

UNIVERSITY OF TARTU  
Institute of Computer Science  
Conversion Master in IT

**Viktorija Suslova**  
**LLMs as a Tool for User Experience Research: A  
Comparison of Synthetic and Real-World Data**  
**Master's Thesis (15 ECTS)**

Supervisors:  
Yana Halas, MSc  
Grace Eden, DPhil

Tartu 2025

# LLMs as a Tool for User Experience Research: A Comparison of Synthetic and Real-World Data

## Abstract:

This thesis examines whether large language models (LLMs) can effectively support early-stage User Experience (UX) research by generating synthetic user data. The study compares two datasets within the domain of online food delivery services: one from five semi-structured interviews with women aged 35-45 in Tallinn, and another generated using ChatGPT following the same interview guide.

Interviews from both datasets were thematically analysed, and the resulting personas and user scenarios were assessed through a side-by-side review focusing on emotional nuance, demographic accuracy, coherence, and the realism of goals, tasks, and interactions.

Results show that LLMs can produce coherent, structured, and context-relevant outputs that capture common user concerns such as navigation, delivery timing, and order modifications. Synthetic data proved useful for generating plausible personas and scenarios quickly, offering advantages in speed, consistency, and thematic breadth. However, the outputs often lacked the emotional depth, context-specific details, and behavioural variability present in real-world data. LLMs tended to generalise user behaviour, repeat similar points across participants, and occasionally introduce plausible but unfounded “false positives” that could misdirect research. Persona and scenario comparisons reinforced these trends, showing fewer concrete tasks, omission of some steps, and the inclusion of elements not mentioned by real-world participants.

The findings suggest that LLMs can complement, but not replace, traditional qualitative methods. They are best used to accelerate exploration and extend research reach when time, budget, or participant access is limited, provided outputs are refined and validated by human researchers to ensure they reflect genuine user needs.

**Keywords:** Large Language Models, UX Research, Personas, Synthetic Data, Human-Centered Design, AI-generated content, ChatGPT

**CERCS:** P170 Computer science, numerical analysis, systems, control

# LLM-id kui kasutajakogemuse uuringu tööriist: sünteetiliste ja pärismaailma andmete võrdlus

## Lühikokkuvõte:

Käesolev magistr töö uurib, kas suured keelemudelid (LLM-id) suudavad tõhusalt toetada varajases etapis kasutajakogemuse (UX) uurimist, sünteetilisi kasutajaandmeid luues. Uurimus võrdleb kahte andmestikku veebipõhiste toidukullerriteenuste valdkonnas: ühte viiest poolstruktureeritud intervjuust Tallinna 35–45-aastaste naistega ning teist, mis loodi ChatGPT abil, kasutades sama intervjuujuhendit.

Mõlema andmestiku intervjuud analüüsiti temaatiliselt ning saadud persoonasid ja kasutajastenaariume hinnati üks-ühele võrdluses, seejuures keskendudes emotsionaalsetele nüanssidele, demograafilisele täpsusele, sidususele ning eesmärkide, ülesannete ja interaktsioonide realistlikkusele.

Tulemused näitavad, et LLM-id suudavad toota sidusaid, struktureeritud ja kontekstipõhised väljundeid, mis kajastavad levinud kasutajamuresid nagu navigeerimine, kohaletoometamise aeg ning tellimuse muutmine. Sünteetilised andmed osutusid kasulikuks usutavate persoonade ja stsenaariumide kiireks loomiseks, pakkudes eeliseid kiiruse, järjepidevuse ja temaatilise ulatuse osas. Siiski puudusid väljunditel sageli reaalses andmetes esinevad emotsionaalsed sügavused, kontekstipõhised detailid ja käitumuslik varieeruvus. LLM-id kaldusid üldistama kasutajate käitumist, kordasid osalejate vahel sarnaseid punkte ning tõid aeg-ajalt sisse usutavaid, kuid alusetuid „valepositiivseid“ tulemusi, mis võivad uurimistööd valesti suunata. Personade ja stsenaariumide võrdlused kinnitasid neid tendentse, näidates vähem konkreetseid ülesandeid, teatud sammude väljajätmist ning elementide lisamist, mida reaalsed osalejad ei maininud.

Järeldused viitavad, et LLM-id võivad täiendada, kuid mitte asendada traditsioonilisi kvalitatiivseid meetodeid. Neid tasub kasutada eelkõige uurimise kiirendamiseks ja uurimisulatuse laiendamiseks olukordades, kus aeg, eelarve või ligipääs osalejatele on piiratud, eeldusel, et tulemusi täiustavad ja valideerivad inimuurijad, et tagada nende vastavus tegelikele kasutajavajadustele.

**Võtmesõnad:** suured keelemudelid, kasutajakogemuse uuring, persoonad, sünteetilised andmed, inimkeskne disain, tehisintellekti loodud sisu, ChatGPT

**CERCS:** P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine

# Table of Contents

1. Introduction.....	6
2. Background.....	8
2.1 Overview of LLM.....	8
2.2 User Research Practices in UX.....	9
3. Related work.....	12
3.1 Synthetic Data in UX Research.....	12
3.1.1 Simulating User Interviews with LLMs.....	14
3.1.2 Generating Personas and Scenarios with LLMs.....	15
4. Methodology.....	17
4.1 Data Collection.....	18
4.1.1 Real-World Data Collection.....	18
4.1.2 Synthetic Data Collection.....	19
4.2 Data Analysis.....	23
4.2.1 Data Preparation.....	23
4.2.2 Thematic Analysis.....	24
4.2.3 Data Comparison.....	25
5. Findings.....	27
5.1 Real-World Data Insights.....	27
5.2 Synthetic Data Insights.....	28
5.3 Comparative Analysis.....	30
5.3.1 Interview Results Comparison.....	31
5.3.2 Persona and Scenario Comparison.....	38
6. Discussion.....	43
6.1 Recommendations.....	45
6.2 Limitations.....	46
7. Future Work.....	48
8. Conclusion.....	49
References.....	50
Appendices.....	56
Appendix I: Interview guideline.....	56
Appendix II: Interview consent form.....	58
Appendix III: Interview questions.....	60

Appendix IV: Interview responses.....	61
Appendix V: List of Activities Performed to Choose AI Tool for UX Interview .....	70
Appendix VI: List of AI tools for simulation UX interviews.....	71
Appendix VII: List of Activities Performed to Choose AI Tool for Creating UX Persona	74
Appendix VIII: List of AI tools for creating UX personas .....	75
Appendix IX: List of Activities Performed to Choose AI Tool for creating Scenario .....	78
Appendix X: List of AI tools for creating UX scenarios .....	79
License .....	81

# 1. Introduction

As artificial intelligence (AI) has evolved from its origins, it has touched almost every sector, influencing how organisations operate, make decisions, and engage with technology. One of the most notable changes is the acceleration of information processing, reshaping how people and systems interact. AI tools are now embedded across industries, influencing how tasks are completed, services are delivered, and decisions are made. For organisations, adapting to these developments is essential to remain competitive [1].

In product and service design, AI is increasingly influencing how teams operate. Within this, User Experience (UX) research plays a critical role in uncovering what users need, the challenges they face, and the expectations they bring to an experience. This knowledge drives better product decisions and improves satisfaction by making designs more intuitive and relevant [2].

Among AI systems, large language models (LLMs) have attracted particular attention. Trained on vast datasets, they can generate and interpret text in ways that resemble natural human conversation. Models such as GPT-4, Claude, and Gemini are already being tested in design workflows to create personas, simulate interview responses, and produce user scenarios. Their ability to process language quickly makes them especially appealing for early-stage research, where ideas must be explored and iterated at speed [3].

However, speed does not guarantee depth. LLMs may miss emotional nuance or the richness of lived experience. Although adoption is expanding across fields, their full strengths and limitations are still being uncovered. Researchers caution against overestimating their intelligence or intent, noting that these systems can produce convincing yet shallow responses that overlook ethical or contextual complexity [4]. This reflects the views of Norman and Shneiderman, who argue that while technology evolves quickly, meaningful integration into design requires reflection, user-centred thinking, and iterative testing [5, 6].

The main aim of this research is to examine whether LLMs can effectively contribute to UX research by producing synthetic user data. Two research questions guide the study:

**RQ1.** *To what extent can LLMs be used to generate synthetic user research data such as interviews, personas, and user scenarios?*

**RQ2.** *How does synthetic data compare to real-world data for understanding user needs?*

To answer these, two datasets were created: one from real-world participants and one from synthetic users, both centred on experiences with online food delivery service. This domain was chosen for its broad familiarity, ensuring participants could draw on first-hand experience. The real-world dataset comprised five semi-structured interviews with women aged 35-45, while the synthetic dataset was generated with ChatGPT using the same interview guide. Both datasets were thematically analysed [46] and used to create personas and user scenarios, enabling direct comparison.

A consistent evaluation framework was applied to both datasets. Interviews were analysed across four key areas: order cancellations, suggested improvements, pain points, and most valued features. Personas and scenarios were assessed for demographic accuracy, grammatical quality, emotional nuance, and the realism of the tasks and interactions described.

The thesis is structured as follows: Chapter 2 introduces the core concepts of LLMs and established UX research methods. Chapter 3 reviews existing studies on synthetic data in UX contexts. Chapter 4 outlines the research design, detailing the data collection and analysis processes. Chapter 5 presents the comparative findings. Chapter 6 discusses the implications, limitations, and recommendations for practice. Finally, Chapter 7 summarises the contributions and identifies opportunities for future research.

This study contributes to the growing debate on AI's role in UX workflows. Rather than positioning LLMs as replacements for traditional research, it evaluates their potential as complementary tools - particularly when time, participant access, or budget are constrained - while emphasising the continued importance of human oversight.

This thesis made use of OpenAI's ChatGPT<sup>1</sup> for grammar refinement, wording and structural suggestions.

---

<sup>1</sup> OpenAI. (2025). ChatGPT (August 8 version) [Large language model]. <https://chat.openai.com/>

## 2. Background

To understand the potential role of LLMs in UX research, it is necessary to first establish a shared foundation of key concepts and methods. This chapter introduces the technical principles behind LLMs, outlining how they generate content, and the limitations that influence their outputs. It also reviews established UX research methods, highlighting how data is typically collected, analysed, and transformed into design artefacts such as personas and scenarios. Together, these sections provide the conceptual background that informs the research design presented in Chapter 4.

Section 2.1 introduces LLMs, outlining their underlying architecture, training processes, key capabilities, and examples of current state-of-the-art models, as well as their strengths and limitations. Section 2.2 reviews established UX research practices, with particular attention to qualitative methods such as interviews, persona development, and scenario creation, and explains how these artefacts inform design decisions throughout the product development process.

