned during the introduction of POTATO. In the Kontron Transportation GmbH use case, the focus is indeed on search and filtering, and this is where POTATO could prove useful. Without POTATO, the process for determining resolved tickets would require considerably more time. The filtering function itself is an independent system that performs the task of filtering when categorical information about the documents is available—in this case, provided by POTATO. This enables the filtering and sorting based on such information.

Cost-Saving Estimations and Hourly Rates

Since cost savings are a key focus of the simulation, it was important to base calculations on an average hourly wage. As Kontron Transportation did not provide specific wage information, an average of €22 per hour was used, based on multiple sources (Statistics Austria, o. D., Sharma, 2024). This amount reflects a general wage in Austria and is not specific to the IT sector. The UX testing participants included IT professionals who likely earn above the average wage, but the simulation was based on the previously mentioned average hourly wage. The choice of a general wage is grounded in two main considerations. First, the tool is not limited to IT applications, so it was important to ensure that the results are as generalizable as possible. Second, the thesis assumes that completing a mechanical task, primarily requiring natural language understanding, does not need the expertise of a highly paid IT professional. The subprocess automated by POTATO replaces a manual workflow that can typically be handled by an administrator, whose salary is lower.

4.1.2 Emphaty Map

Category	Short Definition	Details
Says	Potential Quotes from users	- "It takes hours to find relevant tickets for older products." - "Not all solved tickets contain useful solutions." - "The manual filtering process is time-consuming." - "We need a system to better categorize and find solutions"
Thinks	Insights into the internal thoughts or concerns of the user that may not be directly expressed.	"How can we streamline this process to save time?" - "There must be a way to automatically find the tickets with solutions." - "This search process could be more efficient with better filtering." - "POTATO could reduce our search time, but will it work as expected?"
Does	The physical actions or behaviors that users perform during the process.	- Manually searches the internal ticket database for similar issues. - Spends time reading through multiple tickets to check for solutions. - Collaborates with other employees to solve more complex tickets. - Spends hours searching for relevant tickets for older products.
Feels	The emotional responses or attitudes	Frustrated: "It's taking too long to find solutions for older tickets." - Impatient: "We need quicker filtering for relevant tickets." - Hopeful: "POTATO could really improve this process." - Overwhelmed: "Too many irrelevant tickets slow down the work."

Table 2: Kontron Transportation GmbH Emphaty Map
(Own representation based on Yeung Ng, 2023)

An empathy map is a tool that helps capture a more complete understanding of users by highlighting their thoughts, feelings, actions, and spoken words. It enables to view the experience from the user's perspective, providing key insights into their motivations, challenges, and behaviors (Yeung Ng, 2023). Table 2 below offers a potential interpretation of how employees at Kontron Transportation GmbH experience the current ticket management process—what they say, think, feel, and do. This map is based on the use

case summary, shedding light on their day-to-day challenges and behaviors, and providing useful insights for improving the system's overall effectiveness.

4.1.3 User Persona

User personas are profiles that represent a group of users who have similar needs or behaviors. Personas help teams understand their target audience and share insights with others. They also encourage designers to think from the user's perspective, preventing them from designing only for themselves (Chang, 2021).

The user personas (see Figure 1) created in this thesis are data-driven representations based on insights gathered from the pre-questionnaire and testing data. While these personas are fictional in the sense that they do not represent the actual individuals who participated in the testing, and their images are not real, they are firmly grounded in the data collected.



Figure 1: User Personas for POTATO/Kontron Transportation GmbH created with Canva

Persona 1 (*fictional*): Emma (26-year-old Employee)

Age: 26

Job Role: Junior Technical Support - IT Support

Experience: Moderately confident with AI tools

Personality: Quick learner, thrives in dynamic environments, and loves getting tasks done efficiently and quickly.

Frustrations: She finds it frustrating when searching for previous tickets in the SQL database takes too long. The manual ticket filtering process feels inefficient,

Goals: Complete tasks with minimal hassle, always looking for faster solutions

Behaviour: Enjoys multitasking, and quickly picks up on systems like POTATO. She's open to adopting new tech.

Quote: "I just want to get things done fast without unnecessary steps."

Persona 2 (*fictional*): Mark (35-year-old Manager)

Age: 35

Job Role: Product Manager

Experience: Moderately experienced with AI

Personality: Goal-driven, balancing efficiency with accuracy, and focused on managing his team well.

Frustrations: Struggles to manage competing priorities and can get overwhelmed by too many tickets or unresolved issues.

Goals: Keep the team productive and resolve tickets while staying on track with project timelines.

Behavior: He's open to new tech but likes to integrate it thoughtfully into existing workflows.

Quote: "I need to make sure my team works efficiently while ensuring the job is done right."

Persona 3 (*fictional*): John (55-year-old Employee)

Age: 55

Job Role: Technical Support, IT Support

Experience: Limited AI experience

Personality: Detail-oriented and methodical, prefers to take his time fully understanding new software and tasks.

Frustrations: Struggles with fast-paced environments that expect him to quickly adapt to new systems. He hopes POTATO will help make finding solutions easier, but he's cautious about adopting new technology too quickly.

Goals: Deliver high-quality, precise work.

Behavior: He's slower to adopt new tech, favoring a stable, well-understood process.

Quote: "I take my time to ensure the work is done properly, even if it means taking a little longer."

4.1.4 ADONIS Simulation - POTATO - Manual Version

The following simulation (see Figure 2) shows how tickets are currently managed manually at Kontron Transportation GmbH. It outlines the key steps employees take, from assigning and prioritizing tickets to reviewing and searching for similar cases.

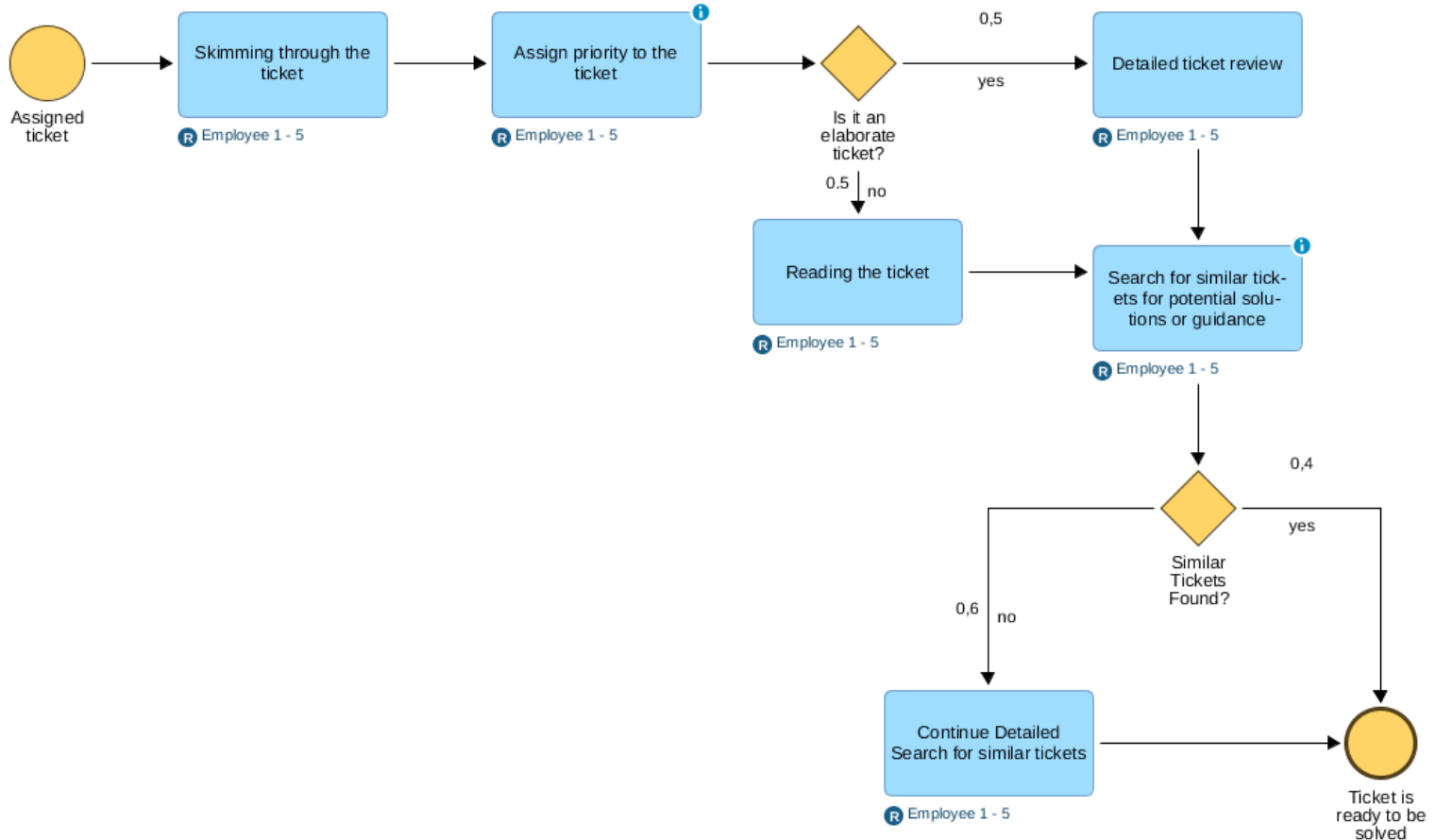


Figure 2: Process Simulation Manual for Ticket Management at Kontron Transportation GmbH(Simulation tool: Adonis)

The Process starts with a ticket being assigned to Employee 1-5, who first skims through it to get a general understanding. The term "employee 1-5" in this simulation refers to a group of five employees. Based on this, they assign priority to the ticket, as Kontron Transportation GmbH distinguishes between Priority 1, 2, and 3 when setting priority levels. At this point, there's a decision to determine whether the ticket is an elaborate one, with a 50 percent chance either way. If it is, the group moves on to perform a detailed review of the ticket. If it's not, they simply proceed to read the ticket more

carefully. After reading or reviewing the ticket, the group looks for similar tickets that might offer solutions or guidance. With an assumed probability of 0.4, they find a match, and the ticket is ready to be solved. If not, with an assumed probability of 0.6, they continue searching in more detail until a solution is found. The simulation assumes that there is always a ticket that can be found which is either similar or serves as a template, providing help or inspiration for solving the current ticket. Therefore, the simulation does not include a scenario where no similar ticket can be found.

Note: The probability used in the simulation (e.g., 50 percent, 60 percent, 40 percent) are not based on empirical data provided by the company. Instead, these values were assumed and estimated based on theoretical considerations.

Estimated times for each task can also be included. The task durations in this simulation are estimated as follows: skimming through the ticket takes about 3 minutes, with another 3 minutes allocated to prioritizing the ticket. A more in-depth reading is expected to take 5 minutes, while a detailed review requires roughly 25 minutes. Searching for similar tickets that could offer solutions is estimated at 30 minutes, and, if no match is found, a further detailed search averages around 3 hours. These times are informed by an interview with Kontron Transportation GmbH, with reasonable estimates provided for smaller tasks where exact data was not available.

To better understand the following simulation, there are a few important points to mention. It's important to note that more than one person can work on a ticket at the same time. In this scenario, five employees are assumed to collaborate on a single ticket. Based on the interview with Kontron Transportation GmbH, an average of 25 employees work in a department, managing around five tickets per day.

Note: If five tickets arrive in a day, it does not necessarily mean that five employees work together on a single ticket. It is possible for one employee to handle all five tickets, while the remaining 24 employees in the team work on their own tasks. However, for the simulation, it was assumed that each ticket is handled by a team of 5 employees. With 5 tickets per day, all 25 employees are evenly assigned across the tickets. The simulation is therefore based on a single ticket, with each task representing the actions taken by the five employees assigned to that ticket.

Additionally, it's important to note that workflows can vary by department, depending on how involved the manager is with the team's work. In this particular simulation, we assume that the manager assigns the tickets, after which the team works independently. The simulation presented above represents the individual contributions of the team.

It's also worth highlighting that this simulation is limited to the ticket preparation phase. It specifically simulates how a ticket reaches the point where it is ready to be worked on. This means that the ticket management process does not end with the preparation of a given ticket. The actual resolution of the ticket begins after the ticket preparation phase. However, in this thesis, the simulation is limited to the ticket preparation process. The reasoning behind this is that the time it takes to solve a ticket can range from a few hours to several years according to the interview conducted with Mehdi Masinovic. It's difficult to calculate an average for such a wide range, and excluding outliers wouldn't

provide statistically reliable results due to insufficient data. Furthermore, the aim of the simulation is to show how POTATO could be integrated into the current workflow. POTATO's role is not in solving the ticket itself, but in improving the efficiency of finding similar tickets in the database and classifying and filtering them. Therefore, this simulation, like the manual version and the automated version (which will be discussed in later chapters), focuses on the ticket preparation phase.

Before getting into the simulation results, it's also useful to point out that hourly wages can be assigned to employees in the simulation. This approach adds a layer of insight, allowing process efficiency and the current workflow to be evaluated through cost metrics. As mentioned in the interview presentation chapter, this Ticket Management Simulation uses an average hourly wage of €22 per employee. The €22 hourly rate represents a gross salary and does not include additional components of personnel costs.

We may now proceed to examine the results of the manual ticket management process simulation (see Figure 3).



Name ↑	Execution time	Waiting time	Resting time	Transport time	Cycle time	Non-personnel costs	Personnel costs	Total costs
Manuel Ticket Preparation - Kontron AG	02:37:56	00:00:00	00:00:00	00:00:00	02:37:56	0.00	289.56	289.56

Figure 3: Result Analysis Overview manual - Ticket Management at Kontron Transportation GmbH (Simulation tool: Adonis)

The simulation for manual ticket preparation at Kontron Transportation GmbH was conducted with 10,000 calculation cycles, 220 working days per year, and 8 working hours per day. The mean execution time for one ticket preparation is estimated to be 2 hours, 37 minutes, and 56 seconds. The personnel costs per ticket amount to €289.56.

Since, according to the Kontron Transportation GmbH process description, an average of 5 tickets are created daily for a group, or let's call it department. Based on the simulation, preparing 5 tickets daily requires $5 * 2.37$ hours, which totals 11.85 hours per day. According to the data simulated for one ticket, this results in a daily cost of $€289.56 * 5$, which means a cost of €1,447.8 per day. Since precise knowledge of annual expenses is essential for a company's cost calculation, the base simulation costs were also calculated for one year. The simulation considers 222 working days per year, so $11.85 * 222 = 2,630.7$ hours are spent annually per department on ticket preparation. This translates to a cost of $1,447.8 * 222 = €321,411.6$ annually (see Table 3).