### 2.1 Overview of LLM

LLMs are sophisticated AI systems designed to handle and produce human language in a way that mirrors natural conversation. They play a central role in Natural Language Processing (NLP), a discipline focused on enabling machines to understand and respond to human language with awareness of both context and nuance. LLMs support a wide range of well-known NLP applications, such as machine translation, sentiment detection, automated text condensation, and answering user queries [7, 8].

These models are typically built on deep learning architectures, most often using transformer-based neural networks. They are trained on vast, diverse datasets drawn from sources such as websites, books, and news articles. During training, the model learns to predict the most likely next word in a sequence based on the surrounding context [9, 3]. Repeating this process billions of times gradually fine-tunes the model's internal parameters, giving it a broad understanding of grammar, meaning, and context. As a result, it can generate coherent and contextually appropriate replies for a broad variety of prompts.

A pivotal breakthrough in the development of LLMs was the introduction of the transformer architecture [10]. Unlike earlier models that handled text in a step-by-step manner, transformers use self-attention mechanisms to examine all words in a sentence simultaneously,

identifying relationships between them regardless of position. This enabled more efficient handling of long or complex texts, and the transformer has since become the standard architecture for nearly all modern language models.

Today's most advanced systems are often referred to as state-of-the-art (SOTA) models, meaning they achieve leading results on recognised benchmarks such as MMLU, BIG-bench, and HumanEval. Examples include OpenAI's GPT-4 and GPT-4o, which demonstrate strong reasoning abilities and support multimodal inputs such as text, images, and audio. Other major contenders are Anthropic's Claude 3 family, Google DeepMind's Gemini 1.5, Meta's LLaMA 3, and Mistral's Mixtral, each offering distinctive strengths [11]. These models are increasingly capable of working with long input contexts, following complex instructions, and producing high-quality, domain-specific outputs.

Modern LLMs can do far more than predict the next word in a sentence. They can compose essays, summarise information, translate between languages, generate and explain code, and tackle multi-step reasoning problems. Many are fine-tuned using reinforcement learning with human feedback (RLHF), a process in which human preferences shape the model's behaviour to make responses more helpful, accurate, and aligned with user needs [12].

While these models are often highly effective, they are not without limitations. They can produce inaccurate information, generate biased or inappropriate content, or fail to interpret a question in the way the user intended. As with any powerful tool, thoughtful use is essential to ensure outputs are reliable and ethically sound [12].

## **2.2 User Research Practices in UX**

User research is a meaningful part of product and service design. Its primary purpose is to understand users' needs, behaviours, and goals so that products and services can be designed to serve them effectively. By grounding design decisions in real-world insights, user research improves usability, accessibility, and overall satisfaction [13, 5].

Research in UX involves a range of activities, each aimed at uncovering different aspects of user needs and challenges. Methods such as interviews, field observations, and workshops are often used to explore the broader context in which users operate, revealing pain points, motivations, and opportunities for improvement [14].

In the early stages of a project, this work relies heavily on qualitative methods such as user interviews, persona creation, and scenario design. These techniques provide a holistic view of

user perspectives but can be time-consuming and resource-intensive. While larger datasets can yield richer insights, collecting and analysing them demands substantial effort. Interviews, in particular, are highly valued because they uncover deep insights into users' goals, tasks, frustrations, and expectations.

Once interview data is collected, it must be synthesised into actionable design artifacts. A persona usually captures key details such as demographics, goals, frustrations, skills, and the context in which a product is used, giving design teams a clear and relatable picture of their users. Goodwin [15] describes personas as focused, believable profiles that bring target user groups to life and guide design choices. Likewise, the Nielsen Norman Group [47] highlights how personas distill important user characteristics into an easy-to-grasp format, helping teams keep real user needs front and centre throughout the design process.

Personas are often paired with scenarios – short narratives that illustrate how a persona might interact with a product or service in a specific situation. Each scenario outlines the context, the user's goals, the steps they take, and the challenges they encounter, always from the user's perspective. This helps designers identify what the system must do to support those goals effectively [16].

Scenarios transform static personas into living stories, making them more actionable. Without scenarios, personas remain abstract; with them, they become dynamic tools for defining requirements and communicating design intent.

Based on the personas and scenarios, teams can generate ideas aimed at creating meaningful products that address identified user needs. These ideas are then translated into prototypes, which are evaluated to uncover usability issues using methods such as moderated usability tests, task analysis, and performance metrics [17]. Following this evaluation, post-launch studies assess whether the product continues to meet user needs, often through analytics, surveys, and follow-up interviews [18].

While these practices have long been central to UX, they demand significant time and resources. This has led to growing interest in integrating LLMs into workflows – particularly for early-stage activities such as simulating interviews, drafting personas, or sketching scenarios. The goal is to explore whether AI can accelerate the process, extend research capabilities, and reduce the barriers to collecting and structuring user insights.

By mapping out the fundamentals of both LLM technology and UX research practices, this chapter has set the stage for the literature review in Chapter 3. Understanding the strengths and constraints of these domains is essential for assessing whether synthetic data can be meaningfully compared with insights from real-world participants.

### **3. Related work**

This chapter looks at the current state of research on synthetic data generation, with a particular focus on its use in UX. It reviews recent work on AI-assisted research workflows, considers the benefits and risks of using LLM-generated content, and highlights how prompt design has emerged as a key factor in shaping output quality. In doing so, it identifies gaps this thesis aims to address - especially around how reliable and realistic synthetic interviews, personas, and scenarios are when compared to their real-world counterparts.

Section 3.1 defines synthetic data and explores its use in different fields before focusing on its growing role in UX research. It outlines how LLMs can create realistic user responses, personas, and scenarios to support early-stage design when traditional research is constrained. Section 3.1.1 examines LLM-simulated user interviews, showing how they can produce realistic transcripts that feed into persona and scenario development. It summarises existing studies on their strengths and weaknesses, underscores the importance of well-crafted prompts, and frames them as a supplement rather than a replacement for human-led research. Section 3.1.2 looks at generating personas and scenarios with LLMs, drawing on comparative studies and advanced techniques. It notes their ability to scale artefact production in exploratory design, while warning against generic or stereotypical results and stressing the value of human review.

Taken together, these sections show that LLMs are already finding a place in UX research, offering new ways to speed up early-stage work and spark creative ideas. At the same time, they make it clear that AI outputs still need careful interpretation, and that human expertise remains essential for turning them into insights that truly reflect user needs.

#### **3.1 Synthetic Data in UX Research**

This section reviews key studies exploring synthetic data generation in UX workflows, focusing on personas, scenarios, and simulated interviews. Synthetic data refers to data that is artificially generated rather than collected directly from users, yet it aims to reflect realistic patterns in structure, tone, or statistical properties.

This approach has already proven useful in domains such as healthcare and finance, where privacy restrictions or the sensitivity of data make traditional collection methods difficult. For instance, in healthcare, synthetic patient records have been used to test systems and algorithms while protecting real patient identities [19]. Similarly, in data science and machine learning,

synthetic datasets are often created to fill gaps, simulate rare scenarios, or balance imbalanced data distributions [20].

In UX research, synthetic data is gaining traction as a practical way to simulate user responses, model hypothetical behaviors, and generate plausible personas using generative tools like LLMs. These methods are particularly valuable during early design phases - especially when it is difficult, expensive, or time-consuming to involve real-world participants. As Wang et al. [21] point out, synthetic data creation at scale can significantly reduce time and cost while still supporting exploratory design work. By offering flexibility and speed, LLM-generated content can help bridge gaps in early-stage research without immediately relying on full recruitment cycles.

LLMs such as GPT-4, Claude, Gemini, and Perplexity represent significant progress in the field of natural language processing. These models are capable of generating coherent, context-aware, and human-like text across a broad range of topics [22]. Their advanced language generation abilities have led to adoption in diverse professional domains, including UX design [23]. Tools like ChatGPT have also popularized dialogue-based interaction, making AI systems more intuitive and accessible for a wider range of users, which in turn supports more streamlined and interactive workflows in UX research.

Recent research is beginning to position LLMs not just as design tools, but as active collaborators in the creative process. In one standout study, Kocaballi [24] put ChatGPT through its paces in multiple roles - designer, user, and even the product itself - within a hypothetical design project. The model didn't just participate; it generated personas, staged fictional interviews, proposed fresh design ideas, built usage scenarios, and even critiqued a prototype app. While ChatGPT handled many of these tasks impressively, the study also flagged some clear gaps: it sometimes forgot key details, produced incomplete answers, and offered limited variety in its output. Together, these results suggest that LLMs can meaningfully enrich early-stage UX workflows - so long as their weaknesses are understood and managed.

Different LLMs excel in different areas, so researchers and designers benefit from adopting flexible, tool-agnostic workflows rather than committing to a single model. Such approaches allow teams to choose the LLM that best fits the specific task - one model might be ideal for quickly producing large volumes of content during ideation, while another may deliver more accurate results for precise information retrieval or fact-based tasks. Building research

processes that leverage the unique strengths of multiple models fosters adaptability and can improve both the quality and efficiency of UX work. Ahmed and Imran highlight the value of selecting AI tools strategically, taking into account factors such as task complexity, desired output, and user context [25].

### **3.1.1 Simulating User Interviews with LLMs**

Traditionally, user interviews in UX research have required direct access to participants, careful recruitment, and time-intensive scheduling. While this method remains the gold standard for gathering authentic, context-rich insights, it can be constrained by budget, availability, and logistical challenges. As a result, the early stages of design - where rapid feedback can guide idea generation - often face bottlenecks.

With the emergence of LLMs such as ChatGPT, GPT-4, Claude, and Gemini, researchers now have the ability to simulate user interviews through detailed, context-specific prompts. These models can generate responses that mimic the characteristics, goals, and behaviors of a targeted user group, making them a valuable tool for developing initial hypotheses, identifying pain points, and iterating on product ideas - particularly when real-world participants are unavailable or resources are limited [27, 28].

Recent studies illustrate the potential of this approach. Zhang et al. [26] demonstrate how GPT-4 can be prompted to produce rich, interview-style transcripts that resemble human-authored dialogue, providing structured input that can jumpstart persona creation or scenario planning. Hasanah and Yuniarto [30] assessed GPT-3.5's performance in simulated survey interviews, reporting a 76.16% accuracy rate in capturing key response patterns. Hämäläinen et al. [31] similarly found that GPT-3 could produce questionnaire responses closely aligned with those of real-world participants, suggesting the feasibility of emulating user-like behavior in structured contexts.

More advanced frameworks have begun to push this capability further. Liu et al. [37] developed the CLUE framework (Conversational LLM-based User Experience) using GPT-4o for real-time, dialog-based interviews, showing strengths in generating surface-level feedback but weaker performance in detecting emotional subtleties. Wuttke et al. [38] tested GPT-4 as an adaptive interviewer in a controlled survey setting, finding comparable engagement and

credibility to human interviewers, but with responses that were often more generic and less responsive to unexpected inputs.

Despite these advances, limitations remain. LLM-generated interviews can lack emotional depth [4], spontaneity, and cultural nuance [32]. Meyer et al. [33] observed that GPT-3 outputs underperformed in variability and depth when compared to real-world data, particularly in tasks requiring spontaneous elaboration. Roberts et al. [34] caution that over-reliance on synthetic interviews may inadvertently narrow the diversity of perspectives and reinforce generalized or biased assumptions. Liu et al. [35] further note that LLMs may struggle to capture sensory or experiential feedback - critical for understanding certain interaction contexts.

Prompt design emerges as a critical factor influencing output quality. Well-crafted, role-specific prompts can improve contextual accuracy, while vague instructions tend to produce stereotypical or overly generalized responses. This aligns with recent findings showing that only about half of performance gains after switching to a more advanced AI model come from the model itself, while the other half is driven by how effectively users adapt their prompts to the system [48].

In practice, LLM-simulated interviews are best viewed as complementary tools rather than replacements for human-led sessions. When integrated thoughtfully - guided by clear research goals and paired with human oversight - they can accelerate early-stage exploration, provide structured starting points for analysis, and extend the reach of UX teams working under time or resource constraints. However, their use should be balanced with traditional methods to ensure depth, authenticity [29], and to preserve diverse perspectives [36].

### **3.1.2 Generating Personas and Scenarios with LLMs**

Traditionally, the development of personas and scenarios in UX design has relied on extensive qualitative research and expert synthesis. This process, while valuable, is often labor-intensive and time-consuming. With the rise of LLMs, parts of this workflow can now be accelerated - or in some cases, partially automated. However, important concerns remain around the authenticity, emotional depth, and contextual fidelity of synthetic outputs. Given that personas and scenarios both emerge from qualitative insight and play complementary roles in design, this section examines their joint creation using LLMs as part of evolving UX research practices.

In a blind study, Schuller et al. [39] compared personas generated by ChatGPT with those created by UX professionals. The results indicated that, when guided by well-structured

prompts, LLM-generated personas were rated as equally usable and believable by practitioners. The authors emphasized that prompt specificity - such as including user demographics, tone of voice, and structural format - plays a critical role in improving the relevance and clarity of the results.

Building on this, Buchmüller et al. [40] proposed a framework that uses LLMs to generate not just personas, but also simulate user experiences and feedback scenarios. Their approach enables rapid and iterative ideation, grounded in structured - though artificial - empathy. The study highlights the importance of role-specific prompting and scenario design to preserve contextual accuracy and realism in generated content.

Zhang et al. [41] introduced a more advanced system that integrates GPT-4, DALL·E 2, and knowledge graphs to produce multi-modal personas combining textual and visual elements. Aimed at UX education, their system allowed students to explore and develop personas more efficiently, improving engagement, content accuracy, and diversity when compared to conventional classroom methods.

These practices position LLMs as useful tools for scaling persona creation, especially in the exploratory stages of design. However, even with their structured output, persona templates and scenarios produced by LLMs can drift toward generic or stereotypical representations if not directed with carefully crafted, context-specific prompts. Human judgment therefore remains essential - not only to refine and validate these outputs, but also to ensure they capture diverse perspectives and accurately represent the intended user groups [42].

The reviewed literature suggests that while LLMs show promise as research aids, their use in UX practice is still limited and mostly documented through isolated examples. This underscores the value of the structured, side-by-side evaluation presented in Chapter 4, which compares real-world and synthetic datasets using consistent methods and criteria.

## 4. Methodology

This chapter describes the methodological approach taken to investigate the research questions outlined in Chapter 1. It details the processes for collecting and analysing both real-world and synthetic datasets, ensuring that comparisons are valid and transparent. The chapter also explains the tools, sampling strategies, and evaluation criteria applied to the data, as well as the rationale for selecting specific AI systems for generating synthetic interviews, personas, and scenarios.

The process was designed to enable a systematic comparison between real-world and synthetic content. To keep the context familiar and relevant, both datasets focused on experiences with online food delivery services. This ensured that participants could draw from real-world situations, while the LLM-generated responses could be meaningfully compared against authentic user perspectives.

Figure 1 outlines the main stages of the study.

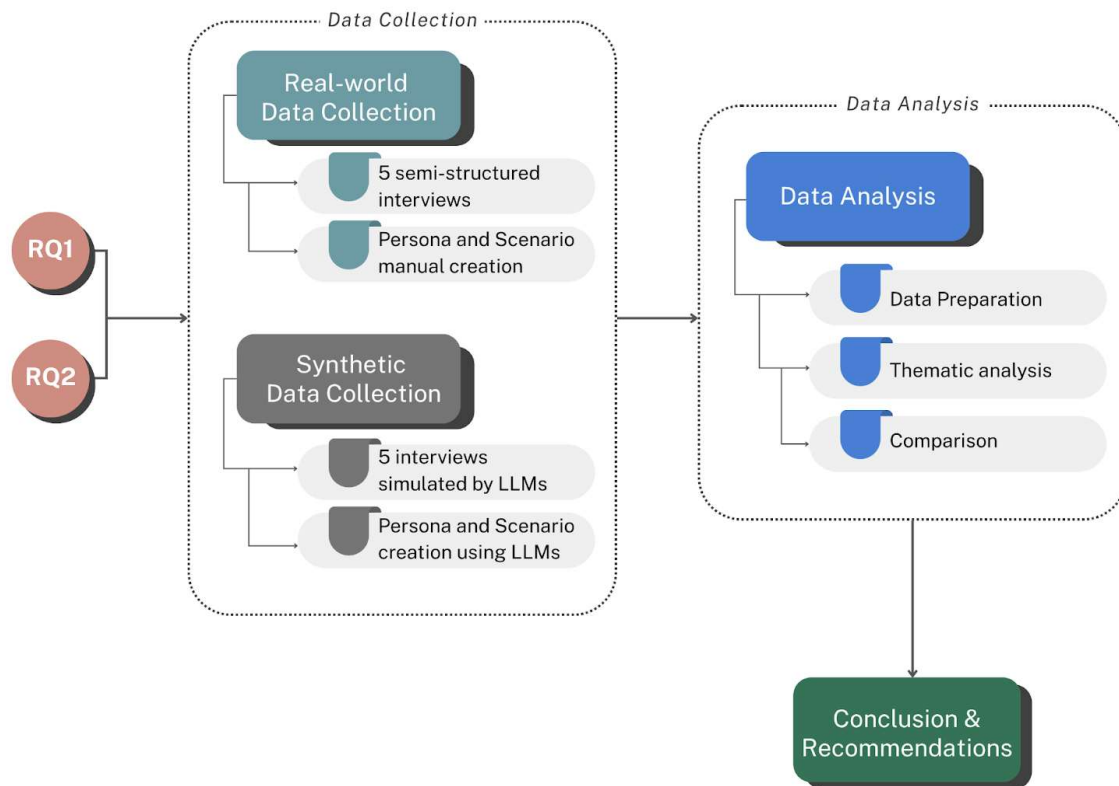


Figure 1. Research process.

Section 4.1 describes the parallel collection of real-world and synthetic datasets, including interviews, personas, and scenarios, along with the selection and use of LLM-based tools, guided by structured prompts where appropriate. Section 4.2 outlines the preparation and analysis of both datasets, detailing the thematic analysis in ATLAS.ti<sup>2</sup> and the criteria-based comparison of interviews, personas, and scenarios to ensure a fair and replicable evaluation.

## 4.1 Data Collection

The study collected two parallel datasets to enable a direct comparison between real-world and synthetic user insights. The first dataset came from real-world participants through semi-structured interviews, while the second was generated synthetically using LLM (ChatGPT). Both followed the same thematic focus - online food delivery services - ensuring a consistent basis for analysis. The process for each dataset included three key stages: conducting interviews, creating personas and user scenarios. By mirroring these steps across both real-world and synthetic sources, the research ensured that differences in output could be attributed to the origin of the data rather than the method of production.

### 4.1.1 Real-World Data Collection

**Interview process:** Five semi-structured interviews were conducted with female participants aged 35–45 residing in Tallinn, all of whom had experience using online food delivery services. Convenience sampling was used due to the exploratory nature of the study and the need to recruit relevant participants within a limited timeframe. While this approach limits the generalisability of the findings, it enabled the collection of targeted insights from individuals who met the study’s criteria.

An interview guide was developed to explore the following:

- Useful and difficult app features
- Experiences with order modification
- Order cancellation
- Navigation issues
- Feature suggestions and general feedback

Details of the conducted interviews can be found in the appendices: the interview guideline is included in Appendix I, the participant consent form is provided in Appendix II, and the full

---

<sup>2</sup> <https://atlasti.com/>

list of interview questions is available in Appendix III. These materials clarify the exact topics discussed with participants and support the transparency of the research process.

Interviews were recorded using the built-in voice recording feature on a mobile phone. The same device was then used to transcribe the recordings through its automatic speech-to-text functionality. After transcription, the text was reviewed and manually corrected for clarity and accuracy. All transcribed data was anonymized to protect participant privacy and was then organized for systematic coding during the qualitative analysis process.

**UX persona creation:** Using insights from the interviews, detailed UX personas were developed, including demographics, frustrations, goals, needs, and expectations. Although personas traditionally include a representative image to make them more relatable and engaging for stakeholders, as recommended by Nielsen Norman Group [47], in this research the real-world personas did not feature one. This decision was made to focus solely on textual data, ensuring a more efficient and consistent comparison between datasets.

**User scenario creation:** Common user scenario was developed to represent realistic contexts, tasks, and challenges associated with online food delivery services. Each scenario focused on a specific user goal and described the typical interaction steps within the app to achieve that objective.

#### 4.1.2 Synthetic Data Collection

**Interview process:** the synthetic data collection initially started with the overview of the existing interview simulation tools. A full record of the activities undertaken to identify relevant applications can be found in Appendix V.

Table 1. Analyzed AI tools for simulating UX interviews.

Tool's name	Best for	Link
<b>ChatGPT (OpenAI)</b>	General-purpose, quick-turnaround idea testing, refining interview scripts, early-stage concept validation	<a href="https://openai.com/chatgpt">https://openai.com/chatgpt</a>
<b>User Interview Simulator</b>	Simulating realistic user personas for early product feedback and question refinement	<a href="https://www.yeschat.ai/">https://www.yeschat.ai/</a>
<b>Delve AI's Synthetic Research</b>	Scalable, multi-persona simulated interviews for directional insights before live testing	<a href="https://www.delve.ai/">https://www.delve.ai/</a>

<b>UXPressia</b>	Testing and refining customer journeys, messaging, and UX flows through live persona interactions	<a href="https://uxpressia.com/ai-persona-chat">https://uxpressia.com/ai-persona-chat</a>
<b>Heysho</b>	Gathering diverse persona perspectives quickly for early-stage concept validation and research training	<a href="https://heysho.com/en/tool/user-interview.html">https://heysho.com/en/tool/user-interview.html</a>

Table 1 summarises the AI tools identified and reviewed for their potential to support conversational research tasks, with the full analysis available in Appendix VI. These tools enable AI-driven user research interviews, allowing UX researchers to collect insights efficiently, reduce manual effort, and simulate realistic user interactions without relying on real-world participants.