Again, these numbers, and costs reflect the effort to prepare one ticket for further processing. The actual solving of the ticket begins only after this preparation phase.

In this simulation, the cycle time and execution time are the same because there are no additional assumed delays such as waiting time, transport time, or resting time

	Calculation	Results
1 Ticket Preparation Time		2.37 hours
1 Ticket cost		289.56 €
Daily Preparation Time	5 x 2.37 hours	11.85 hours
Daily Cost	5 x 289.56 €	1,447.8 €
Annual Preparation Time	11.85 x 222 days	2,630.7 hours/year
Annual Cost	1,447.8 x 222 days	321,411.6 €

Table 3: Ticket Preparation Time and Cost Calculations - Manual

between tasks. Cycle time represents the total time it takes to complete the entire process, including any potential delays or pauses. Since no such delays were given in this simulation, the cycle time is the sum of the execution times for each task.

per year per month **per day** per process

Simulation parameters Export results

	Name ↑	Roles	# executions	Execution time	Waiting time	Resting time	Transport time	Non-personnel costs	Personnel costs	Total costs	Capacity
Manuel Ticket Preparation - Kontron AG											
	Assign priority to the ticket	Employee 1 - 5	1.00	00:03:00	00:00:00	00:00:00	00:00:00	0.00	5.50	5.50	0.01
	Continue Detailed Search for...	Employee 1 - 5	0.59	01:47:02	00:00:00	00:00:00	00:00:00	0.00	196.22	196.22	0.22
	Detailed ticket review	Employee 1 - 5	0.50	00:12:24	00:00:00	00:00:00	00:00:00	0.00	22.72	22.72	0.03
	Reading the ticket	Employee 1 - 5	0.50	00:02:31	00:00:00	00:00:00	00:00:00	0.00	4.62	4.62	0.01
	Search for similar tickets for ...	Employee 1 - 5	1.00	00:30:00	00:00:00	00:00:00	00:00:00	0.00	55.00	55.00	0.06
	Skimming through the ticket	Employee 1 - 5	1.00	00:03:00	00:00:00	00:00:00	00:00:00	0.00	5.50	5.50	0.01
	Summary (6)		4.59	02:37:56	00:00:00	00:00:00	00:00:00	0.00	289.56	289.56	0.33

Figure 4: Result Analysis Activities manual - Ticket Management at Kontron Transportation GmbH (Simulation tool:Adonis)

Breaking down the tasks, the most time-consuming one is the detailed search for similar tickets, which takes 1 hour and 47 minutes. This accounts for the highest personnel cost of €196.22 per ticket.

4.1.5 User Problem Formulation

The table below (Table 4) summarizes the main issues in Kontron Transportation GmbH's ticket preparation process, based on the previously shown simulation. Each issue is detailed by type, root cause, and expected solution. The focus is on the most costly phase, "Continue Detailed Ticket Research," and areas where POTATO could streamline tasks and reduce manual effort.

No.	Type of Issue	Root Cause	Expectation
1	High preparation costs	Manual ticket preparation without automation	Introduce a system to streamline ticket preparation, reducing manual effort and costs: POTATO Cost of five employees
2	Excessive time spent on "Continue detailed search for similar tickets" (most costly phase)	Ineffective search engine makes it difficult to find relevant similar tickets	Automate this step with POTATO, which can classify tickets, list only relevant ones. Plus add filtering functions
3	High personnel costs	Five employees manually working on the ticket	Optimize task allocation or automate steps to reduce the number of employees needed per ticket.

Table 4: User problem formulation

The first point is a general observation of high preparation costs, driven by steps that could be automated. Currently, preparing one ticket requires on average 2 hours, 37 minutes, and 56 seconds of effort from five employees, resulting in daily personnel costs of €289.56. Introducing a system to automate parts of this process would cut down on both the manual workload and costs.

Excessive time spent on "Continue Detailed Search for similar Research," is the most costly phase, accounting for about 67.75 percent of the total preparation time (1 hour and 47 minutes) . This phase is therefore the most resource-intensive, both in terms of time and money. This issue arises from the fact, as previously mentioned, that the current database at Kontron Transportation GmbH does not function in a way that facilitates employees in processing and resolving individual tickets. Not every ticket has a solution, and even if there is description for a previous ticket, it may not contain information that is relevant or directly applicable to the current ticket. As a result, finding one or more tickets with descriptions that are actually useful can take several hours, as employees sift through numerous tickets. Here, an automated solution would be highly beneficial. POTATO could classify the tickets and, with a filtering function, allow employees to quickly identify which tickets and information are relevant, eliminating the need to manually read through countless tickets to find one that resembles their own.

Finally, high personnel costs arise because five employees work manually on each ticket. Notably, as mentioned in the interview, an average of five tickets arrive per day, with a team of 25 employees, meaning that an average of five employees work on a single ticket. Further information, interviews, and research would be required to determine whether an average of five employees per ticket is truly necessary. However, this issue is not within the scope of the current thesis. Furthermore, if the most time-consuming part of the process were automated with POTATO, personnel costs would be significantly reduced. However, the necessity of having five employees per ticket remains uncertain.

4.2 Ideate

Now we are moving into the ideate stage, where, after analyzing the issues and defining key problem areas, we focus on generating potential solutions. This phase aims to explore ideas and practical methods for addressing the challenges identified in Kontron Transportation GmbH's ticket preparation process. As mentioned earlier, a potential solution would be to automate the "detailed search for similar ticket " phase with POTATO, streamlining this most time-consuming part of the process.

This chapter will introduce the must-have, should-have, and could-have features that POTATO needs, to automate the "detailed search for similar tickets" process task. Additionally, it will cover which of these features were realized in the prototype used for the UX testing session with Kontron Transportation GmbH employees. Following this, a new ADONIS simulation will be presented, showcasing the automated version of the ticket preparation process with POTATO integrated into the workflow. The results of this simulation will be shown, followed by a comparison of values from the automated ticket management process versus the manual ticket management process.

4.2.1 MSCW Model

The MoSCoW method breaks down project requirements into four groups, making prioritization easier and more effective. As represented below in Table 5, firstly, there are the Must-haves. These are the absolute essentials—the project can't work without them. If any Must-have requirement or feature is missing, the project will fail to deliver value. Next, we have Should-haves, which are important features that make a big difference but aren't critical. The project will still function without them, but it won't be as strong or efficient.

Then come the Could-haves. These are the nice-to-have extras that add value but aren't necessary. If time or budget becomes tight, these are the first things that can be cut. Finally, there are Won't-haves. These are items left out for now, either because they are not a priority or could complicate the project too much. They may be revisited later, but for now, they are excluded to keep the project focused (Schicker, 2023).

Based on the table, the functionality and features of POTATO have been categorized into these four groups. Before diving into these details, it is important to mention that we are

M	Must have	Essential elements that the project cannot succeed without.
S	Should have	High-priority features that add significant value but are not critical for the project's core.
C	Could have	Nice-to-have features, but can be excluded if necessary.
W	Won't have	Features that won't be included now but might be considered in the future.

Table 5: MoSCoW method (Own representation based on Schicker,2023)

not refining an existing software product but developing a new, not yet released system. POTATO can be tailored to the specific use cases of any interested company. With this in mind, POTATO's features have been categorized and prioritized accordingly.

The **Must-have** features for POTATO include input data, specifically the simulated dataset that represents a ticket database.

Another essential feature is Annotation, which involves marking specific words in sentences during the training phase to indicate whether a ticket is solved, contains a solution description, etc. This annotation process helps POTATO establish its rule set and classify tickets accordingly. Another key must-have feature is Perform Inference, which follows the annotation step and generates output, such as listing useful tickets similar to the given one. Additionally, essential navigation features include buttons for modifying annotations and the filtering function. A separate system is needed within POTATO to allow ticket filtering, with the filter options: "problem described" and "solution provided."

The **Should-have** category includes features that are beneficial and add value but can be excluded if necessary. For Kontrog AG's use case, these might include expanding the filtering options, but also adding functionality to annotate multiple words simultaneously. The **Could-have** features involve enhancements like allowing annotation of the words for multiple sentences. Basic UI design for POTATO also falls into this category. The use case scenarios and UX testing of the prototype primarily focused on the training phase, testing the software's structure and functionality rather than the user interface. Thus, basic UI features are categorized as Could-haves, as they received less emphasis during this phase.

Finally, **Won't-have** features include the complete development of POTATO's UI and more refined annotation capabilities. These features would focus on fine-tuning and perfecting POTATO, which are goals for future development stages.

The goal for Kontron Transportation GmbH is to have an effective filter system that sorts out relevant tickets. However, as mentioned earlier, the filter alone cannot identify which tickets are resolved without POTATO. To achieve this, POTATO needs to be trained to classify tickets based on a set of rules, as this is a human-in-the-loop learning

system. Therefore, during UX testing, this rule based training process took priority, and a prototype was developed to test the must-have functionalities needed only for the training. The prototype will be shown later in chapter 4.3

4.2.2 Automated Simulation

We have now reached the point in the simulation where the automated version is presented. To briefly recap, the previous simulation demonstrated the manual process of Kontron Transportation GmbH's Ticket Management. Now, the same process will be simulated, but this time with the integration of POTATO.

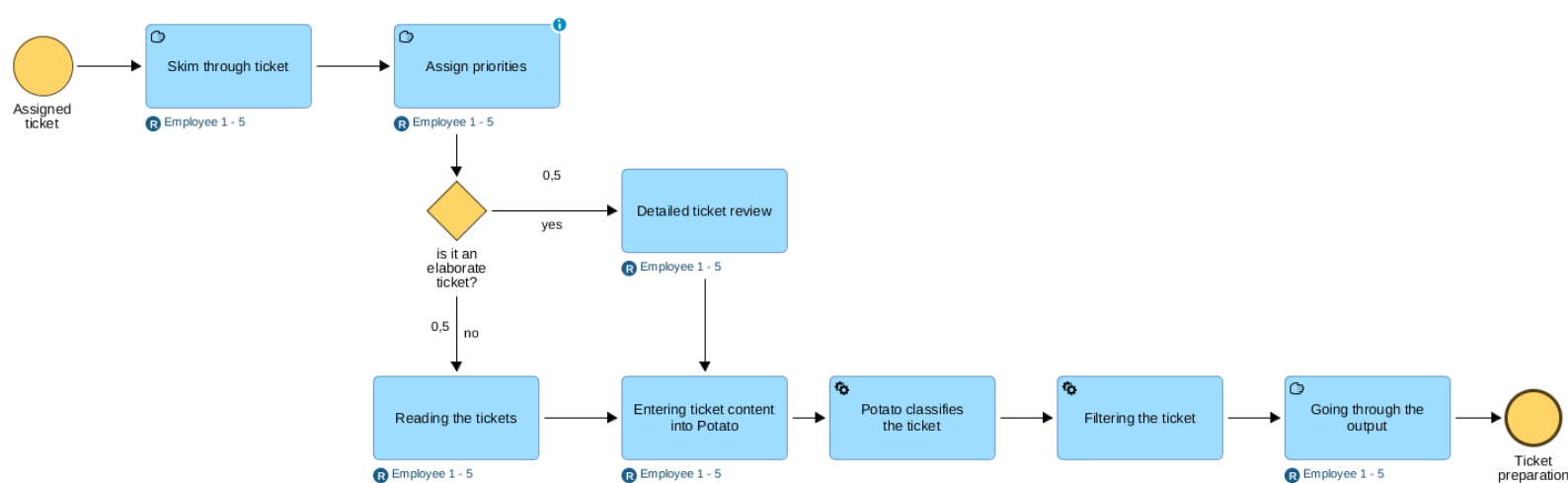


Figure 5: Process Simulation Automated for Ticket Management at Kontron Transportation GmbH (Simulation tool: Adonis)

The automated simulation process for ticket preparation at Kontron Transportation GmbH starts with one's assigned ticket entering the workflow. The first step involves quickly skimming through the ticket to gain a general understanding of its content. This step takes about 3 minutes, providing a brief overview before the ticket proceeds to the Assign Priorities stage. Here, employees allocate priority levels based on urgency and impact, which also takes around 3 minutes.

At the next decision point, the ticket is evaluated to determine if it is an elaborate ticket. There is a 50 percent chance (probability of 0.5) that the ticket is complex, in which case it moves to a Detailed Ticket Review. This review involves a more thorough analysis, and requires about 25 minutes. If the ticket is not elaborate, it moves to the Reading the Tickets step, where employees spend approximately 5 minutes understanding the details. Until this point, the process aligns with the manual version. When the ticket reaches this phase, the assigned person manually enters the ticket into POTATO and adds it

to the database. Once entered, POTATO automatically classifies and filters the ticket. This process takes around 5 seconds. The classified ticket is then filtered. The filtering naturally depends on the filters that have been set for the tickets, and the results will be provided based on these settings. A filtering task was estimated to take 30 seconds per ticket, as it includes the time needed to select the criteria for filtering the tickets. Finally, employees go through the output generated by POTATO to ensure accuracy. The Ticket Preparation is then marked as ready.

Similar to the manual version, this simulation is designed for processing one ticket, with a team of five employees working on it. The results are calculated based on the hourly wage of these five employees, set at €22 per hour per person. It is important to note again that this value was not provided by Kontron but is an average derived from online sources, as mentioned earlier.

Note: Again, this €22 hourly rate represents a gross salary and does not include additional components of personnel costs. Similarly, the probabilities used in the simulation were not derived from company-provided empirical data. Rather, they were based on assumptions and theoretical estimations.

Just like in the manual approach, the involvement of management is assumed only up to the point of ticket assignment. After that, the simulation operates under the assumption that the employees carry out their tasks independently.

Following the description of the automated ticket management process, the next section will address the simulation results. The simulation for automated ticket preparation was conducted also with 10,000 calculation cycles, 220 working days per year, and 8 working hours per day.

The simulation results in Figure 6 provide an overview of automated ticket management process at Kontron Transportation GmbH. The execution time amounts to 34 minutes and 34 seconds per ticket. There is no waiting time, resting time, or transport time, all of which are recorded as 0, and therefore, as in the manual simulation, the **cycle time** is identical to the execution time. As mentioned before, cycle time refers to the total time taken to complete one entire process, from start to finish, including any waiting or delays. *Execution time*, on the other hand, is the actual time spent performing the tasks within the process. The personnel costs for one ticket are calculated at €62.29.

Overview	Activities	Capacity	End events & paths					
per year	per month	per day	per process					
⚙️ Simulation parameters 📄 Export results ℹ️								
Name ↑	Execution time	Waiting time	Resting time	Transport time	Cycle time	Non-personnel costs	Personnel costs	Total costs
Process Simulation Automated for Tick...	00:34:34	00:00:00	00:00:00	00:00:00	00:34:34	0.00	62.29	62.29

Figure 6: Result Analysis Overview automated - Ticket Management at Kontron Transportation GmbH (Simulation tool: Adonis)

Table 6 explains the time and cost calculations for the automated ticket preparation process. Preparing one ticket takes 34 minutes and 34 seconds, with a personnel cost of €62.29 per ticket. Since 5 tickets are processed each day, the total preparation time comes to 2 hours and 53 minutes, costing €311.45 daily. Over the course of a year, with 222 working days, the total preparation time adds up to 639 hours and 29 minutes annually, with a total yearly cost of €69,141.9. The daily values in the simulation are useful because they provide a tangible view of the time and energy investment on a daily basis. On the other hand, the yearly simulation data offer a better perspective from a budgeting standpoint, showing the total cost associated with a specific task or process over the year.

	Calculation	Results
1 Ticket Preparation Time	34 minutes and 34 seconds	0.58 hours
1 Ticket cost		62.29 €
Daily Preparation Time	5 x 0.58 hours	2.88 hours
Daily Cost	5 x 62.29 €	311.45 €
Annual Preparation Time	2.88 x 222 days	639.48 hours
Annual Cost	311.45 x 222 days	69,141.9 €

Table 6: Automated Ticket Preparation Time and Cost Calculations

Overview		Activities	Capacity	End events & paths									
per year		per month	per day	per process									
⚙️ Simulation parameters												📄 Export results ⓘ	
# ↑	Name ↑		Roles	# executions	Execution time	Waiting time	Resting time	Transport time	Non-personnel costs	Personnel costs	Total costs	Capacity	
📅 🍌 Process Simulation Automated for Ticket Management at Kontron AG													
	🍌 Assign priorities	👤 Employee 1 - 5	1.00	00:03:00	00:00:00	00:00:00	00:00:00	00:00:00	0.00	5.50	5.50	0.01	
	🍌 Detailed ticket review	👤 Employee 1 - 5	0.50	00:03:00	00:12:28	00:00:00	00:00:00	00:00:00	0.00	22.87	22.87	0.03	
	🍌 Entering ticket content into P...	👤 Employee 1 - 5	1.00	00:03:00	00:00:00	00:00:00	00:00:00	00:00:00	0.00	5.50	5.50	0.01	
	🍌 Filtering the ticket	No role assigned	1.00	00:00:30	00:00:00	00:00:00	00:00:00	00:00:00	0.00	0.00	0.00	0.00	
	🍌 Going through the output	👤 Employee 1 - 5	1.00	00:10:00	00:00:00	00:00:00	00:00:00	00:00:00	0.00	18.33	18.33	0.02	
	🍌 Potato classifies the ticket	No role assigned	1.00	00:00:05	00:00:00	00:00:00	00:00:00	00:00:00	0.00	0.00	0.00	0.00	
	🍌 Reading the tickets	👤 Employee 1 - 5	0.50	00:02:30	00:00:00	00:00:00	00:00:00	00:00:00	0.00	4.59	4.59	0.01	
	🍌 Skim through ticket	👤 Employee 1 - 5	1.00	00:03:00	00:00:00	00:00:00	00:00:00	00:00:00	0.00	5.50	5.50	0.01	
	Summary (8)		7.00	00:34:34	00:00:00	00:00:00	00:00:00	00:00:00	0.00	62.29	62.29	0.07	

Figure 7: Result Analysis Activities automated - Ticket Management at Kontron Transportation GmbH (Simulation tool: Adonis)

In the following figure (see Figure 7), the table outlines the key activities of the automated ticket management process at Kontron Transportation GmbH. The Detailed Ticket Review emerges as the most time-consuming and expensive task, taking 12 minutes and

28 seconds per ticket and causes a personnel cost of €22.87. In contrast, the POTATO Classifies the Ticket step is the fastest and cheapest, lasting only 5 seconds with no associated personnel cost.

Before diving into the comparison and evaluation of the two simulations, there is another essential part of the comparison to address: the training phase. The training phase for an AI software is the step where the system learns to perform its tasks accurately. During this phase, the AI model is provided with large amounts of data and taught how to analyze and make decisions based on that data. This process involves feeding the AI labeled examples, allowing the model to understand patterns, rules and relationships (Géron, 2019).

The previous simulation represented the process where POTATO is applied after the training phase. However, to properly answer Research Question 1, which focuses on the cost and time savings from automation, it's also important to include the time and costs from the training phase in the overall comparison.

The following section will therefore present the training process of POTATO, similar to the structure of the previous two (see Figure 8).

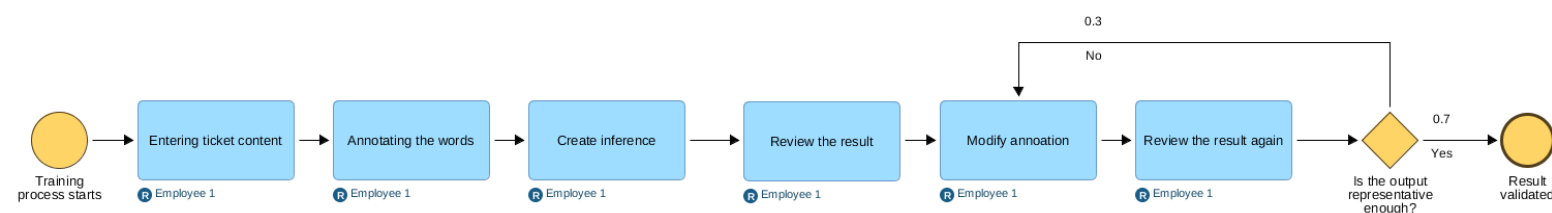


Figure 8: POTATO training phase simulation

The training phase starts with Entering Ticket Content, where the ticket is chosen and added to the system. Next, there's Annotating the Words, where key words from the sentences in that ticket are tagged and marked. These chosen words always represent certain characteristics of the ticket, such as whether it is solved or not. The marked words imply and help POTATO understand how the employee knows if the ticket is solved.

Then comes Create Inference, where POTATO automatically generates inferences from the annotations. This step is automated and doesn't require any manual interpretation. Once the inference is generated, Review the Result comes into play, where the output is checked to see if it meets the expected outcome. If it needs improvement, the process moves to Modify Annotation, where the annotations are refined to "teach" POTATO how to provide more accurate responses. After this, another review takes place. A decision is then made: Is the output representative enough? If yes, with a 70 percent chance, the process concludes with Result Validated. If not, there's a 30 percent chance it loops back to Modify Annotation, and the refining process continues until POTATO delivers a

proper output. The probabilities used in the simulation were not provided as empirical data by the company but were instead estimated based on theoretical reasoning and assumptions.

The above process presents the journey of a single ticket, handled by one person. However, training POTATO cannot be accomplished with just one ticket. The number of tickets required to annotate words and effectively train POTATO depends on various factors. These factors include the complexity of the tickets but also the precision and routine of the person doing the annotation. Therefore, it is not possible to specify an exact number, but a range can be given. The cost and time of the training phase will be provided as an interval. The first simulation represents the lower, optimistic estimate, while the second shows the pessimistic version (see Figure 9 and 10). The optimistic simulation assumes that 50 tickets are sufficient for POTATO, the pessimistic version, on the other hand, assumes the processing of 300 tickets.

per year per month **per day** per process

Simulation parameters Export results

Name ↑	Execution time	Waiting time	Resting time	Transport time	Cycle time	Non-personnel costs	Personnel costs	Total costs
POTATO _ training phase simulation	1d 03:07:24	00:00:00	00:00:00	00:00:00	1d 03:07:24	0.00	244.72	244.72

Figure 9: Optimistic simulation of the POTATO training phase

The result for the optimistic simulation of the POTATO training phase shows an execution time of 1 day, 3 hours, and 7 minutes. Since we work with standard 8-hour workdays instead of full 24-hour days, this duration translates to over three workdays. Specifically, it spans 27 hours in total, which is equivalent to three complete workdays plus an additional three hours on the fourth day.

The total cost for this execution amounts to €244.72, covering the personnel expenses involved in the training phase.

per year per month **per day** per process

Simulation parameters Export results

Name ↑	Execution time	Waiting time	Resting time	Transport time	Cycle time	Non-personnel costs	Personnel costs	Total costs
POTATO _ training phase simulation	8d 02:39:21	00:00:00	00:00:00	00:00:00	8d 02:39:21	0.00	1,466.43	1,466.43

Figure 10: Pessimistic simulation of the POTATO training phase

The result for the pessimistic simulation of the POTATO training phase indicates an execution time of 8 days, 2 hours, and 39 minutes. Considering that a standard workday is 8 hours, this translates to 66 hours and 39 minutes in total. Spread across a work schedule, this execution time would extend over eight full workdays and almost three more hours into a ninth day.

The total cost associated with this execution amounts to €1,466.43. Both simulations are based on an hourly wage of €22 per hour, which has been used to calculate the total personnel costs.

Overall, the range for this training phase spans from €244.72 to €1,466.43, with the time varying between 27 hours and 66 hours, 39 minutes (see Table 7).

Scenario	Execution Time	Equivalent Workdays	Hourly Rate	Total Cost
Optimistic	27 hours	3 full days + 3 hours	€22 /hour	€244.72
Pessimistic	66 hours and 39 minutes	8 full days + 2 hours 39 minutes	€22 /hour	€1,466.43

Table 7: POTATO Training Phase Costs and Time Estimates

So again the optimistic and pessimistic estimates are important because there are two factors that cannot be precisely predicted. The first is how much training POTATO will require, as this depends on the complexity of the tasks. The second is how long it will take for users to become fluent in using the tool. Additionally, the software may need to be retrained from time to time.

4.2.3 Comparison: manual vs automated simulation results

This chapter presents a comparison between manual and automated text processing, utilizing POTATO as the automated solution. This analysis is essential for addressing the research question:

R1: How does the utilization of POTATO for text processing tasks enhance work efficiency compared to manual text processing methods?

The data is broken down into time and cost savings to illustrate the impact of automation. The results show a substantial increase in efficiency, both in terms of time and cost (see Table 8)

Time Efficiency

For 1 Ticket, manual processing takes 2 hours and 22 minutes, while POTATO completes the task in just 34 minutes. This results in a time savings of 1 hour and 48 minutes per ticket, representing a 75.5 percent reduction. On a daily basis, handling 5 ticket preparation manually requires for the employees (on average 25 employees working in a department) 11 hours and 51 minutes, compared to just 2 hours and 53 minutes with POTATO. This means 8 hours and 58 minutes are saved each day, translating to a 75.7

Metric	Manual Processing	Automated with POTATO	Reduction
1 Ticket Preparation Time	2.37 hours (2 hours, 22 minutes)	0.58 hours (34 minutes)	1.79 hours (1 hour, 48 minutes, 75.5 percent)
1 Ticket cost	€289.56	€62.29	€227.27 (78.5 percent)
Daily Preparation Time	11.85 hours (11 hours, 51 minutes)	2.88 hours (2 hours, 53 minutes)	8.97 hours (8 hours, 58 minutes, 75.7 percent)
Daily Cost	€1,447.8	€311.45	€1,136.35 (78.5 percent)
Annual Preparation Time	2,630.7 hours (2,630 hours, 42 minutes)	639.48 hours (639 hours, 29 minutes)	1,991.22 hours (1,991 hours, 13 minutes, 75.7 percent)
Annual Cost	€321,411.6 /year	€69,141.9 /year	€252,269.7 /year (78.5 percent)

Table 8: Comparison of Manual vs. Automated Processing with POTATO

percent reduction in daily preparation time. Over the course of a year (222 workdays), manual processing totals 2,630 hours and 42 minutes, whereas POTATO only uses 639 hours and 29 minutes. The annual time savings amount to 1,991 hours and 13 minutes, maintaining a 75.7 percent reduction.

Cost Efficiency

The cost of processing 1 Ticket manually is €289.56, while POTATO reduces this to €62.29, saving €227.27 per ticket, or 78.5 percent. On a daily basis, the manual cost is €1,447.8, but with POTATO, it drops to €311.45, yielding daily savings of €1,136.35, also a 78.5 percent reduction. Over a year, the cost of manual processing is €321,411.6, whereas automation reduces this to €69,141.9, resulting in annual savings of €252,269.7, which is a 78.5 percent reduction.

In summary, using POTATO in Kontron Transportation GmbH's ticket management process significantly improves productivity. The automation reduces processing time by more than 75 percent and lowers costs by nearly 79 percent. This shows that POTATO not only speeds up ticket handling but also leads to substantial cost savings.

While the automated approach provides significant time and cost savings, it comes with an initial training cost as mentioned earlier. Based on the earlier simulation results, the training cost can range from €244.72 (optimistic scenario) to €1,466.43 (pessimistic scenario).

With this calculation, Research Question 1 has been answered successfully, and the results align with the expected outcomes.

Beyond the training costs, if a company intends to use POTATO, it must also account for both fixed and ongoing costs. The fixed cost covers the setup and customization of the system for the specific tasks of the company, while the ongoing cost involves a lower monthly support fee. These costs are highly dependent on the company and require individual calculations. They vary based on the company's tasks, their scale, and complexity, as well as the associated monthly support requirements. Since POTATO is still under development and does not yet have a standardized pricing model, it is currently being deployed in individually priced consultancy projects.

As the research questions of this thesis do not cover these aspects and primarily focus on improving efficiency in terms of time and cost, the pricing of POTATO in the market could serve as an additional valuable topic for future research. This would complement and extend the scope of this thesis. Another further potential research area, or question remains: how do these results translate into the real-world business scenario for Kontron Transportation GmbH, and how can they be practically applied?