For this study simulated user interviews were conducted using ChatGPT. ChatGPT was chosen because it demonstrated the most consistent ability to produce human-like, context-aware responses suitable for simulating user interviews. The interviews were carried out in a guest browser session using the free version of GPT-4, without logging into an account or adjusting advanced settings, to reflect a typical user experience. To guide the simulations, prompts were structured using the RTF (Role, Task, Format) framework. A typical prompt example:

*"As a UX researcher, I'm trying to better understand the users of online food delivery services. Act like a 37-year-old married woman with two children, living in Tallinn and working as a treasury specialist. Answer the following user interview questions..."*

The simulated interviews were then transcribed and added to ATLAS.ti for thematic analysis [46].

**UX persona creation:** The creation process started with the initial exploration involved identifying existing tools. This was conducted via both traditional search engines (Google) and interaction with ChatGPT to uncover available design-support tools and workflows. A full record of the activities undertaken to identify relevant applications can be found in Appendix VII.

Table 2. Analyzed AI tools for creating UX personas.

Tool's name	Best for	Link
<b>AI User Persona Generator by Bot Memo</b>	Quick, free creation of text personas for later visual or data-backed refinement	<a href="https://botmemo.com/ai-user-persona-generator">https://botmemo.com/ai-user-persona-generator</a>
<b>ChatGPT (OpenAI)</b>	Flexible, conversational creation of text-based personas that can be iteratively refined through interactive prompts	<a href="https://openai.com/chatgpt">https://openai.com/chatgpt</a>
<b>Writecream</b>	Quick, free generation of text-based personas that can be refined and visually designed later	<a href="https://www.writecream.com/user-persona-generator">https://www.writecream.com/user-persona-generator</a>
<b>PersonaGenerator</b>	Quick, no-cost creation of ready-to-use text personas for early design, marketing, or product planning	<a href="https://www.personagenerator.com/">https://www.personagenerator.com/</a>
<b>Creately</b>	Team-based creation of visually designed personas using drag-and-drop templates for workshops, presentations, and strategy sessions	<a href="https://creately.com/lp/user-persona-creator">https://creately.com/lp/user-persona-creator</a>

Table 2 summarises the AI tools analysed for creating UX personas, outlining their primary strengths and providing direct access links. The full analysis available in Appendix VIII. These tools automate and enhance persona creation, ranging from quick, no-cost generators for producing initial text-based personas to more advanced platforms that support visual design, iterative refinement, and team collaboration.

LLMs were evaluated for their ability to create personas, and as a result, the AI tool - *AI User Persona Generator* by Bot Memo<sup>3</sup> was selected.

**User scenario creation:** the synthetic data collection initially started with the overview of the existing scenario creation tools. A full record of the activities undertaken to identify relevant applications can be found in Appendix IX.

---

<sup>3</sup> <https://botmemo.com/ai-user-persona-generator/>

Table 3. Analyzed AI tools for creating UX scenarios.

<b>Tool's name</b>	<b>Best for</b>	<b>Link</b>
<b>ChatGPT (OpenAI)</b>	Quick, conversational scenario ideation from persona inputs. Requires human refinement for accuracy	<a href="https://openai.com/chatgpt">https://openai.com/chatgpt</a>
<b>GenerateStory</b>	Rapid creation of varied, text-based scenarios that can be adapted for narratives, games, or learning activities	<a href="https://generatestory.io/scenario-generator">https://generatestory.io/scenario-generator</a>
<b>LeanScope</b>	Creating detailed, research-backed user scenario maps to inform design decisions and prioritize enhancements	<a href="https://www.leanscope.ai/en/user-scenarios">https://www.leanscope.ai/en/user-scenarios</a>

Table 3 presents the tools identified and evaluated for their potential to support the creation of user scenarios in UX research. The full evaluation, including detailed functionality comparisons, is provided in Appendix X. The review found that the number of tools specifically designed for scenario creation is limited. In contrast, there is a far wider range of tools available for related outputs such as customer journey maps and user stories.

In the real-world data process, user scenarios were developed as an integral part of each persona. These scenarios illustrated typical interaction steps, and goals that users encounter while using an online food delivery service. To ensure a consistent approach when working with synthetic data, a similar structure was required for generating comparable scenarios.

The platforms identified during the initial exploration had limited flexibility and often relied on predefined structures that did not allow for detailed customization of user context or tasks. As a result, they were not suitable for generating narratives aligned with the persona-driven format used in the real-world data scenarios.

Given these constraints, ChatGPT was selected to generate synthetic scenarios. The model allows users to define rich context and specific user characteristics, making it possible to prompt for realistic and tailored interaction flows. Prompts were crafted to generate not only personas, but also add scenarios, which helped the model generate scenarios that mirrored those created from interview-based data. A typical prompt example:

*"You are a product manager. Develop a user persona based on interview answers providing the name of the persona, the quote of such persona, demographic facts and description of such persona. Include main frustrations, goals, needs and expectations. Also add a section describing a person's typical day (scenario), their tasks using the product, and how they would interact with it. Generate the onepager picture with all this information divided in sections".*

## **4.2 Data Analysis**

This section explains how the collected datasets were processed, coded, and compared to address the research questions. It outlines the preparation steps taken to ensure consistency between real-world and synthetic interviews, the thematic analysis approach applied in ATLAS.ti, and the structured criteria used to evaluate personas and scenarios. Together, these methods establish a clear and replicable framework for examining similarities, differences, and potential complementarities between real-world and synthetically generated UX research outputs.

### **4.2.1 Data Preparation**

Before conducting the formal analysis, the real-world and synthetic datasets were reviewed and prepared separately to ensure consistency and clarity. Real-world interview transcripts were cleaned, manually corrected for clarity and accuracy where necessary, and anonymized. Synthetic data generated by ChatGPT underwent the same preparation process, including review and manual corrections.

To support side-by-side comparison, each data type was organized into a structured table (see Appendix IV), with responses grouped under the same interview questions. Real-world and synthetic responses were deliberately placed in separate sections of the table to allow for direct comparison on each topic. The relevant codes were assigned to each interview participant as follows:

- **H#** – a real-world interview and its participant number
- **LLM#** – a synthetic interview generated by the LLM and its corresponding number

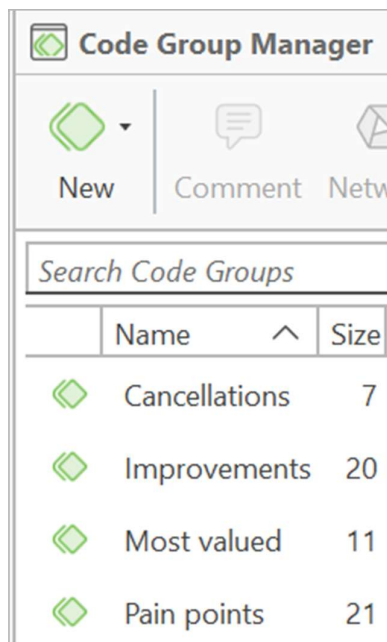
This preparation enabled a fair and transparent evaluation of how each data type addressed key research areas, such as order cancellations, suggested improvements, pain points, and most valued features. Maintaining this separation made it possible to trace differences and

similarities in user insights across the two sources. The both datasets were then imported into ATLAS.ti for thematic analysis, as described in the following subsection.

### 4.2.2 Thematic Analysis

The transcripts from both the real-world interviews and LLM-generated interviews were analyzed in ATLAS.ti using a reflexive thematic analysis approach [46]. Reflexive thematic analysis, as outlined by Braun and Clarke, is a flexible, iterative method for identifying and interpreting patterns of meaning in qualitative data. It acknowledges the active role of the researcher in generating codes and themes, and is well suited to exploring nuanced experiences - making it an appropriate choice for examining differences between real-world and synthetic user data.

Each dataset was coded separately to preserve distinctions during the initial coding phase. Recurring patterns were identified based on both the frequency of specific issues and their relevance to the user experience with online food delivery services. Based on these patterns, the resulting codes were grouped into four categories reflecting the main themes for the analysis (Figure 2).



Code Group Manager		
New	Comment	Netw
Search Code Groups		
	Name ^	Size
◇	Cancellations	7
◇	Improvements	20
◇	Most valued	11
◇	Pain points	21

Figure 2. Four main themes from interview data identified through thematic analysis in ATLAS.ti.

This grouping enabled a more targeted comparison of the datasets, making it easier to identify differences and similarities. Each category was assigned a distinct color in ATLAS.ti to support visual clarity during analysis (Figure 3).

Name	
◇ Cancellations	Orange
◇ Improvements	Green
◇ Most valued	Blue
◇ Pain points	Red

Figure 3. Color coding used in ATLAS.ti to analyze interview data.

The content of each category was then examined to identify commonalities and divergences between the real-world and synthetic data. In addition to examining the overall distribution of codes across both datasets, the density of coding within each category and the presence of unique codes were also analyzed.

### 4.2.3 Data Comparison

To compare the outputs from the real-world and synthetic datasets, an in-depth analysis of interview data was carried out alongside a structured, criteria-based review of personas and scenarios. This approach combined qualitative depth with side-by-side visual comparison, ensuring that findings were both nuanced and systematically derived.

**Interview data** from both real-world and synthetic sources was merged into a single table (see Appendix IV) and analysed in two complementary ways:

1. **Respondent-by-respondent review:** Each participant's answers across all questions were examined to identify tone, recurring frustrations, repeated themes, emotional shifts, and unexpected insights, creating a fuller picture of each respondent as an individual.
2. **Question-by-question comparison:** The perspective was then reversed, comparing how different respondents answered the same question. This revealed fresh angles, contrasts, and subtle nuances that might remain hidden in a purely respondent-focused review.

**Personas and scenarios** were assessed on four key dimensions:

- **Overall appearance:** As personas and scenarios are key artefacts in product design and guide important development decisions, it was essential to assess whether the outputs were well-structured, easy to interpret, grammatically correct and retained all critical information.
- **Demographic accuracy:** Since demographic details were predefined in both tools, these were checked to ensure they were correctly preserved and not altered during persona or scenario generation.
- **Emotional nuances:** Emotional tone and nuance were analyzed separately, as they are critical for understanding user needs. Loss of these elements could lead to misinterpretation of needs and, ultimately, products that fail to meet user expectations.
- **Scenario and task realism:** Scenarios were reviewed to determine whether the tasks and flows reflected realistic user interactions with the online food delivery service or whether they introduced unrealistic or impractical steps that could not be implemented in a real-world context.