The calculation shows that using POTATO frees up 8.97 hours, which is almost equivalent to a full workday for one employee. If we spread this time evenly across 25 employees—which admittedly raises questions about how representative that is—it averages out to 0.36 hours (about 21.5 minutes) of freed-up time per person. The real question becomes: how should this extra time be used? Even though work hours are saved, the company still has to pay employees for the standard 8-hour workday. One approach could be increasing ticket processing capacity or optimizing costs, which may free up workforce for other tasks. More detailed information on further research ideas can be found in later chapters.

4.3 Prototype

The next chapter will focus on presenting the prototype. The prototype to be presented is an actual real version that was used during UX testing at Kontron Transportation GmbH. Since it was not feasible to test the entire ticket management process, the testing session concentrated on the training process of POTATO. As a result, the prototype includes the functionalities necessary for this purpose. It is also important to mention that, as this is not a finished product and the priority during prototyping was on UX—focusing on the structure and functionality of the software rather than UI design elements—the prototype should be reviewed with this context in mind.

Figure 11 displays the "NewPotato Demo" interface, specifically the Annotate tab, where a sentence is being manually annotated. The first tab, "Add Sentence," allows sentences

to be added to the system for further processing and annotation. Annotation is the process of labeling specific parts of a sentence to provide structured data for AI training. In this example, the sentence "The alarm was cleared after 3 min but reoccurred again at 8h51" is being annotated. The sentence is divided into individual tokens, with each word or punctuation separated and sequentially numbered. The token "alarm" has been highlighted and marked as PRED (predicate), indicating its importance as a key term in the sentence. Additionally, labels such as ARG1 and ARG2 are available to define roles or relationships between tokens. This page includes Submit and Reset buttons, which allow annotations to be finalized or revised. The Submit Annotation button completes the process.

NewPotato Demo

Add Sentence **Annotate** Inference Load

Select a sentence to annotate:

The alarm was cleared after 3 min but reoccurred again at 8h51.

0_The 1_alarm 2_was 3_cleared 4_after 5_3 6_min 7_but 8_reoccurred 9_again 10_at 11_8h51 12_.

Submit Annotation

Annotation added successfully.

Current Annotations:

≡ pred	≡ args	≡ triplet	≡ delete
1	,	alarm(,)	☐

Figure 11: Prototype 1 - New POTATO Demo

The next page (see Figure 12) shows the continuation of the annotation process from the previous step. The token "alarm" has been marked as PRED (predicate), and the annotation has been successfully submitted, as indicated by the green confirmation message. Below, the Current Annotations section lists the completed annotation.

The third image (see Figure 13) shows the Inference step, which is designed to classify sentences based on previously annotated data. At the top, a list of the data base, all

the given sentences, are shown, with checkboxes allowing the selection of individual sentences or all sentences for classification. At the bottom, the Knowledge Graph section shows the inferred results, summarizing the system's output. This functionality works by identifying sentences similar to the ones that have been annotated previously, based on the annotated words or patterns. For instance, if the goal is to find tickets marked as solved, the system will use the annotated sentences to infer and list all similar sentences that indicate solved tickets.

NewPotato Demo

Add Sentence **Annotate** Inference Load

Select a sentence to annotate:

The alarm was cleared after 3 min but reoccurred again at 8h51.

PRED x

0_The 1_alarm 2_was 3_cleared 4_after 5_3 6_min 7_but 8_reoccurred 9_again 10_at 11_8h51 12_.

Submit Annotation

Annotation added successfully.

Current Annotations:

pred	args	triplet	delete
1	,	alarm(,)	☐

Figure 12: Prototype 2 - New POTATO Demo

It is possible to switch back and forth between the windows. If the output is not as expected or does not provide the precise solution one anticipated, it is possible to go back and correct or refine the annotation to help POTATO establish a more accurate rule set. If the annotation is deemed correct, the Perform Inference function can be used, and POTATO will store the results. As previously mentioned, annotating 50 to even 300 tickets may be necessary for POTATO to work with a reliable rule set and classify sentences accurately and dependably.

After presenting the prototype, the focus shifts to the testing process and its evaluation. This evaluation will explore how well the prototype aligned with Kontron Transportation GmbH's processes, the extent to which POTATO could be customized to meet the

company's specific use case during the training phase, and how users rated the software based on pre specified criteria and their experiences.

Add Sentence Annotate **Inference** Load

Sentences to Classify

☐ Show Rules

select	sentence	annotations
☐	At ZKVG level, Oxetofzj live, Fault AMJ408 seen on XVJL for the 2 QOQ => appeared ea	
☐	Here the corresponding fault on QOQ Amstetten - same exist on Linz ZKYTTZ2FP	
☐	Fault seen 15/9 on KUM Linz : HFPKMTEEXNC_EAI.	
☐	This kind of fault had mnfdiq been seen during migration of other network (ex Ifkopn	
☐	: alarm HUJX TUYV490 QWU BILLING FILE TRANSFER (GS-NVSYJX4TM/CPZ-CGBT/STRE	
☐	The alarm was cleared after 3 min but reoccurred again at 8h51.	
☐	Our reference: Notification SPRINT 50626146 21.09.2020 07:57:02 & 21.09.2020 08:51:	
☐	Alarm WX: 205841 & 20586	
☐	Could you please investigate?	
☐	Brs Hi all, the alarm ZCO PmNoData	

Classify all sentences or selected sentences?

☒ All
☐ Selected

Classify

Knowledge Graph

Inferred:

text	pred	args	annotated
jzmsc un ticket de test Hello, CustomerA tried to eliminate the alarm (External QYQX alarm GMUPLX	None	None, None	
What is the cause of the alarm and how to eliminate it when the release command does not eliminat	None	None, None	
Data in the error report : 10.128.3.129 >kgmsgljq jql login XSZ_NJ4 >mpsol >17 (Bqi) >10 (T	None	None, None	
The FNG is releasing the call.	None	None, None	releasing(,)

Figure 13: Prototype 3 - New POTATO Demo *Note: The artificial data shown was manually constructed by Kontron Transportation to reflect the nature of tickets.*

4.4 Evaluation

This chapter presents the evaluation approach and the results of the user testing conducted with the implemented prototype. The user test involved five employees from Kontron, including one manager and four employees responsible for ticket management and resolution within their department. The participants had varying levels of familiarity with AI: some had no prior knowledge, while others had minimal to moderate experience using AI tools. All participants interacted with the new POTATO prototype, which was introduced in the previous chapter.

The primary objective of the evaluation was to address the following two remaining research questions:

R2: Within the current UX/UI design of POTATO, which features could be improved or adapted to enhance user-friendliness for individuals interacting with the platform?

R2.1: To what extent can POTATO be customized to support the specific business tasks of the collaborating company?

4.4.1 Evaluation Approach

The test design was based on the guidelines from the Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests (Rubin Chisnell, 2008):

Orientation script:

The orientation script introduced participants to the testing process and provided an overview of what to expect during the session. To ensure consistency, it was presented to the group beforehand using a PowerPoint presentation. This gave participants an opportunity to ask questions if anything was unclear. The actual testing, however, was conducted individually and anonymously. Participants drew numbers to identify them, and each person completed the test one at a time based on their assigned sequence.

Background questionnaire :

The background questionnaire was designed to collect testing relevant personal information about the participants. The detailed results will be discussed in the Results section, but the questions focused on topics such as prior experience with AI tools, familiarity with POTATO, and whether the task became clear after reviewing the orientation script. These areas were crucial for understanding participants' behavior and performance during the testing process.

Data collection instruments:

Both performance and preference data were collected. For performance data, the following aspects were monitored for each task:

- Whether the task was completed successfully.
- Whether participants needed assistance.
- How long it took to complete the task.
- Major challenges or obstacles encountered.
- Notable actions or behaviors during the task.

In addition to performance data, preference data was collected after participants completed their tasks. This allowed them to evaluate and rate the software in specific categories. The preference questionnaire focused on four key areas:

- Clarity and Understandability
- Usability and Navigation
- Customizability
- Autonomy

Preparatory training materials:

These materials provided users with the necessary background information to complete the tasks during the user test. The software was demonstrated live using an example task to familiarize participants with its functionality. This was accompanied by the PowerPoint presentation to clarify the research goals and motivations behind the testing.

Questionnaire:

Participants were presented with 17 statements distributed across four categories: Understandability, Usability and Navigation, Customizability, and Autonomy. They rated their agreement with each statement on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Following the recommendation of Rubin and Chisnell (2008), these statements were repeated after each task to track changes in ratings over time as participants gained experience. The set of 17 statements was asked three times, once after completing each of the three tasks. After the third task, the final questionnaire included three open-ended questions. These allowed participants to provide their feedback and share their opinions. The session for each participant concluded with answering these open-ended questions.

Task Scenarios:

As mentioned earlier, participants were required to complete three predefined tasks. The tasks were presented in text form, and the instructions for each subsequent task were revealed only after the successful completion of the previous one. The tasks were completed onsite at the Kontron Transportation office.

Test Protocol:

The test protocol defined the structure and agenda for the user test. It ensured the testing process was reproducible and the results were comparable across participants. Of course, before finalizing the agenda, to ensure everything works as expected, we conduct a pre-trial where we test the planned tasks, the questionnaires, and the software to confirm they function as designed and as desired.

4.4.2 Background Questionnaire Results

The Background Questionnaire (see Table 9) provides insights into participants' prior experience with AI tools, familiarity with the POTATO software, their understanding of the software's purpose after seeing a demo, their confidence in completing the upcoming tasks, and their clarity on how to rate statements using a Likert scale. The questionnaire used a likert scale from 1 to 5, where 1 indicates the lowest level of familiarity or agreement, and 5 represents the highest.

The results reveal a wide range of prior experience with AI tools among participants. While one user (User 1) reported extensive experience, rating it as 5, another (User 5) indicated no prior experience, with a score of 1. The remaining participants rated themselves between 2 and 3, suggesting a mix of moderate and minimal experience with AI.

Familiarity with the POTATO software was consistently low, with most participants rating their familiarity as either 1 or 2, and only one user (User 2) gave a slightly higher score of 3. These results match the study's expectations, as the participants had no prior experience with POTATO. Their feedback is beneficial for understanding how well the system supports first-time users.

After the demo, most participants rated their understanding of the software's purpose between 4 and 5. This shows that the demo effectively closed the knowledge gap, even for those with little experience.

Confidence in completing tasks with POTATO was generally high, it was between 4 and 5. However, one participant (User 4) rated their confidence lower, with a score of 3. Finally, all participants clearly understood how to rate statements on a Likert scale, with ratings of 4 or 5.

This was a valuable takeaway from the background questionnaire, as it ensured that the usability testing results would not be influenced or skewed by factors such as participants misunderstanding how to answer or rate the questions, or being unclear about the task descriptions or the basic introduction to the software. The background questionnaire

	User 1	User 2	User 3	User 4	User 5
How would you rate your prior experience with AI software in general? <i>1 - No experience .. 5 - Extensive experience,</i>	5	2	3	3	1
How familiar are you with our software, POTATO? <i>1 - Not familiar at all .. 5 - Extremely familiar,</i>	2	3	2	1	1
Now that I've seen the demo, I understand what this software is used for.*	4	5	5	4	4
I feel confident about completing the tasks I will receive.*	5	4	4	3	4
I am clear about how to rate the provided statements using the likert scale.*	5	5	5	4	5
<i>* 1 - Strongly Disagree .. 5 - Strongly Agree</i>					

Table 9: POTATO Background questionnaire results

proved to be effective, allowing us to proceed with the testing under the assumption that it would focus on the intended and relevant factors.

4.4.3 Tasks Results

Participants were given three tasks to complete during the experiment. The first section outlines these tasks, including how much time participants needed to complete them, whether they were successful, and if they required assistance. This is followed by an analysis of the user-generated preference data, based on the questionnaires participants filled out after each task and that we gathered continuously throughout the experiment.

Task 1

This task was assigned to five non-experts, all employees working in IT, with varying levels of experience with AI.

Task definition

The first task consisted of several subtasks, including opening the input data, identifying a ticket marked as solved, annotating the relevant words or phrases, and performing an inference. The main objective was to determine how well users could navigate through the provided input data, select a ticket marked as solved, and identify the specific words or phrases that led to this conclusion. Additionally, the task sought to test their ability to perform an inference based on the annotated information, ensuring that they understood the process and could follow the instructions effectively. Participants received the following instructions:

“ANNOTATION WITH ONE TICKET | INFERENCE

- 1:Open the input data.
- 2:Select any ticket where it indicates that the ticket is considered solved.
- 3:Annotate a word (or words) in the sentence that helped you decide that the ticket is solved.
- 4:After annotation, perform inference”

To successfully complete the task, participants had to finish all the steps outlined in the instructions within a time limit of 20 minutes. This time limit was established based on insights gained during the pilot test. The results in Table 10 show that all five participants successfully completed the task within the given time.

	User 1	User 2	User 3	User 4	User 5
Completed successfully	Yes	Yes	Yes	Yes	Yes
Assistance needed	Yes	Yes	Yes	Yes	Yes
Time needed	7m 15s	8m 9s	5m 10s	12m 15s	17m 55s

Table 10: POTATO User test task 1 performance data

However, all users required assistance at some point during the process. The completion times varied significantly, with User 3 completing the task the fastest in 5 minutes and 10 seconds, while User 5 took the longest at 17 minutes and 55 seconds. These differences in completion time could indicate varying levels of familiarity with the software or differences in understanding the task requirements. However, it should be noted that a quick completion time does not necessarily indicate a well-executed training process, as much depends on the annotated words and how effectively POTATO can be trained using them. Speed, therefore, does not automatically equate to efficiency.

What did users need assistance with?

User 1 quickly figured out how to annotate and did so extensively, but they initially struggled to find a solved ticket, spending about three minutes on this step. User 1 also needed clarification on whether multiple words could be annotated. Despite these early challenges, the annotation process went smoothly once they got started, highlighting that the tool becomes easy to use with some initial guidance.

User 2 had some trouble understanding how to open a ticket at first. After couple of minutes user 2 used the software confidently. However, the user were confused by the three annotation colors and initially misunderstood their purpose, showing that some features were not entirely intuitive.

User 3 struggled to understand the task and how to view the full ticket. While user 3 asked many questions, these were logical and showed an effort to understand the software better. It can be said the interface is functional but lacks the intuitive design needed for new users to navigate it independently. However, as mentioned earlier, this testing

primarily focuses on UX rather than UI. Of course, it is difficult to completely separate and evaluate these two aspects independently, but the thesis aims to shed light specifically on UX elements.

User 4 experienced frustration and needed help finding a solved ticket. While user 4 annotated multiple tickets simultaneously, only one word in the first ticket was annotated, suggesting limited engagement with the task. User 5 hesitated to interact with the software, and needed guidance. User 5 frequently asked for step-by-step instructions.

Task summary

There were no issues with performing the inference, and participants were able to complete it intuitively after the annotation step. Overall, based on the first round or the first task, it can be said that the core functionality of the software worked well. Analyzing the artificial dataset within the POTATO environment proved to be very slow. A significant portion of the time spent completing the first task was used to find a ticket that could be classified as solved. At this point, we were curious to see how the participants' familiarity and confidence would evolve in the next phases of the testing. Sometimes, participants inquired about information that was already provided in the description.

Task 2

This task was also assigned to five non-experts, all employees working in IT, with varying levels of experience with AI.

Task definition

The second task, "Correcting Annotation and Inference," was designed to test how well participants could review the output of their previous work, identify errors, make necessary corrections to their annotations, and then perform an inference based on the updated results. This task focused on participants' ability to reflect on their work and adjust their inputs to improve accuracy.

Participants received the following instructions:

“CORRECTING ANNOTATION | INFERENCE

- 1:Review the output.
- 2:Based on your output, modify / correct your annotation
- 3:Perform inference”

Participants were asked to review the output generated from their initial annotations, correct or modify the annotated words, and perform an inference using the revised data. Compared to the first task, this required more critical thinking, as participants had to rethink their work and figure out how to improve it.

The annotated words play a crucial role in helping POTATO interpret and categorize textual content. The more meaningful, "better" the words annotated by the users, the easier it is for POTATO to process the data, leading to a more effective training process. For this reason, we asked users to review the output generated by POTATO based on their annotations to help them understand which words are worth annotating and identify the key terms that contribute to the success of the training. This task was designed to help users internalize this concept and assess how confidently they could correct their annotations while actively using the software.

The results show (see Table 11) that all five users successfully completed the task, though every participant needed some assistance. Completion times varied widely. User 1 completed the task the fastest, in just 2 minutes and 29 seconds, while User 5 took significantly longer, at 10 minutes and 49 seconds.

	User 1	User 2	User 3	User 4	User 5
Completed successfully	Yes	Yes	Yes	Yes	Yes
Assistance needed	Yes	Yes	Yes	Yes	Yes
Time needed	2m 29s	4m 59s	3m 36s	6m 32s	10m 49s

Table 11: POTATO User test task 2 performance data

What did users need assistance with?

By the time users reached the second task, there was a visible improvement in how they interacted with the software. Some participants used it confidently, while others were still hesitant, but overall, their understanding and familiarity were clearly improving. User 1 approached the task with confidence, just a little assistance was needed. User 1 seemed focused on identifying patterns and annotated many words, but their selections might not generalize well for broader use cases.

User 2 understood where to find the output but misunderstood the goal of the task, thinking they needed to add new annotations rather than correct and expand the existing ones. User 3 struggled to find the output at first, mistaking the dataset for the actual results. They asked questions like, "I annotated word X, how can I see tickets containing X?" which shows they were engaged but unsure of how the system worked. User 4 initially thought the dataset was the output as well but eventually figured it out with a little assistance and completed the task with growing confidence. While user 4 annotated fewer words than other users, they demonstrated a better understanding of the workflow compared to Task 1.

User 5 continued to face challenges, forgetting to click the submit button and struggling to complete the task independently. User 5 seemed to be hesitant to interact with the software, possibly (also) due to language comprehension issues. User 5 marked multiple independent words in a sentence without realizing that they would be treated as a single pattern.

Task summary

Many users felt more confident navigating the software compared to Task 1, and the second task showed clear progress in how users interacted with POTATO. Participants like User 4 and User 2 grew more confident and demonstrated a better understanding. However, there were still challenges. Many users were confused about the difference between the dataset and the output, which slowed them down. Modifying existing annotations also proved unintuitive for some, particularly User 2 and User 5.

Task 3

As before, task 3 was assigned to five non-experts, all employees working in IT, with varying levels of experience with AI.

Task definition

In Task 3, participants were asked to annotate data from two tickets and perform an inference based on their annotations. While similar to Task 1, this task added complexity by requiring users to work with two tickets instead of just one. The aim was to test their ability to handle multiple data inputs.

Participants received the following instructions:

“ANNOTATION WITH 2 TICKETS | INFERENCE

- 1:Go back to the input data.
- 2:Annotate using 2 tickets.
- 3:Perform inference."

The results show (see Table 12) that all participants successfully completed the task, but all of them still needed some assistance. Completion times varied widely, with User 2 finishing the fastest at 3 minutes and 11 seconds and User 5 taking the longest at 8 minutes and 47 seconds. Despite the need for help, the overall completion times were faster compared to Task 1, indicating that participants were becoming more familiar with the software and the workflow. The main difference from Task 1 was the increased complexity of handling two tickets. Users generally adapted well to the added challenge.

	User 1	User 2	User 3	User 4	User 5
Completed successfully	Yes	Yes	Yes	Yes	Yes
Assistance needed	Yes	Yes	Yes	Yes	Yes
Time needed	4m 52s	3m 11s	4m 6s	5m 8s	8m 47s

Table 12: POTATO User test task 3 performance data

What did users need assistance with?

User 1 handled annotations confidently, focusing more on annotating word combinations rather than individual words, and carefully reviewed which ticket to choose, demonstrating a thoughtful approach. However, user 1 were slightly confused by the submission button, which required clicking twice, and asked if annotations were preserved when switching between documents. While the overall performance showed competence and minor uncertainties.

User 3 was much more confident in selecting tickets compared to earlier tasks. Performing inference came easily, and user 3 didn't require assistance until the very end when moving to the questionnaire. However User 3 consistently annotated only a few words. For User 5, inference was no longer a challenge. However, still needed assistance finding a solved ticket. Despite this, the annotation skills improved compared to previous tasks.

User 4 confidently searched within tickets and performed inference well. However, user 4 only annotated one word at a time. While becoming more confident by the end of the task, the user expressed frustration throughout the process, which may have affected their overall experience. User 2 demonstrated the strongest confidence and intuition while using the software. Although user 2 had some trouble understanding the task itself, the user used POTATO effectively and made corrections to their annotations multiple times. This user was the only participant to perform several inferences during the task and check the outputs afterward. This intuitive use of the tool stood out.

Task summary

Participants showed notable improvements in confidence. Particularly with annotations and inference. However, challenges remained for some, especially with understanding task instructions and navigating the interface. User 2 stood out as the most intuitive user, actively engaging with the software and exploring its capabilities. This is an interesting observation because User 2 rated their prior experience with AI tools as a 2 on a 1-5 scale. There were other users, including User 4, who rated their experience higher. But on the other hand, after the demo, User 2 also provided high ratings on the pre-questionnaire, particularly in areas assessing their confidence and understanding of the tasks. In contrast to that, Users 5 and 4 still required assistance with certain steps, suggesting areas where the interface could be further refined.

Performance data summary

The testing sessions highlighted both strengths and areas where the software could improve. Participants became noticeably more confident as they progressed through the tasks, showing that the tool becomes easier to use over time. However, a few recurring issues stood out that could be improved to make the software more user-friendly.

One of the challenges was the visibility of tabs. Several users didn't even notice they were there, or were not clearly "communicated" which caused confusion. This shows the interface could benefit from clearer visual cues. Another issue was that many participants created very specific annotations, which may not generalize well. Adding a feature to

suggest broader, more general patterns could help steer users in the right direction and make the annotation process more efficient. Furthermore, it wasn't immediately clear to all users after creating an inference, which table represented the output and which was the dataset. In addition, issues like users annotating two tickets in the first task, even though only one was required, were likely because the users didn't read the instructions carefully enough or because the instructions themselves were not entirely clear. As a result, assistance was sometimes needed, but this was primarily due to uncertainty around the instructions rather than any flaws in POTATO's UX or UI design.

Overall, the software has potential, and it's clear that users get more comfortable with it as they go. By improving navigation clarity, offering smarter annotation suggestions, and implementing better UI, the tool could become more intuitive.

In the next chapter, we'll dive into the preference data collected from the questionnaires, answered by the users. This will provide even more insights into how they felt about the software and their overall experience using it.

4.4.4 User Preference Data

Participants were presented with statements after the first task and again after each subsequent task. They rated these statements on a Likert scale ranging from one (strongly disagree) to five (strongly agree). This approach allowed us to track changes in participants' preferences over the course of the test. The statements were divided into four main categories: Understandability, Usability and Navigation, Autonomy, and Customizability. The questions were presented to users in a randomized order.

Understandability

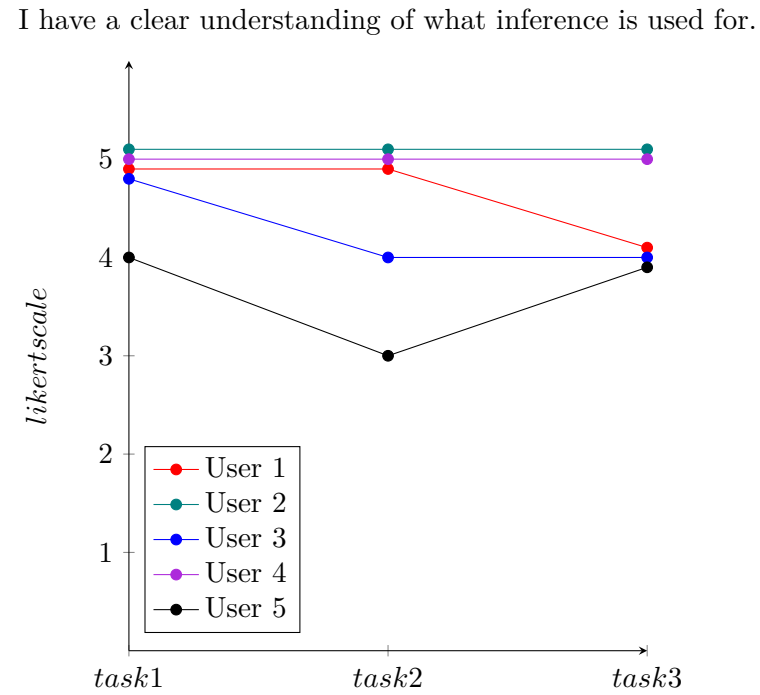
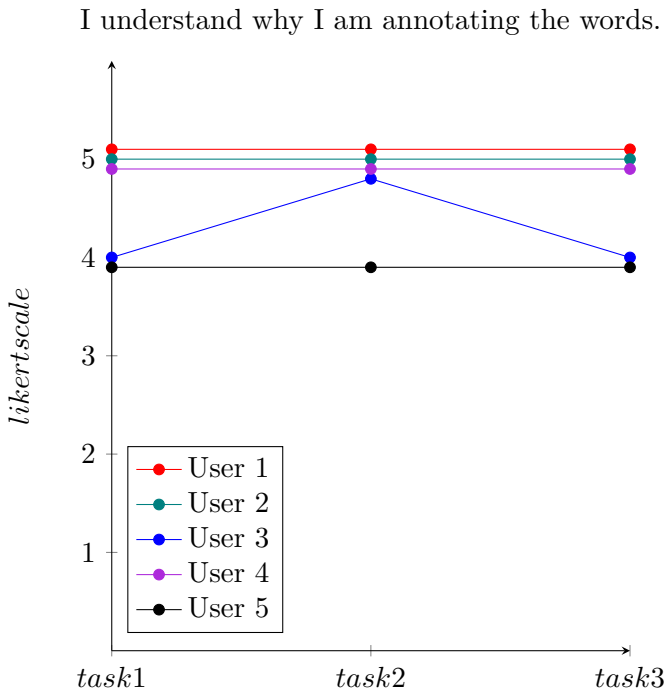
For the statement "I understand why I am annotating the words," most users consistently rated their understanding highly. Users 1, 2, and 4 gave a score of 5 across all tasks. User 5 started with a 4 in Task 1 maintaining that score till Task 3. However, User 3's responses fluctuated, starting at 4, rising to 5 in Task 2, and dropping back to 4 in Task 3.

For the second statement, "I have a clear understanding of what inference is used for," there was more variability. User 2 and 4 consistently rated their understanding as 5, showing confidence throughout. Users 3 started strong but experienced a dip in Task 2. User 1, meanwhile, began with a 5 but dropped to 4 in the final task. User 5 struggled the most, starting with a 4 in Task 1, dropping to 3 in Task 2, and improving slightly to 4 in Task 3.

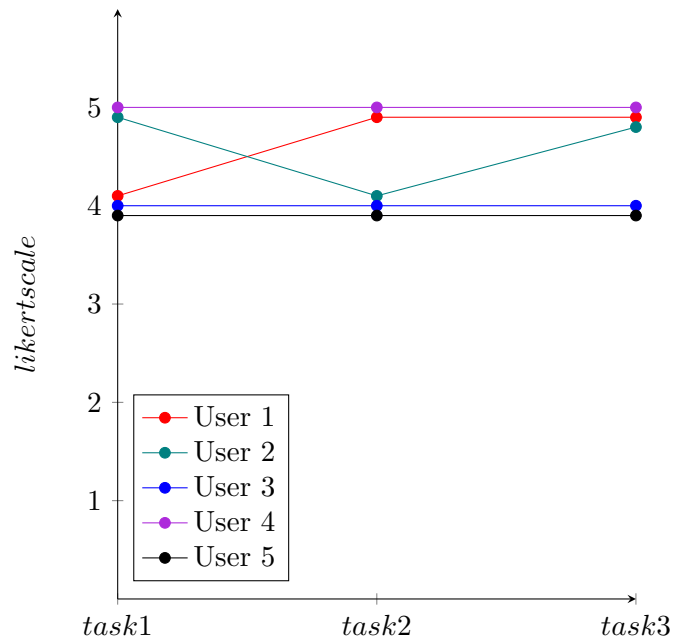
The third statement, "I understand how the words I mark relate to the output I receive after performing inference," showed mixed results. User 4 maintained a perfect score of 5, showing a strong confidence. User 1 improved from 4 after Task 1 to 5 after Task 2 and stayed there through Task 3. User 3's understanding remained steady at 4 across all tasks. User 2 started at a score of 5, dropped to 4 after task 2, but improved again after task 3, with a score of 5.