By combining these perspectives, the methodology captured both the qualitative richness of the interviews and the practical usability of the design artefacts in a consistent, comparable way. This structured approach provides a solid foundation for the comparative analysis in Chapter 5, helping to ensure that any differences observed reflect the nature of the data itself rather than variations in the research process.

## 5. Findings

This chapter presents the results of the comparative analysis between the real-world and synthetic datasets. It is organised around the four key themes identified during coding - order cancellations, suggested improvements, pain points, and valued features - and extends to the evaluation of personas and scenarios generated from each dataset. The aim is to provide a clear, side-by-side account of where synthetic data aligns with, diverges from, or supplements the insights drawn from real-world participants.

Section 5.1 discusses insights from real-world interviews, outlining recurring themes such as order cancellations, suggested improvements, pain points, and valued features, and how these informed manually created personas and scenarios. Section 5.2 examines synthetic data generated with LLM (ChatGPT), applying the same thematic structure and evaluating LLM-generated personas and scenarios. Section 5.3 compares both datasets side by side, identifying thematic overlaps, divergences, and differences in emotional nuance and contextual detail.

### 5.1 Real-World Data Insights

Thematic analysis of five semi-structured interviews with real-world participants revealed several recurring user needs and frustrations:

- **Order cancellations:** Four out of five respondents had issues with orders cancellations - either by themselves or by the restaurant- and due to delays or unclear updates (see Appendix IV respondents H2, H3, H4, H5).
- **Suggested improvements:** Parents requested more child-friendly options, nutritional information, and packaging for hot food. Suggestions also included clearer labeling for allergens or spice levels. Each responded provided original suggestions on the improvements.
- **Pain points:** Although most users found the app “easy enough,” some (see Appendix IV respondents H1, H4) reported that complex menus and excessive categories were overwhelming.
- **Most valued features:** Users appreciated the ability to schedule deliveries and refine search to order quickly (one or both of these features were mentioned by 80% of respondents).

From these responses, a detailed persona and a corresponding user scenario were manually developed based on the real-world interview data (see Figure 4).

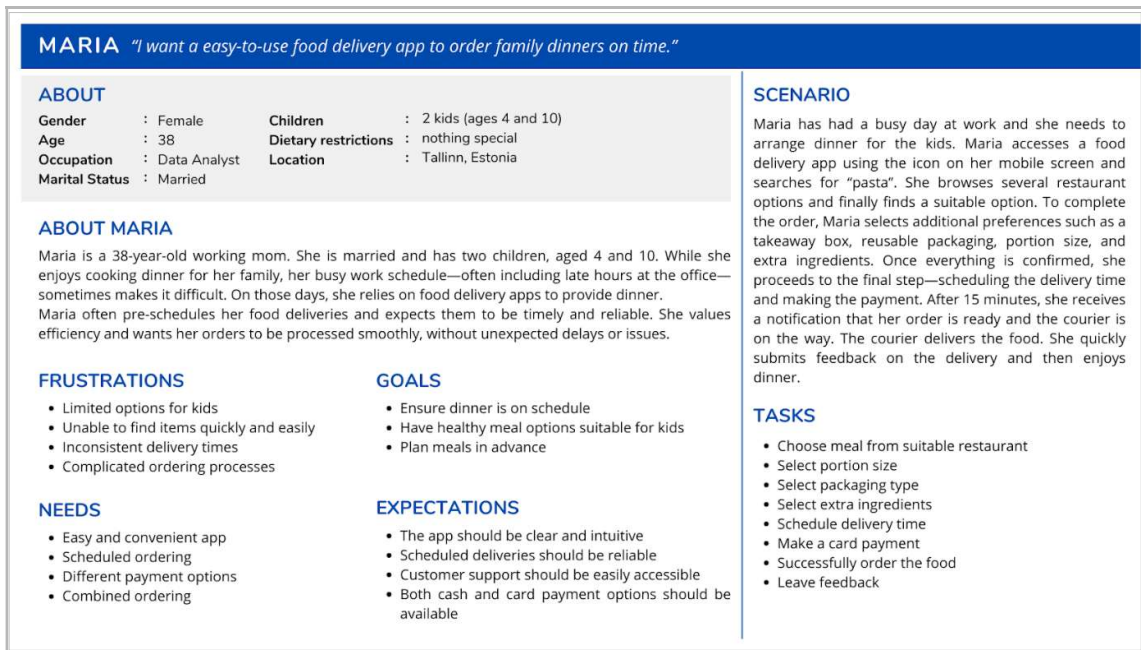


Figure 4. Persona and scenario generated based on real-world data.

The persona reflects the demographic background, frustrations, goals, needs and expectations commonly expressed by participants, while the scenario illustrates a typical task flow.

The following section presents a parallel analysis based on synthetic data and explores how LLMs performed in simulating user interviews and generating comparable personas and scenarios.

## 5.2 Synthetic Data Insights

LLM (ChatGPT) was prompted to simulate responses to the same interview questions.

The LLM-generated interviews were coherent and on-topic, providing notable insights on the main themes:

- **Order cancellations:** 100% of LLM responders have cancelled the orders. Most synthetic users described abandoning orders due to slow or confusing online food delivery service functionality, or perceived delivery uncertainty.
- **Suggested improvements:** 3 out of 5 LLM interviews mentioned that customer support service should be improved. Also 100% mentioned better filtering options to optimize search.
- **Pain points:** Challenges around filtering the items, unclear modification flows, and payments issues were common across the LLM responses. 100% of LLM respondents

criticized complex navigation and ineffective search filters, particularly for dietary preferences.

- **Most valued features:** 90% of the respondents consistently mentioned the value of features like real-time order tracking or estimated delivery times and saved favorites.

From these responses, detailed persona were derived using an AI tool *AI User Persona Generator* by Bot Memo (see Figure 5).

**Ideal User Persona: "Busy Mom in Tallinn"**

**Demographics:**

- Age: 36-38 years old
- Gender: Female
- Marital Status: Married
- Family: Has children
- Location: Tallinn

**Psychographics:**

- Values family time and convenience.
- Health-conscious but often pressed for time.
- Enjoys a variety of cuisines but may prioritize kid-friendly options.
- Likely to be tech-savvy, using smartphones for shopping and meal planning.

**Behaviors:**

- Frequently orders food online, particularly during busy weekdays or weekends.
- Engages with social media platforms for recommendations and reviews.
- Prefers services that offer quick delivery and easy-to-navigate apps.
- Likely to subscribe to newsletters or loyalty programs for discounts and promotions.

**Pain Points:**

- Struggles to find time to cook healthy meals amid family responsibilities.
- May feel overwhelmed by meal planning and grocery shopping.
- Concerned about the quality and nutritional value of food delivered.
- Seeks reliable services that cater to family needs, including dietary restrictions.

**Challenges:**

- Managing a busy household while balancing work and family life.
- Finding meals that satisfy both adult tastes and children's preferences.
- Navigating the plethora of food delivery options available in the market.
- Ensuring food is delivered on time and meets quality expectations.

**Preferred Marketing Channels:**

- Social media (Facebook, Instagram) for visual appeal and community engagement.
- Email marketing for personalized offers and reminders.
- Influencer partnerships that resonate with parenting and lifestyle themes.
- Local online forums or parenting groups for word-of-mouth recommendations.

Figure 5. Persona generated using an AI tool based on LLM interview answers, .txt format.

Since the AI tool used for generating personas did not include an option to create or integrate a user scenario, an alternative approach was required to complete the synthetic output in line with the format used for the real-world data. To maintain consistency - where scenarios were developed as part of each persona - ChatGPT was selected to generate the synthetic scenarios. ChatGPT was chosen for its ability to support precise prompt construction, allowing the inclusion of detailed user background, goals, and daily activities. This level of customization enabled the creation of realistic scenarios that mirrored the narrative structure and interaction flow observed in the real-world data.

The prompt was designed to instruct the model not only to generate a persona but also to include scenario descriptions linked to specific tasks and user needs (see Section 4.1.2 for a typical prompt example). From these outputs, a complete synthetic persona and scenario were produced; however, for the purposes of this research, the focus remained on the scenario component (see Figure 6).

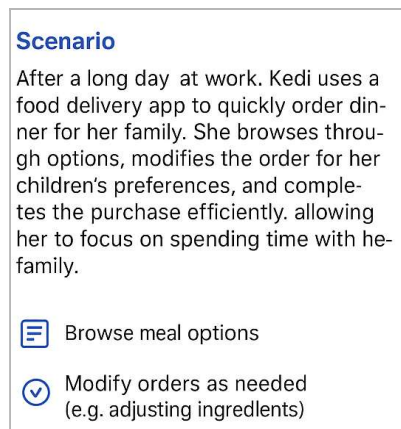


Figure 6. Scenario generated using ChatGPT based on synthetic data.

The figure presents both a description of the typical interaction steps and goals that users encounter while using an online food delivery service.

The following section compares the real-world and synthetic datasets side by side, highlighting key similarities as well as notable differences in the insights generated across interviews, personas, and user scenarios.

### 5.3 Comparative Analysis

This section presents a structured comparison of the real-world and synthetic datasets generated through user interviews, personas, and scenarios. The aim is to assess how closely LLM outputs

align with data derived from traditional UX research practices. By evaluating the content side by side, this chapter explores the strengths and limitations of LLM-generated insights and their potential role in UX research.

Understanding how synthetic data compares to real-world input is important for UX researchers considering the integration of AI tools into their workflows. This analysis helps determine whether LLM-generated content can supplement or even replace traditional methods in early-stage UX research, and under what conditions such integration may be most effective.

### 5.3.1 Interview Results Comparison

To evaluate how closely the LLM-generated responses aligned with real-world participant insights, a side-by-side comparison of interview results was conducted. Each dataset - real-world and synthetic - was analysed against four recurring criteria that emerged during the thematic analysis, as shown in Table 4 below.

Table 4. Key differences between real-world and LLM-generated interviews.

Theme	Real-world interviews	LLM-generated interviews
<b>Order cancellations</b>	Commonly due to app/restaurant errors	All respondents cancelled. Mainly due to app slowness or unclear navigation
<b>Suggested improvements</b>	Different improvements requested	Often mentioned better customer service
<b>Pain points</b>	No big issues, some minor comments	Lack of clarity in order modification, poor search, payments issues
<b>Most valued features</b>	Scheduling, search by meal type	Reordering, tracking and scheduling

Table 4 summarises the main similarities and differences across these areas. While certain themes, such as scheduling and navigation, appeared in both datasets, the nature and frequency of the concerns often differed. Synthetic responses tended to emphasise issues more strongly and consistently across participants, whereas real-world responses showed greater variation in tone and detail. The table offers a clear snapshot of these patterns, setting up the deeper analysis that follows on how consistent each data type is for UX research.

Figure 7 below shows the overlapping codes that emerged in both datasets as a result of the thematic analysis.