Overall, users demonstrated a good understanding of the role of annotated words, how

the inference process works, and the connection between the two. This understanding is essential for the training process. The only noticeable drop occurred after Task 2, where one user gave a score of 3. However, this score improved in the subsequent task. In addition, Task 2, which required users to critically evaluate and correct their annotations, appeared to cause temporary uncertainty. Despite this, most scores either stabilized or improved, and users regained their confidence. Interestingly, there was no significant difference in results between users with high or moderate prior experience using AI tools in generals and those with lower experience. For instance, User 2, who rated the prior experience as just a 2, was one of the most consistent participant, consistently giving a score of 5 in the "Understandability" category. This suggests that familiarity and prior experience with AI tools was not a determining factor.



"I understand how the words I mark relate to the output I receive after performing inference."

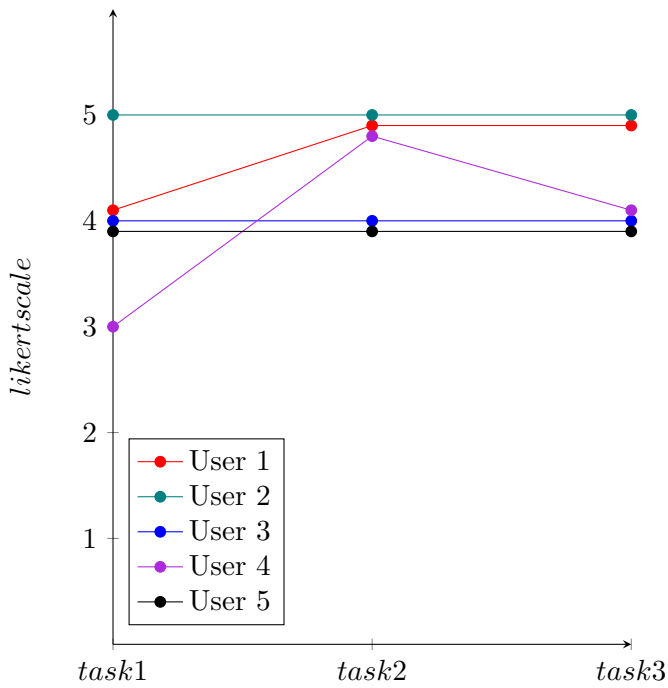


Usability and Navigation

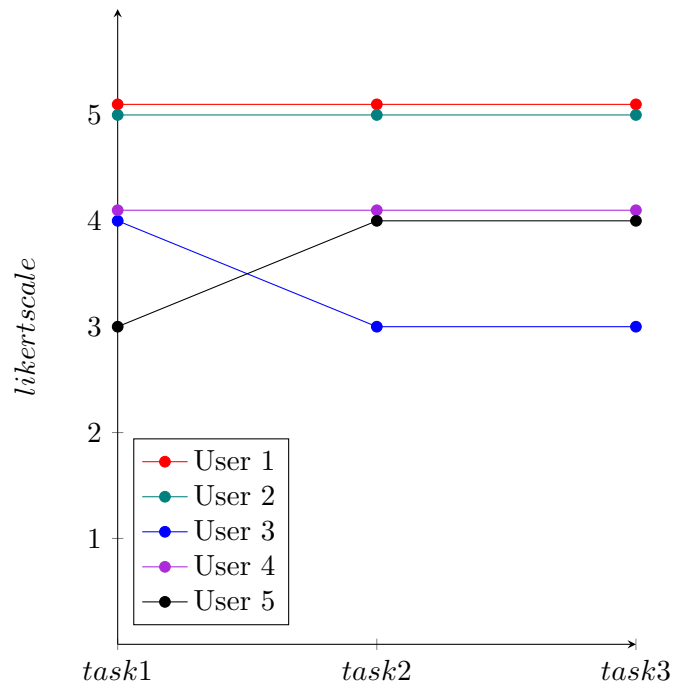
This session focuses on the software's usability and navigation. The first graph on "Response times don't hinder my workflow" highlights a positive aspect of the software. User 2 consistently rated it a perfect 5, and User 1 showed improvement over time. User 4 begins with a rating of 3, improves to 5 in task 2, but then drops back to 4 by task 3, reflecting some inconsistency in their experience. Users 3 and 5 remain consistent at 4, respectively, throughout all tasks. The graph for "The software runs smoothly" shows consistently high ratings from Users 1 and 2 (5 across all tasks). User 4 rated it slightly lower but steady at 4. User 3's satisfaction declined from 4 to 3, while User's 5 rating improved over time, rising from 3 to 4. Overall, the data indicates that response times are satisfactory for a majority of users.

The graph, "The software's navigation structure is consistent (same) across different screens," shows divided opinions. User 1 rated this aspect highly, with scores rising to 5 by the end. User 2 improved from 3 to 4. However, Users 3 remained neutral. Other participants showed slight gains, finishing with scores in the 3 to 4 range. The fourth graph, "The interface is consistent (looks the same everywhere)," paints a similar picture of variability. User 1's ratings improved over time, ending at a perfect 5. However, User 2, who started with a 5, dropped slightly as tasks progressed. User 3 showed minor improvement but remained low at 3. Other users either stayed neutral or experienced small declines. These findings point to some progress in interface design but also reveal areas where consistency could be improved.

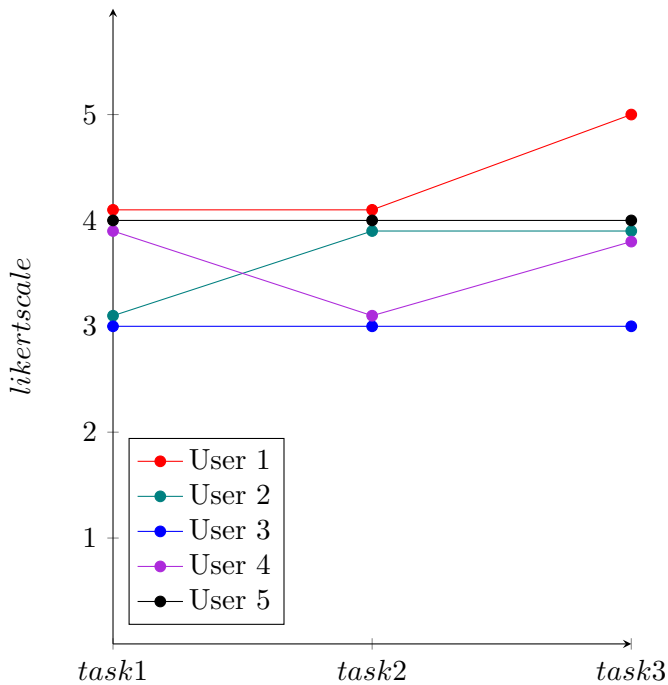
Response times don't hinder my workflow.



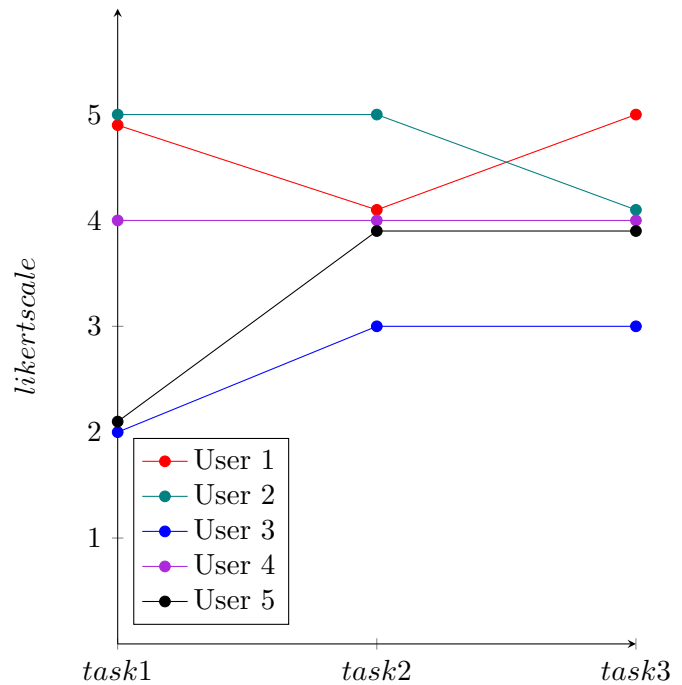
The software runs smoothly.



The software's navigation structure is consistent (same) across different screens.



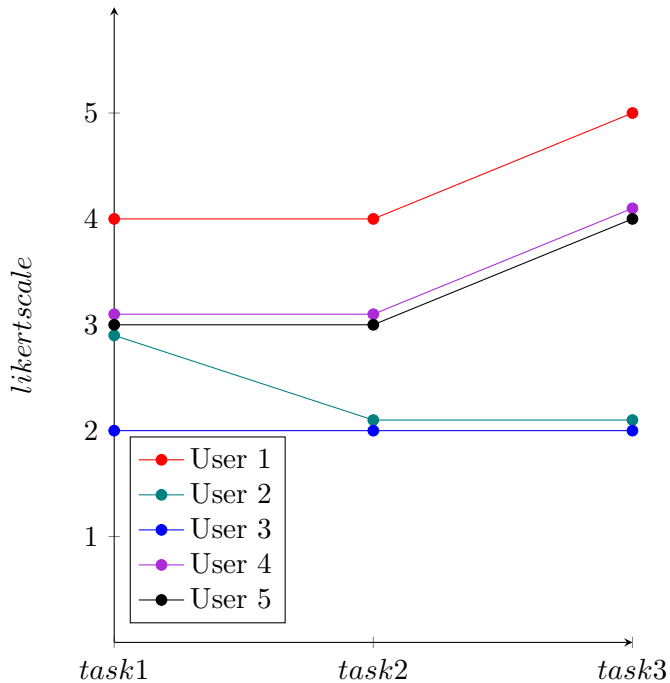
The interface is consistent (looks the same everywhere)



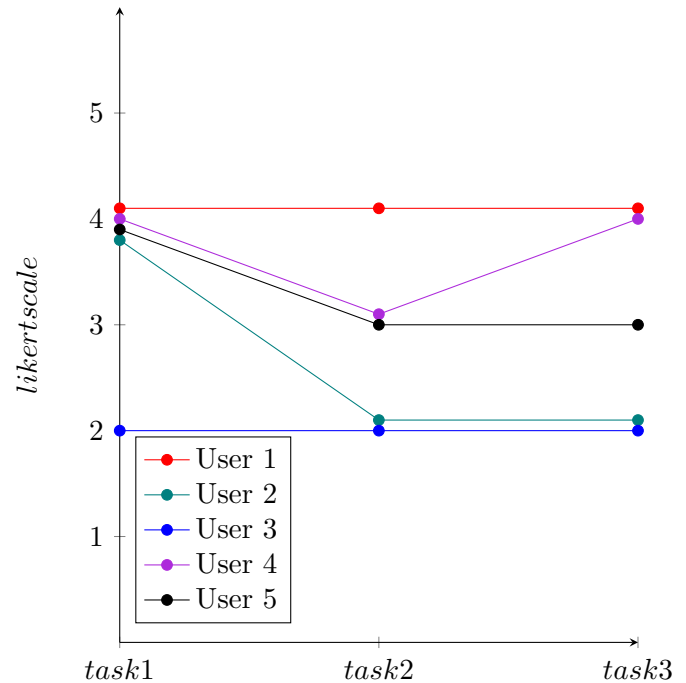
The next graph, "Navigation within the software is straightforward," shows some encouraging signs of improvement, but also ongoing struggles for certain users. User 1 showed a steady increase in satisfaction, eventually rating navigation a 5, while User 2 and 3 found it consistently poor, staying at a 2. The other participants showed slight improvements, ending at 4. In the graph focused on "I think the software is easy to use," perceptions were also highly polarized. User 1 maintained a consistently high score of 4, while Users 2 and 3 both rated it poorly at 2 throughout. User 4 initially dropped in satisfaction but recovered somewhat by the end, while User 5 stagnated at a neutral 3.

Users having significantly different experiences depending on their familiarity or expectations, and the use of the software still creates notable challenges for some users, especially early on.

Navigation within the software is straightforward.

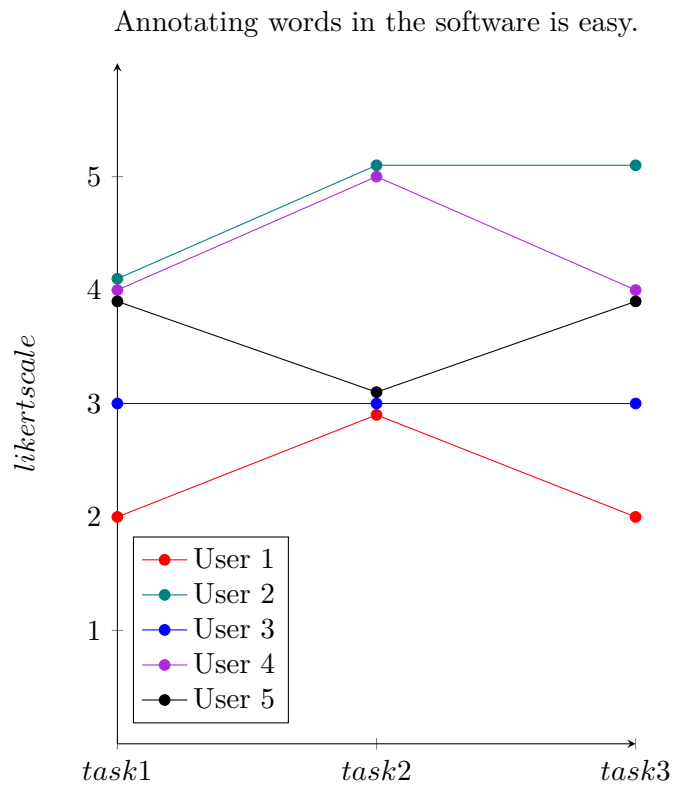
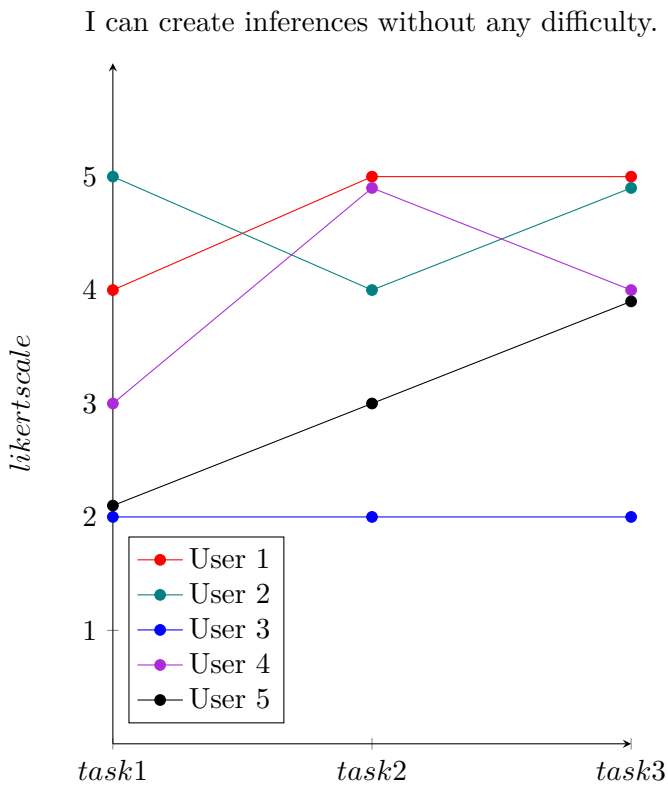


I think the software is easy to use.



The seventh graph, which examines "I can create inferences without any difficulty," demonstrates also mixed results. Users 1 and 2 found it easy, with ratings of 4 and 5 respectively. User 3, on the other hand, struggled from the start and maintained a low score of 2. User 5 showed a significant improvement, starting at 2 and with an end score of 4. This variability, according to their ratings, suggests that while inference creation works well for some, others encounter difficulties. This is interesting because, based on the observations during the session, the "create an inference" task caused no issues for most users. When it did, the problems were significantly less severe compared to any other subtask.

The last graph in this category, addresses "Annotating words in the software is easy.". User 2 rated this feature as excellent. However, User 1 was unsatisfied, with a score of 2. While some users, like User 4, showed slight improvements early on, their ratings eventually fell back. Others maintained moderate but unimpressive ratings, such as User 3's consistent 3. The users preference data support the observations. Some users were more confident and performed well with the annotation, while others were more uncertain. However, by the end, most became more confident compared to where they started. Overall, it would be worthwhile to make the annotation feature and the overall navigation of the software a bit more intuitive and user-friendly.

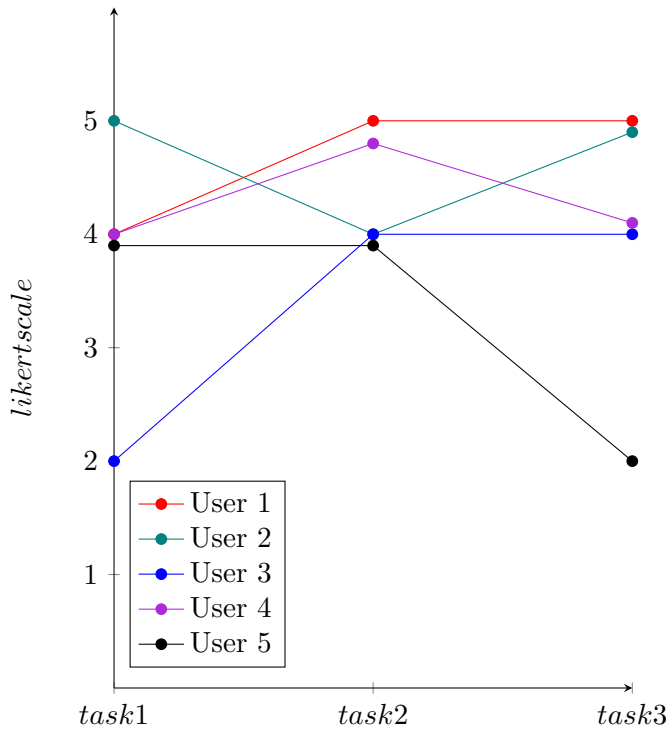


Autonomy

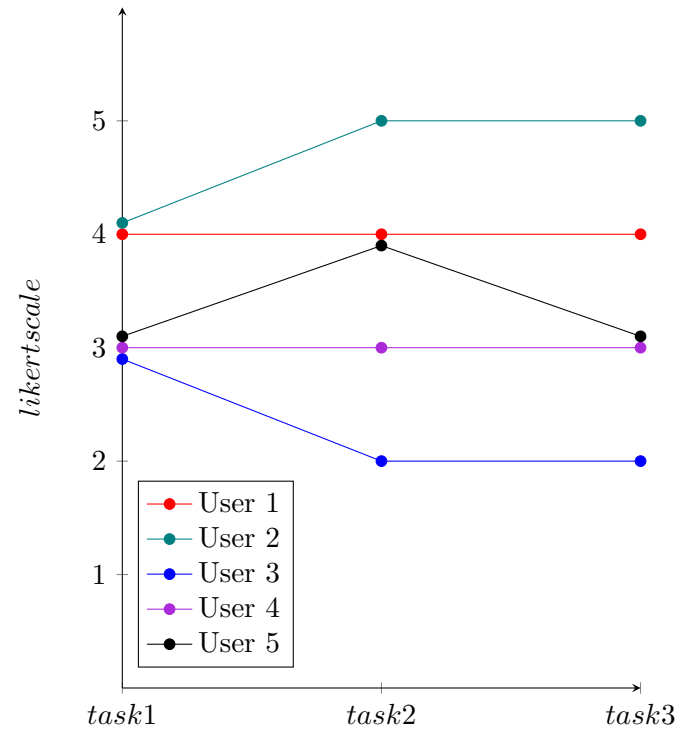
In the first graph, which focuses on the ability to autonomously complete predefined tasks, User 2 rated their experience the highest. User 1 also displayed steady satisfaction, maintaining a rating of 4 throughout. In contrast, other users showed signs of struggle. User 3 started with a moderate score of 3 but dropped to 2, where it remained, while both User 4 and User 5 were in the middle of the rating scale.

The second graph, which evaluates the statement "I can annotate words in the software on my own," highlights also strong differences between the ratings. User 2 remained confident. User 1 and User 4 followed a similar pattern, starting with ratings of 4 and showing slight improvement over time. User 3 showed significant progress, starting at 2 and improving to 4 by task 2. However, User 5 pictures the opposite trend, beginning with a confident rating of 4 but dropping sharply to 2 by the final task.

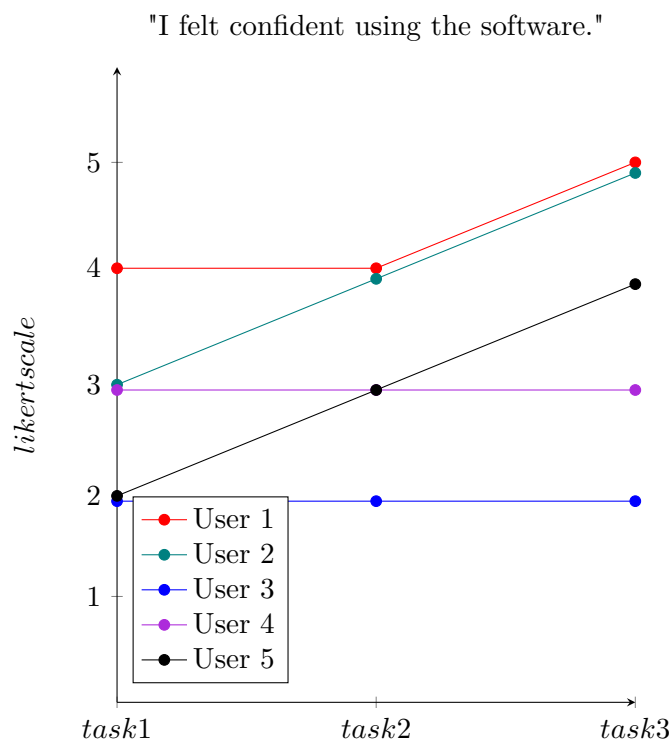
I can autonomously complete predefined tasks in the software.



I can annotate words in the software on my own.



The third graph, which examines overall confidence using the software, reflects a more positive feedback for most users. User 1, 2 and User 5 showed consistent growth in confidence, both ending with a perfect score of 5. User 3, and User 4 remained steady at 2, and 3. These results suggest that users generally gained confidence as they became more familiar with the software, although the pace of improvement varied between individuals. User 3's ratings are worth mentioning because they show a slight contradiction.



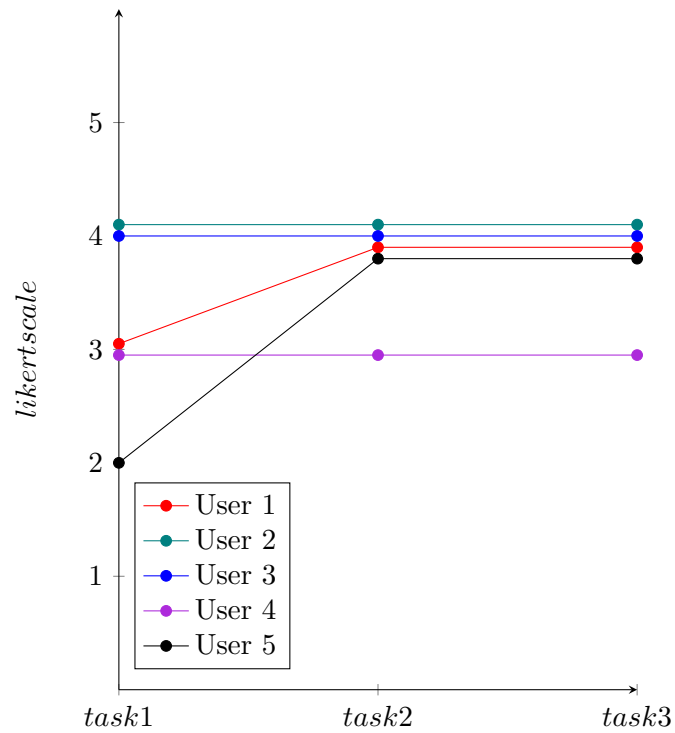
After completing the third task, user 3 rated their confidence in independently handling annotations as a 4. However, when it came to their overall confidence in using the software independently, the user gave a much lower score of 2. This could mean that the responses weren't entirely consistent, or it might suggest that while the user feel confident specifically with annotations, there is still uncertainty about using the software as a whole.

Customizability

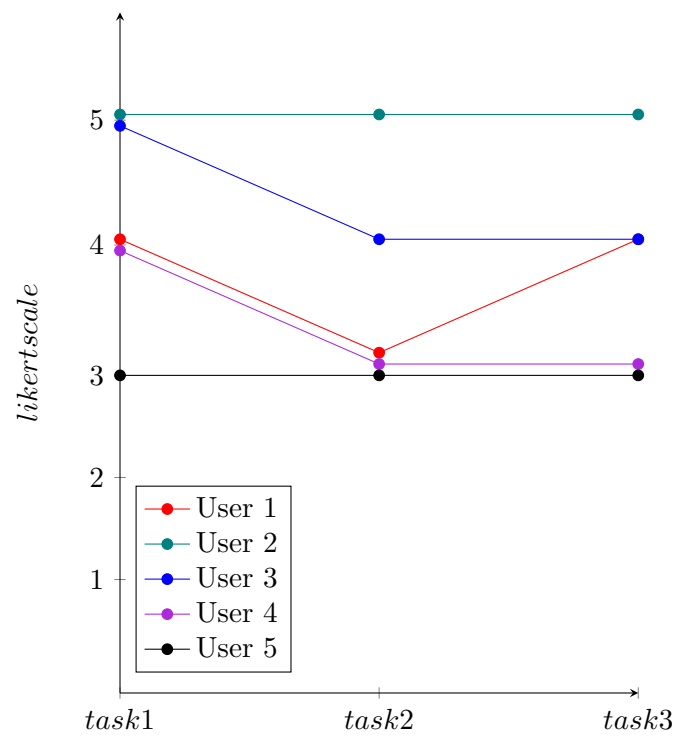
The feedback on customizability highlights generally positive perceptions. For the statement "I believe this software has been well adapted to suit our business task," most users provided good ratings from the beginning, or showed improvement. However, The only one who gave a neutral rating of 3 even after completing all three tasks was User 4. Based on this initial measurement regarding customizability, the outcome can be considered positive.

The second graph, "The software mirrors our business task effectively," shows a similar pattern or tendency as before. Compared to the first graph, there are some differences, but no notable divergences that would suggest inconsistency in the respondents' answers.

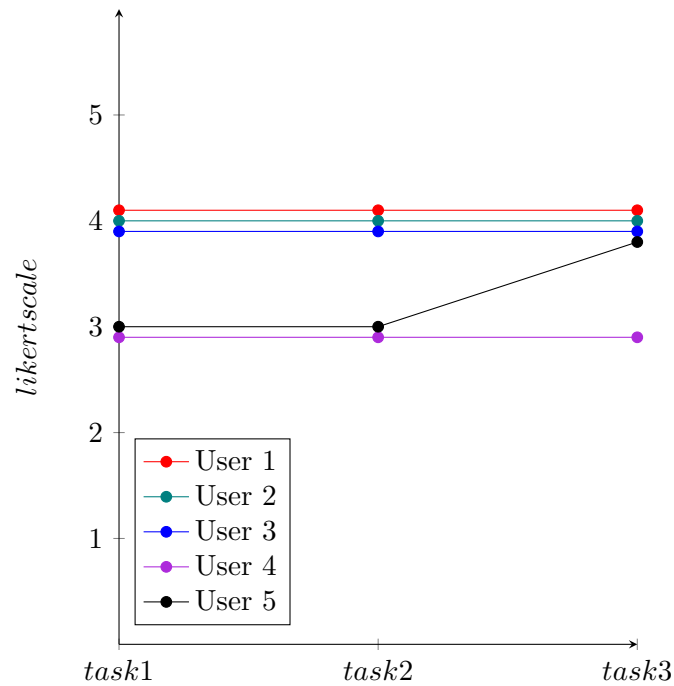
I believe this software has been well adapted to suit our business task.



The software mirrors our business task effectively.



"The software effectively addresses our specific business task."



In the final graph, "The software effectively addresses our specific business task," most users rated this aspect positively but not exceptionally. Users 1, 2, and 3 maintained a steady rating of 4, reflecting a sense of general satisfaction with the software's ability to meet specific needs. User 4 rated it lower at 3, while User 5 started at 3 but improved to 4 by the final task.