	Real-data Gr=40; GS=5		Synthetic data Gr=105; GS=5	
	Absolute	Row-relative	Absolute	Row-relative
○ support Gr=8	1	12,50%	7	87,50%
○ search options Gr=15	2	13,33%	13	86,67%
○ complicated_search_pain Gr=7	1	14,29%	6	85,71%
○ food_labels_improvement Gr=7	1	14,29%	6	85,71%
○ scheduled_delivery Gr=10	2	20,00%	8	80,00%
○ payment_options Gr=4	1	25,00%	3	75,00%
○ click_fatigue Gr=11	5	45,45%	6	54,54%
○ navigation_not_detailed_pain Gr=2	1	50,00%	1	50,00%
○ nutrient_information_improvement Gr=2	1	50,00%	1	50,00%
○ filtering_search_useful Gr=3	2	66,67%	1	33,33%
○ long_delivery_cancel Gr=3	2	66,67%	1	33,33%
<b>Abbreviations</b>				
Gr: Groundedness of codes (number of quotations coded by a code)				
GS: Number of documents in a document group				

Figure 7. Overlapping codes in real-world and synthetic datasets.

Despite some variation in how frequently specific themes appeared in each dataset, Figure 8 highlights clear areas of thematic overlap, suggesting that both real-world and synthetic interviews captured similar high-level concerns. These included topics such as search functionality, repetitive clicking (often referred to as "click fatigue"), and scheduled delivery preferences. At the same time, certain issues were more prominent in one dataset than the other, with some concerns appearing consistently in every interview of that group.

Figure 8 below illustrates the code distribution by individual interview. This visualization offers insight into the thematic density and coverage within each dataset and helps to identify where patterns converge or diverge between real-world and LLM-generated responses.

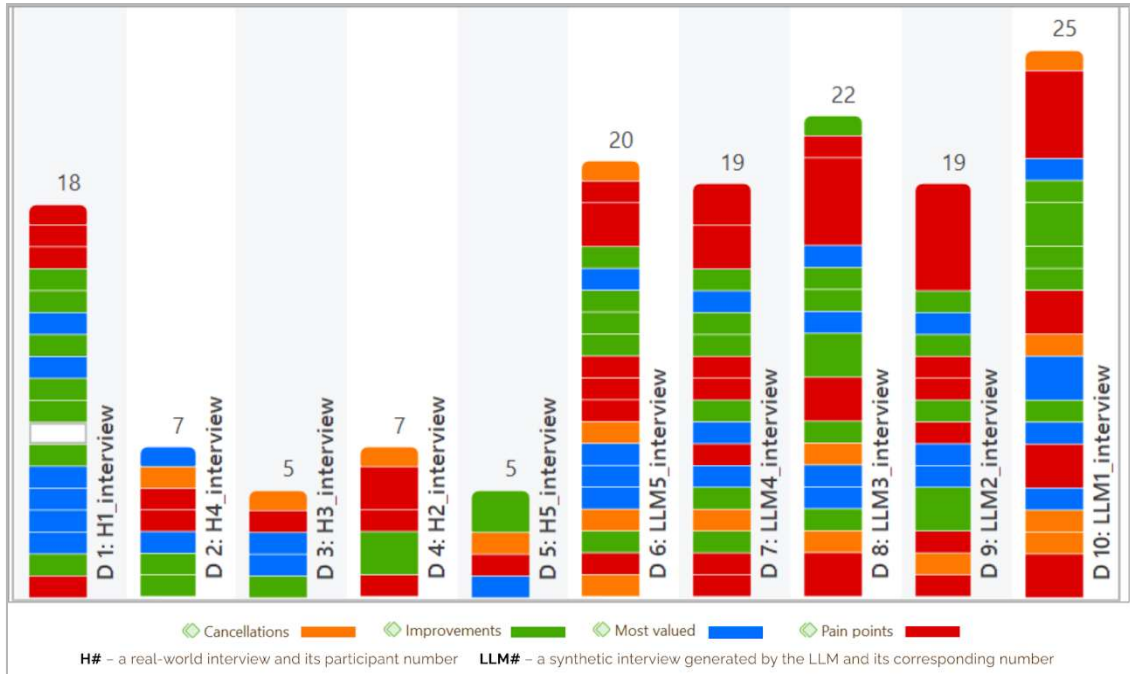


Figure 8. Total count of codes and distribution by real-world and synthetic datasets.

From this figure, several key patterns emerge. The synthetic dataset produced a greater number of codes overall, while the real-world dataset yielded fewer. Despite this difference in volume, the proportion of positive themes (blue and green codes) to negative themes (red and orange codes) remained almost the same across both datasets. This indicates that, although the LLM-generated interviews generated more coded segments, they did not disproportionately emphasise one sentiment over the other compared to real-world interviews.

In addition, LLM-generated responses introduced novel topics (see Figure 9) - loyalty systems or point tracking (code colored in purple), or dietary restrictions as gluten-free or vegan (code marked in yellow) - which were not raised at all in the real-world user interviews.

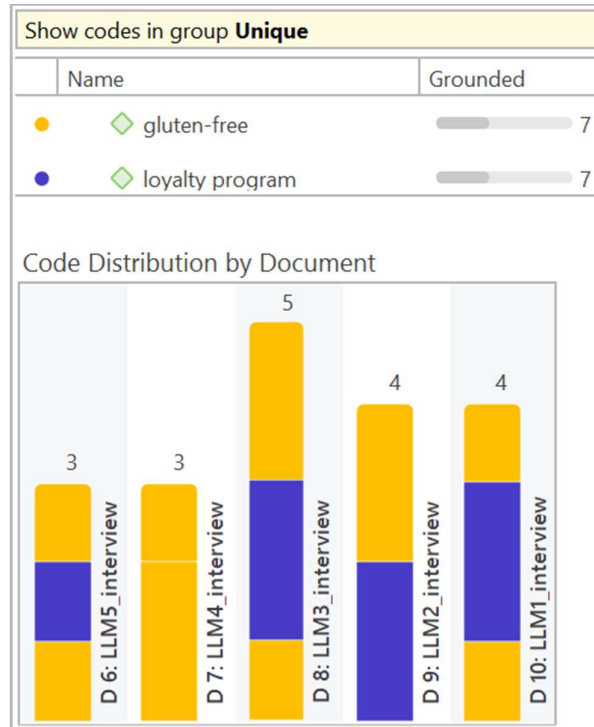


Figure 9. Unique codes identified in LLM-generated interviews.

These findings reflect the LLM's ability to surface broader or less commonly mentioned user concerns, possibly influenced by its training on a diverse range of sources.

Figure 10 below provides a comparison of the most frequently occurring themes - those mentioned more than five times - across both datasets. This analysis focuses on identifying shared concerns, as well as the density and coverage of user insights in each group.

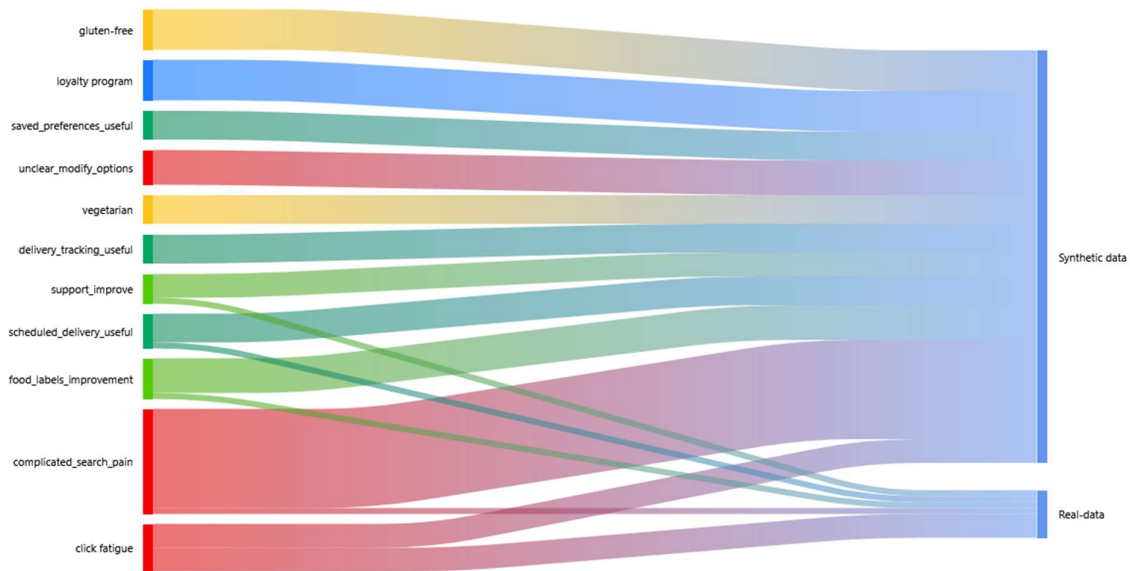


Figure 10. Sankey Diagram, most common themes from interviews.

As the figure illustrates, both datasets include recurring mentions of common topics, such as navigation issues, scheduling, and improvement requests, indicating some overlap in the types of insights captured. However, responses generated by the LLM were more verbose, leading to a higher total number of codes and broader thematic coverage.

This difference is particularly visible in the distribution of themes: the real-world dataset contains fewer repeated themes and tends to be more concise, while the synthetic data includes a wider range of issues, with some themes appearing multiple times across interviews. The real-world dataset covers less than 50% of the themes covered by LLMs showing more unique answers being less repetitive.

A detailed comparison of individual interview responses (see Appendix IV) revealed that LLM-generated answers were generally more structured and complete. However, these responses often lacked the inconsistencies and spontaneous comments that are typical in real-world conversations. For example, one real-world interview participant (H1) mentioned canceling an order by phone despite disliking phone calls - an expression of contradiction and personal nuance that was not replicated in the LLM responses:

*“I had to call the restaurant to make changes because online support in the chat didn’t handle it. I’d like to have the option to resolve issues without calling the restaurant - I prefer not to talk to people”.*

Another real-world participant (H2) clearly stated that their main concern with online food delivery services was the limited range of restaurants available in their location. This frustration was repeatedly expressed throughout the interview and appeared to be a particularly upsetting factor for this user.

*“Ordering and payments are simple. However, there are issues with ordering from certain restaurants that don’t deliver to my location.”* (Question 2)

*“There could be more choices available for my location.”* (Question 6)

*“I would like to have McDonald’s options available for Nõmme.”* (Question 7)

The LLM-generated answers tended to be more general, frequently covering a broader range of features, but without reflecting edge cases or emotional undertones found in real-world interviews, such as frustration over repeated meal cancellations (H3):

*“I never canceled an order, but I had orders canceled on me. One time, no one responded for over 10 minutes, and then I got a message saying the restaurant wasn’t responding, so the order was canceled. Another time, after placing an order, the restaurant called and suggested a replacement.”*

or concerns related to children’s dietary needs (H1):

*“As a mom, I find there aren’t enough options for kids - there should be dedicated choices. I’d also like mandatory labelling for kid-friendly meals (e.g., non-spicy).”*

Noticeable differences also emerged regarding order cancellations. While most real-world interview participants reported rarely canceling orders, the LLM-generated responses consistently assumed that every user had canceled at least once. The reasons given by the LLM were also more generic and repetitive, typically citing slow app performance or confusing interfaces. In contrast, real-world participants provided more personal accounts, such as orders being canceled unexpectedly by the app or restaurant - a situation described as frustrating and inconvenient.