What is interesting about these results is that, unlike the other categories, the ratings here showed less fluctuation or, in other words, less improvement. This aligns more with the idea that time had less of an influence on how users perceived the customizability of the product. In most cases, users gave a positive rating from the start, which remained consistent throughout.

4.4.5 Evaluation Summary

In the following chapter, the summary table for the user preference data will be presented, showcasing the average ratings across all evaluated categories. The purpose of the previous chapter was to address Research Questions 2 and 2a. This chapter will also aim to provide answers to these research questions:

R2: In the context of Potato’s current UX design, which component or feature can be enhanced and adapted to foster a more user-friendly experience for individuals interacting with the platform?

R2.1: How well is POTATO customizable for the collaboration company’s specific business task?

The summary table (see Table 13) provides a overview of user feedback across four key categories: Understandability, Usability and Navigation, Autonomy, and Customizability, along with their respective average ratings.

Understandability stood out as the highest-rated category, with an average of 4.56. This indicates that users found it straightforward to understand the purpose of software, such as annotating words, create inferences, interpreting input and output data.

Customizability came in second, with an average score of 3.71. Users generally felt that the software was well-suited to their business tasks. However, there remains room for improvement in making the software more adaptable to a broader range of requirements.

Autonomy received an average rating of 3.62, highlighting users’ confidence in independently performing tasks. It can be said that users demonstrated a good understanding of what the software does and its purpose, including its sub-functions. But its actual usage and application, particularly in terms of autonomous use, while not showing poor average ratings, can be interpreted as having room for improvement.

Usability and Navigation ranked with an average score of 3.7. Although response times and smooth performance received positive feedback, other aspects, such as navigation, annotation ease, and overall usability, presented challenges. These areas highlight opportunities for refinement to enhance the overall user experience.

The primary area for improvement is usability and navigation. A need for a clearer and more intuitive navigation structure ist markable, which would also lead to the assumption, that a better navigation and usability features would improve autonomy usage of the software. Addressing these issues by making navigation more consistent and annotation features easier to use would significantly improve the platform’s overall usability. To specifically address research question R2a , the Customizability category achieved an average score of 3.71 out of a maximum of 5, based on the ratings of five users of Kontron Transportation GmbH. This is a clear indication that the prototype adequately covered

Statement	Rating
Understandability	
I understand why I am annotating the words.	4.7
I have a clear understanding of what inference is used for.	4.5
I understand how the words I mark relate to the output I receive after making an inference.	4.5
Average:	4.56
Usability and Navigation	
Response times don't hinder my workflow.	4.3
The software runs smoothly.	4.2
Annotating words in the software is easy.	3.6
I can create inferences without any difficulty.	3.7
Navigation within the software is straightforward.	3
I think the software was easy to use.	3.1
The interface is consistent (looks the same everywhere).	3.9
The software's navigation structure is consistent (same) across different screens.	3.7
Average:	3.7
Autonomy	
I can autonomously complete predefined tasks in the software.	3.5
I can annotate words in the software on my own.	3.6
I felt confident using the software.	3.9
Average:	3.62
Customizability	
The software effectively addresses our specific business task.	3.7
I believe this software has been well adapted to suit our business task.	3.6
The software mirrors our business task effectively.	3.9
Average:	3.71

Table 13: Summary of POTATO User Test Ratings Statements rated on a Likert scale from 1 (strongly disagree) to 5 (strongly agree)

Kontron Transportation GmbH's business case. As previously mentioned, the prototype was designed specifically for Kontron Transportation GmbH, focusing on mapping their onboarding ticket management process. However, while this result rates the software's customizability level positively, it also highlights that there is still room for further development.

Conclusion

This thesis examined the explainable AI platform POTATO in collaboration with Kontron Transportation GmbH, focusing on its usability, customizability, and efficiency within the company's onboarding ticket management process. The aim was to evaluate how well POTATO performs in real-world applications and the software's resource-saving potential through automated text-based processes.

To achieve this, three key research questions were addressed using business process simulations with the ADONIS tool and also UX testing.

The first research question explored how POTATO improves work efficiency compared to manual text processing methods. Simulations conducted in ADONIS revealed significant improvements. Ticket preparation time dropped from 2.37 hours to 0.58 hours, a 75.5 percent time saving, while costs per ticket were reduced from €289.56 to €62.29, a 78.5 percent cost reduction. On an annual scale, these improvements resulted in savings of 1,991 hours and over €252,000. These findings demonstrate that POTATO significantly enhances both time and cost efficiency. Furthermore it can be said that costs associated with POTATO are relatively low compared to the potential savings.

The second research question focused on identifying areas of POTATO's user experience that could be improved. This was addressed through UX testing, where users rated various aspects of the software on a Likert scale. The testing focused on four key categories: understandability, usability and navigation, autonomy, and customizability. Each category included topic-specific statements, which were rated by five users (employees of Kontron Transportation GmbH) and served as the basis for the preference data analyzed in this study. Collecting performance data gave us also valuable insights. On one side, it allowed us to measure efficiency under objective criteria: how long users took to complete the tasks, whether they could complete them, and if they needed assistance. These three factors—time, success, and need for help—gave us a way to evaluate task performance. At the same time, observing and documenting user behavior added another layer of understanding. Based on the outcomes, the testing highlighted navigation as

the most critical challenge, which was also reflected in lower ratings for the “autonomy” category. While users generally understood the software’s functionalities, features such as annotating words and interpreting output needs to be improved.

Finally, the third research question evaluated POTATO’s customizability. The use case scenario highlighted the software’s customizability, as the prototype aligned well with an existing business task. But research Question 2.1 places special emphasis on this topic, focusing on measuring the level of customizability. Users rated the software’s customizability positively, with an average score of 3.71 out of a maximum of 5. The software successfully aligned with Kontron Transportation GmbH’s onboarding process, but there is still room to make it more adaptable to diverse needs.

In conclusion, POTATO has shown to be a useful platform for automating text processing tasks, providing clear time and cost savings while being customizable and working well with specific business workflows. However, there is still a need to improve navigation, usability, and customization to fully unlock its potential.

5.1 Limitation and Future Work

One limitation of the thesis is the practical interpretation of the time and cost reductions calculated through the ADONIS simulation. The thesis examines POTATO, with one of the research questions addressing whether it can speed up one of Kontron Transportation GmbH’s processes through automation. The answer is yes, and this also leads to potential cost savings. However, if viewed from an economic perspective, a practical question arises. Even if the workflow is accelerated by POTATO, the company still needs to pay an 8-hour workday per employee, despite an average of x hours per day being freed up for each worker. Two potential solutions exist. The first is workforce reduction or reallocation. The second is utilizing the time freed up by POTATO to prepare additional tickets and increase ticket preparation volume. However, both scenarios raise further questions that require additional research to determine how could firms practically capitalize on the time and cost savings generated by POTATO in a meaningful way.

Another limitation is that due to time and resource constraints during UX testing, it was not possible to create and test a prototype covering the entire ticket management process. Instead, only a part of it, specifically the training process, was examined. It would be a valuable research field to further develop the current prototype and conduct another round of testing with an expanded and improved version.

List of Figures

1	User Personas for POTATO/Kontron Transportation GmbH created with Canva	19
2	Process Simulation Manual for Ticket Management at Kontron Transportation GmbH(Simulation tool: Adonis)	21
3	Result Analysis Overview manual - Ticket Management at Kontron Transportation GmbH (Simulation tool: Adonis)	23
4	Result Analysis Activities manual - Ticket Management at Kontron Transportation GmbH (Simulation tool:Adonis)	24
5	Process Simulation Automated for Ticket Management at Kontron Transportation GmbH (Simulation tool: Adonis)	28
6	Result Analysis Overview automated - Ticket Management at Kontron Transportation GmbH (Simulation tool: Adonis)	29
7	Result Analysis Activities automated - Ticket Management at Kontron Transportation GmbH(Simulation tool: Adonis)	30
8	POTATO training phase simulation	31
9	Optimistic simulation of the POTATO training phase	32
10	Pessimistic simulation of the POTATO training phase	32
11	Prototype 1 - New POTATO Demo	36
12	Prototype 2 - New POTATO Demo	37
13	Prototype 3 - New POTATO Demo <i>Note: The artificial data shown was manually constructed by Kontron Transportation to reflect the nature of tickets.</i>	38

List of Tables

1	Comparison of Design Thinking and DSRM (Own representation based on Suratno and Shafira, 2022b; Peffers et al., 2007)	12
2	Kontron Transportation GmbH Emphaty Map (Own representation based on Yeung Ng, 2023)	18
3	Ticket Preparation Time and Cost Calculations - Manual	24
4	User problem formulation	25
5	MoSCoW method (Own representation based on Schicker,2023)	27
6	Automated Ticket Preparation Time and Cost Calculations	30
7	POTATO Training Phase Costs and Time Estimates	33
8	Comparison of Manual vs. Automated Processing with POTATO	34
9	POTATO Background questionnaire results	42
10	POTATO User test task 1 performance data	43
11	POTATO User test task 2 performance data	45
12	POTATO User test task 3 performance data	46
13	Summary of POTATO User Test Ratings Statements rated on a Likert scale from 1 (strongly disagree) to 5 (strongly agree)	59

Bibliography

- Alexander, C. S., Yarborough, M. Smith, A. (2023). Who is responsible for ‘responsible AI’?: Navigating challenges to build trust in AI agriculture and food system technology. *Precision Agriculture*, 25(1), 146–185. <https://doi.org/10.1007/s11119-023-10063-3>
- Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., Schneider, N. (2013). Abstract meaning representation for sembanking. *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, 178–186. Association for Computational Linguistics. <https://www.aclweb.org/anthology/W13-2322>
- Chang, C. (2021). Improving Access to Free Online Legal Information Through Universal Design: User Personas, User Journeys, a Proposal, and a Prototype. *Legal Reference Services Quarterly*, 40(4), 199–281. <https://doi.org/10.1080/0270319x.2021.2003686>
- Du-Harpur, X., Watt, F. M., Luscombe, N. M., Lynch, M. D. (2020). What is AI? Applications of artificial intelligence to dermatology. *British Journal of Dermatology* (1951), 183(3), 423–430. doi: 10.1111/bjd.18880
- Gémes, K., Kovács, Á., Reichel, M., Recski, G. (2021). Offensive text detection on English Twitter with deep learning models and rule-based systems. *FIRE 2021: Forum for Information Retrieval Evaluation*. <https://hlt.bme.hu/media/pdf/hasoc2021.pdf>
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. In *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O’Reilly Media, Incorporated.
- Kale, A., Nguyen, T., Harris, F. C., Li, C., Zhang, J. Ma, X. (2023). Provenance documentation to enable explainable and trustworthy AI: a literature review. *Data intelligence*, 5(1), 139–162. https://doi.org/10.1162/dint_a0119
- Kornai, A. (2019). *Semantics*. Springer Verlag, Cham, Switzerland. <http://kornai.com/Drafts/sem.pdf>
- Kovács, Á., Gémes, K., Iklódi, E. Recski, G. (2022). POTATO: ExPlainable InfOrmation ExTRAcTion FramewOrk. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2201.13230>

- Lewis, J. R., Sauro, J. (2021). *USABILITY AND USER EXPERIENCE: DESIGN AND EVALUATION*. In *Handbook of Human Factors and Ergonomics* (pp. 972–1015). Hoboken, NJ, USA: John Wiley Sons, Inc. doi: 10.1002/9781119636113.ch38
- Madaio, M. A., Chen, J., Wallach, H. Vaughan, J. W. (2024). Tinker, Tailor, Configure, Customize: The Articulation Work of Contextualizing an AI Fairness Checklist. *Proceedings Of The ACM On Human-Computer Interaction*, 8(CSCW1), 1–20. <https://doi.org/10.1145/3653705>
- Miller, T. (2021). Contrastive explanation: a structural-model approach. *The Knowledge Engineering Review*, 36. <https://doi.org/10.1017/s0269888921000102>
- Nayak, A. C., Sharma, A. (2019). Explaining Black-Box Models Using Interpretable Surrogates. In *PRICAI* (1) (Vol. 11670, pp. 3–15). Switzerland: Springer International Publishing AG. doi: 10.1007/978-3-030-29908-8₁
- Nivre, J., Abrams, M., Agić, Ž., et al. (2018). Universal Dependencies 2.3. *LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL)*, Faculty of Mathematics and Physics, Charles University. <http://hdl.handle.net/11234/1-2895>
- Peppers, K., Tuunanen, T., Rothenberger, M. A. Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal Of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/mis0742-1222240302>
- Recski, G., Iklódi, E., Lellmann, B., Kovács, Á., Hanbury, A. (2024). BRISE-plandok: a German legal corpus of building regulations. *Language Resources and Evaluation*. <https://doi.org/10.1007/s10579-024-09747-7>
- Rubin, J., Chisnell, D. (2008). *Handbook of usability testing: How to plan, design, and conduct effective tests*. John Wiley Sons.
- Rudin, C., Radin, J. (2019). Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. *Harvard Data Science Review*, 1(2). <https://doi.org/10.1162/99608f92.5a8a3a3d>
- Suratno, B. Shafira, J. (2022). Development of User Interface/User Experience using Design Thinking Approach for GMS Service Company. *Journal Of Information Systems And Informatics*, 4(2), 469–494. <https://doi.org/10.51519/journalisi.v4i2.344>
- Wood, L. E. (1997). Semi-structured interviewing for user-centered design. *Interactions*, 4(2), 48–61. <https://doi.org/10.1145/245129.245134>
- Yeung, W., Ng, O. (2023). Using empathy maps to support design-thinking enhanced transdisciplinary STEM innovation in K-12 setting. *International Journal of Technology and Design Education*, 34(4), 1325–1350. <https://doi.org/10.1007/s10798-023-09861-7>

Internetquelle

ETCS | Kontron Transportation. (o. D.). <https://www.kontron.com/ktrdn/company>

Schicker, E. (2023, 18. Juli). How to use MoSCoW in UX research and avoid featuritis. Medium. <https://medium.com/usabilitygeek/how-to-use-moscow-in-uxresearch-and-avoid-featuritis-b19b8b46734d>

Statistics Austria. (o. D.). General income report. STATISTICS AUSTRIA. <https://www.statistik.at/en/statistics/population-and-society/income-and-living-conditions/general-income-report>

Sharma, I. (2024, 24. Dezember). Average salary in Austria: latest salary insights in 2025. Terratern. https://terratern.com/blog/average-salary-in-austria/?utm_source=chatgpt.com