Figure 11 below illustrates these contrasts, highlighting how the cancellation narratives differ between real-world and synthetic interviews.

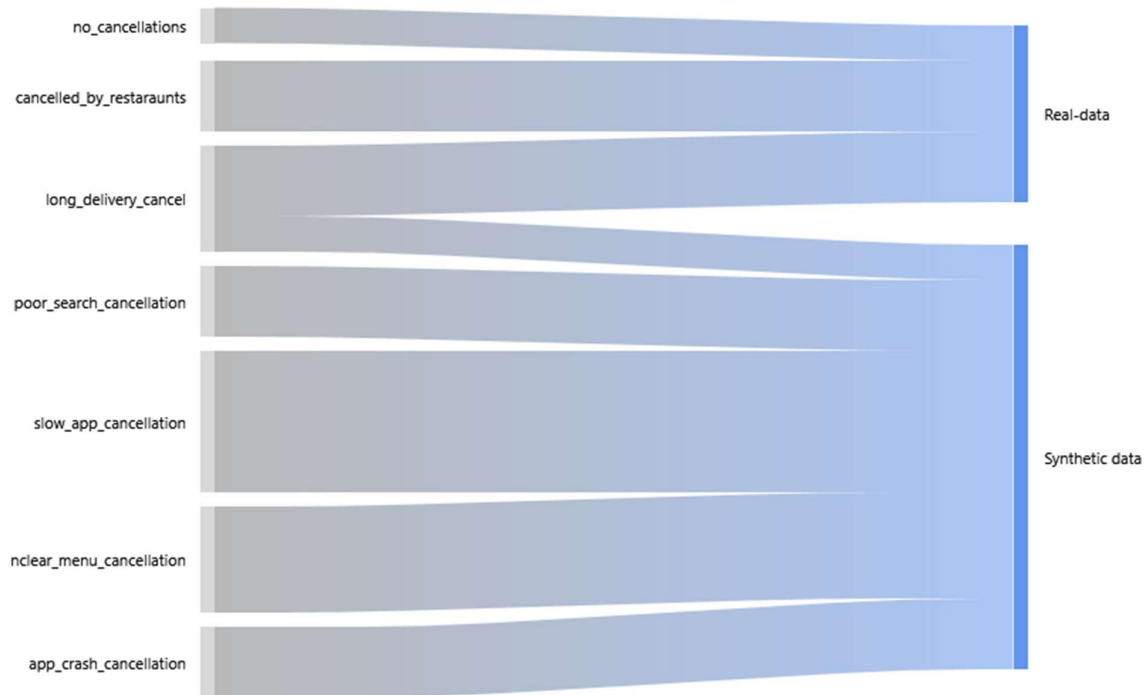


Figure 11. Sankey Diagram, cancellation patterns.

These results indicate that while LLMs tend to generalize from broader usage patterns and user expectations, they can sometimes introduce ideas that real-world participants may overlook or deprioritize. Rather than saying no new use cases emerged, it is more accurate to state that LLMs generated plausible but non-user-specific enhancements - such as loyalty programs - that may reflect wider product trends or inferred design norms rather than lived experience.

Emotional nuance and spontaneity were limited in the LLM-generated responses, which often repeated similar content with varied wording. In contrast, the responses from real-world interview participants demonstrated more emotional depth, personal detail, and spontaneity. Some participants initially said they had no major concerns with online food delivery service but later recalled specific frustrations or suggested improvements as the conversation progressed. A couple of participants even followed up after the interview with further input. One participant contacted the researcher the next day with an idea for a new feature (H1), while another shared a screenshot to illustrate a previously discussed app feature (H3).

In summary:

- LLMs excelled at producing structured, readable, and “plausible” responses.
- Real-world interview participants introduced unpredictability, personal history, and social nuance.
- While synthetic data were useful, they lacked the surprise factor and subtle complexity that emerged from real-world interviews.

The results support the idea that LLM-generated data can complement, but not fully replace, real-world UX research - particularly in early stages of product and service design.

### 5.3.2 Persona and Scenario Comparison

To compare the outputs from the real-world and synthetic datasets a structured, criteria-based review for the personas and scenarios.

The comparison was conducted through a side-by-side visual analysis of the outputs received from both datasets as shown on Figure 12:

**MARIA** *"I want a easy-to-use food delivery app to order family dinners on time."*

**ABOUT**

Gender	: Female	Children	: 2 kids (ages 4 and 10)
Age	: 38	Dietary restrictions	: nothing special
Occupation	: Data Analyst	Location	: Tallinn, Estonia
Marital Status	: Married		

**ABOUT MARIA**

Maria is a 38-year-old working mom. She is married and has two children, aged 4 and 10. While she enjoys cooking dinner for her family, her busy work schedule—often including late hours at the office—sometimes makes it difficult. On those days, she relies on food delivery apps to provide dinner. Maria often pre-schedules her food deliveries and expects them to be timely and reliable. She values efficiency and wants her orders to be processed smoothly, without unexpected delays or issues.

**FRUSTRATIONS**

- Limited options for kids
- Unable to find items quickly and easily
- Inconsistent delivery times
- Complicated ordering processes

**NEEDS**

- Easy and convenient app
- Scheduled ordering
- Different payment options
- Combined ordering

**GOALS**

- Ensure dinner is on schedule
- Have healthy meal options suitable for kids
- Plan meals in advance

**EXPECTATIONS**

- The app should be clear and intuitive
- Scheduled deliveries should be reliable
- Customer support should be easily accessible
- Both cash and card payment options should be available

**Ideal User Persona: "Busy Mom in Tallinn"**

**Demographics:**

- Age: 36-38 years old
- Gender: Female
- Marital Status: Married
- Family: Has children
- Location: Tallinn

**Psychographics:**

- Values family time and convenience.
- Health-conscious but often pressed for time.
- Enjoys a variety of cuisines but may prioritize kid-friendly options.
- Likely to be tech-savvy, using smartphones for shopping and meal planning.

**Behaviors:**

- Frequently orders food online, particularly during busy weekdays or weekends.
- Engages with social media platforms for recommendations and reviews.
- Prefers services that offer quick delivery and easy-to-navigate apps.
- Likely to subscribe to newsletters or loyalty programs for discounts and promotions.

**Pain Points:**

- Struggles to find time to cook healthy meals amid family responsibilities.
- May feel overwhelmed by meal planning and grocery shopping.
- Concerned about the quality and nutritional value of food delivered.
- Seeks reliable services that cater to family needs, including dietary restrictions.

**Challenges:**

- Managing a busy household while balancing work and family life.
- Finding meals that satisfy both adult tastes and children's preferences.
- Navigating the plethora of food delivery options available in the market.
- Ensuring food is delivered on time and meets quality expectations.

**Preferred Marketing Channels:**

- Social media (Facebook, Instagram) for visual appeal and community engagement.
- Email marketing for personalized offers and reminders.
- Influencer partnerships that resonate with parenting and lifestyle themes.
- Local online forums or parenting groups for word-of-mouth recommendations.

Figure 12. Personas generated from both datasets. Left real-world, right LLM-generated.

Persona generation showed relatively few differences, with results closely matching those from real-world data. Generating personas using AI tools largely depends on the specific tool selected and the quality of information obtained during the interviews.

The visual comparison of the results did not reveal any major errors. Minor grammatical issues and formatting inconsistencies should be corrected manually to achieve the best presentation quality. Both personas included demographic sections; however, this was expected, as demographic details were provided as input to the tool. The LLM-generated persona also mentioned factors such as convenience, ordering on busy days, the importance of kid-friendly options, and service reliability - information that was also reflected in the real-world persona.

Some differences were identified during the comparison:

- The real-world persona included four categories: frustrations, goals, needs, and expectations. In contrast, the LLM-generated persona featured five categories: psychographics, behaviours, pain points, challenges, and preferred marketing channels.
- The LLM-generated persona also did not include a persona quote, which is a valuable element as it conveys the user's perspective in their own words, adds emotional depth, and helps connect with the persona on a more human level.
- Similar to the interview results, the LLM-generated persona mentioned loyalty programs (see the "Behaviour" section), which were never mentioned in the real-world interviews and therefore were not reflected in the real-world persona.
- The LLM-generated persona also contained a "Preferred Marketing Channels" section; however, no relevant input was provided for this category, and related themes were not discussed in either the real-world or synthetic interviews.

Similarly to the persona comparison, the scenario comparison was conducted through a side-by-side visual analysis of the outputs from both datasets, as shown in Figure 13:



<p><b>SCENARIO</b></p> <p>Maria has had a busy day at work and she needs to arrange dinner for the kids. Maria accesses a food delivery app using the icon on her mobile screen and searches for “pasta”. She browses several restaurant options and finally finds a suitable option. To complete the order, Maria selects additional preferences such as a takeaway box, reusable packaging, portion size, and extra ingredients. Once everything is confirmed, she proceeds to the final step—scheduling the delivery time and making the payment. After 15 minutes, she receives a notification that her order is ready and the courier is on the way. The courier delivers the food. She quickly submits feedback on the delivery and then enjoys dinner.</p> <p><b>TASKS</b></p> <ul style="list-style-type: none"> <li>• Choose meal from suitable restaurant</li> <li>• Select portion size</li> <li>• Select packaging type</li> <li>• Select extra ingredients</li> <li>• Schedule delivery time</li> <li>• Make a card payment</li> <li>• Successfully order the food</li> <li>• Leave feedback</li> </ul>	<p><b>Scenario</b></p> <p>After a long day at work. Kedi uses a food delivery app to quickly order dinner for her family. She browses through options, modifies the order for her children's preferences, and completes the purchase efficiently, allowing her to focus on spending time with her family.</p> <ul style="list-style-type: none"> <li> Browse meal options</li> <li> Modify orders as needed (e.g. adjusting ingredients)</li> </ul>
--	---

Figure 13. Scenarios generated from both datasets. Left real-world, right LLM-generated.

Both scenarios provide a description of the typical flow and interactions with the app, followed by a list of tasks performed within it. However, the real-world scenario contains more detailed information in both sections.

For example, the typical flow description includes specific personal details such as “*searches for ‘pasta’*” or “*takeaway box*”, whereas the LLM-generated scenario is more general and lacks such nuances.

In addition, the LLM-generated scenario does not capture the full process, ending after the order is placed. In contrast, the real-world scenario also includes post-purchase stages, such as waiting time and food delivery, with details like “*after 15 minutes, she receives a notification.*”

A significant difference is also evident in the “Tasks” section. The real-world scenario lists eight tasks typically performed by the user, while the LLM-generated version contains only two, which resemble general user needs rather than concrete in-app steps.

Table 5 summarizes five key aspects used to compare personas and scenarios from the both - the real-world and synthetic - datasets.

Table 5. Key differences between real-world and LLM-generated personas and scenarios.

Aspect	Real-world data	Synthetic data
<b>Demographics</b>	Accurate, lived, and grounded in real context	Structured and plausible, inferred solely from prompt inputs
<b>Overall appearance</b>	Realistic and coherent in presentation	Minor formatting and grammatical inconsistencies; output requires manual refinement
<b>Emotional nuance</b>	Evident – includes frustration, hesitation, and occasional contradictions	Limited – attempts at emotional depth depend heavily on prompt specificity
<b>Scenario and tasks realism</b>	High – reflects daily routines, constraints, and behavioral variability	Low – generic, surface-level scenarios with minimal contextual detail

The comparison between real-world and synthetic personas reveals clear differences in emotional expression, detail, and originality. Personas developed from actual interviews tended to include more personal emotions, specific insights, and diverse suggestions that reflected real-world participants' experiences. These details often emerged naturally during the conversations, allowing the researcher to identify and shape the most relevant insights to fit the research goals. By contrast, synthetic personas generated using LLMs lack spontaneity and emotional subtlety. Their structure and tone are highly dependent on how precisely the prompts are written, with limited flexibility in narrative depth or behavior diversity.

The tool analysis also revealed that not all AI tools support scenario creation in the format defined for the purposes of this research, although many allow customization of the final output's design. When using tools - particularly ChatGPT - crafting precise prompts is essential to achieve the desired format and content quality. Previous research has shown that only half of the performance gains achieved after switching to a more advanced AI model come from the model itself, while the other half result from how effectively users adapt their prompts - that is, the written instructions guiding the AI - to take advantage of the new system [48]. This reinforces the methodological approach described in Section 4.1.2, where prompt design was treated as a critical factor in ensuring the LLM-generated outputs matched the structure, level of detail, and usability standards set for the real-world scenarios.

The findings reveal both overlaps and notable differences between the datasets. While synthetic outputs often mirrored the thematic structure of real-world responses, they lacked certain context-specific details and emotional nuances.

In summary:

- LLMs produced structured and visually coherent personas and scenarios that aligned with the general themes and flows found in real-world outputs.
- Real-world versions contained richer detail, emotional nuance, and context-specific elements, including post-purchase steps and concrete in-app actions.
- Synthetic outputs were more generic, with omissions such as persona quotes and fine-grained scenario steps, and occasionally introduced elements not grounded in the original data (e.g., loyalty programs).
- While useful for quickly generating plausible artefacts, LLM outputs lacked the depth, variability, and lived-experience detail essential for accurately representing user behaviour and needs.

The results suggest that LLM-generated personas and scenarios can support early-stage design work but should be refined and validated through real-user research before informing critical design decisions. These findings form the basis for the critical discussion in Chapter 6, which considers their implications for integrating LLMs into UX research workflows.

## 6. Discussion

This thesis set out to explore whether LLMs can meaningfully support UX research, focusing on their ability to simulate three core elements of UX workflows: user interviews, personas, and user scenarios. By generating both real-world and synthetic datasets and evaluating them systematically, the study assessed how effectively synthetic content can reproduce the kinds of insights typically derived from traditional qualitative methods.

Both datasets were assessed using clear criteria - emotional nuance, demographic accuracy, coherence, and the realism of user goals and tasks. Looking at them side by side made it easier to see where LLMs perform well in UX contexts and where they fall short, providing practical pointers on when and how these tools can be woven into professional workflows.

To address Research Question 1 - *“To what extent can LLMs be used to generate synthetic user research data such as interviews, personas, and user scenarios?”* - two sets of tools were examined: general-purpose LLMs and specialised UX-focused applications.

For interview generation, several capable options were identified, with LLMs like ChatGPT able to produce coherent, well-structured responses that followed a predefined interview guide. For personas, the chosen tool could create structured profiles with basic demographic and behavioural details, although these often required refinement to achieve realism and avoid overgeneralisation. In contrast, scenario generation revealed the most significant limitations: outputs included overly simplified, or impractical user actions, indicating challenges in modelling realistic, context-rich user flows. Overall, while the tools were technically capable of producing the required artefacts, their reliability and realism varied considerably by artefact type.

To address the Research Question 2 - *“How does synthetic data compare to real-world data for understanding user needs?”* - the study conducted a direct comparison between the two datasets. The analysis showed that synthetic data was consistently coherent, thematically broad, and structurally complete. However, it lacked much of the emotional depth, contextual specificity, and nuanced variability present in real-world responses.

These differences were especially evident in interview outputs. While real-world participants often provided emotionally charged accounts, personal stories, and context-specific

frustrations, synthetic responses tended to be more generic, repeating similar pain points across interviews and limiting the diversity of insights. For instance, while all synthetic participants reported cancelling orders, real-world participants rarely did so. Moreover, real-world participants offered emotionally charged examples, such as frustration with restaurant practices or reflective comments after the interview. Such details added realism and depth to the personas and scenarios developed from their input.

LLMs also tended to generalise user behaviour, increasing consistency but reducing behavioural diversity and limiting the emergence of unexpected use cases. Certain suggestions, such as the inclusion of loyalty programs, appeared to reflect market norms rather than the lived experiences of the studied demographic. As noted by Nielsen Norman Group [45], synthetic insights are best regarded as structured suggestions rather than definitive evidence and should be validated with real-world participants before informing design priorities.

Personas generated using the selected tool generally displayed a polished structure and clear formatting, meeting expectations for overall appearance. However, the content diverged from that of real-world personas. Several sections repeatedly included loyalty programs despite this never being mentioned by real-world participants. Additionally, no personal quotes were incorporated, and the narratives lacked the emotional nuances and lived-experience details present in real-world personas, reducing their depth and authenticity. Such omissions and content deviations have implications for understanding user needs, as personas are often an important input for product and service design decisions.

The scenario comparison reinforced these patterns. While both the real-world and LLM-generated scenarios followed a similar ordering flow, the synthetic version lacked the fine-grained, context-specific details that make scenarios valuable.

Another notable difference emerged during the interview analysis: when real-world participants had nothing to answer, they left a question unanswered or explicitly stated they had no relevant experience. In contrast, the LLM always produced a response - often filling the gap with something that sounded plausible but was not based on real-world experience. While this behaviour can make the dataset appear more complete, it risks introducing “false positives”: problems or needs that seem genuine in the synthetic data but do not actually exist. If taken at face value, such fabricated details could lead a UX research team in the wrong direction, diverting time and resources away from real-world user priorities.

These findings show that while LLMs can produce structured, plausible scenarios and personas for early-stage exploration, their outputs often require careful prompting, targeted refinement, and human interpretation to achieve the depth, emotional nuance, and contextual accuracy needed for reliable UX insights. They can accelerate initial research, extend the reach of UX teams working under time or resource constraints, and even surface plausible ideas that real-world participants might not mention - but these ideas may reflect market trends or inferred norms rather than lived experiences. Understanding these strengths and limitations is essential for integrating LLMs effectively into professional UX workflows.

## **6.1 Recommendations**

Based on the findings of this research, LLMs show the greatest potential in specific areas of UX research. They are particularly effective for:

- Quickly generating draft personas and user scenarios
- Brainstorming potential pain points, feature ideas, or user goals
- Supporting early-stage research when access to real-world participants is limited or delayed

The broad availability and ease of use of AI tools make it possible to quickly simulate an initial persona or scenario based on existing information. Within minutes, a preliminary version of these artefacts can be generated and used as a starting point. While these outputs require thorough review and modification to ensure alignment with real-world insights, they provide a solid foundation for deeper research phases.

Draft personas and scenarios can also support early decision-making, such as identifying which types of participants should be recruited for upcoming studies. By revealing preliminary assumptions about user demographics, behaviours, and needs, they help UX teams prioritise recruitment criteria before engaging with actual users. Similarly, using LLMs to brainstorm potential pain points, features, and user goals can help structure a user research plan. This ensures that all critical topics are addressed in interviews and that the team enters the research phase with a well-defined scope.

In addition, synthetic outputs can act as a springboard for internal discussions and workshops. They offer generalised user information that can spark ideas, align the team's understanding, and accelerate the early design process. This makes it possible to begin UX research at minimal cost and effort while still having useful, structured input at the initial stage.

These strengths align with findings by Hewing and Leinhos, who describe LLMs as highly effective tools for quickly producing relevant UX content - especially when guided by structured prompting methods. By framing prompts with clear roles, context, and structure, researchers can significantly improve both the quality and reliability of LLM-generated material [43].

The findings of this study indicate that LLM-generated outputs are most valuable when used as a supplementary resource rather than a standalone replacement for real-world user research. In practice, synthetic artefacts proved effective for sparking ideas, shaping early design concepts, and identifying preliminary directions for exploration, but they required validation through engagement with actual users before informing final design decisions. This aligns with observations by Kieslinger and Wippler [44], who similarly emphasise that LLM-generated content can enhance, but not replace, the iterative and user-driven nature of design processes.

Because LLMs may generate plausible but unfounded answers when no real experience exists, there is a risk of introducing “false positives” that can misdirect research if accepted at face value. To avoid this, human oversight is essential for reviewing, validating, and contextualising synthetic outputs before they are used to inform product or service design decisions.

When projects involve emotionally sensitive topics or complex social contexts - such as designing for children, health services, or vulnerable populations - direct interaction with real users remains essential. While LLMs can provide early-stage insights, they cannot replicate the empathy, unpredictability, and personal perspective that real-world participants bring to UX research. Ultimately, LLMs should be viewed as complementary tools. They can streamline and accelerate certain aspects of the research process, but the interpretation, validation, and application of results must remain in the hands of human researchers. This ensures that UX decisions stay grounded in genuine human needs and experiences.

## **6.2 Limitations**

This study relied on a convenience sample consisting of five real-world interviews and five simulated interviews generated by a LLM. Consequently, the findings reflect the perspectives of a single, specific user group.

The research also focused exclusively on one type of digital service - an online food delivery service. While this narrow scope allowed for a controlled and focused comparison between real-world and synthetic data, it limits the generalisability of the results. Including a broader

range of user groups, demographic profiles, or service types could offer a more comprehensive understanding of where synthetic data is most effective and where it may fall short.

Another limitation is that the synthetic dataset was generated primarily using ChatGPT 4.0. Although this tool was well-suited to the study's objectives, different language models may yield varied outputs due to differences in their training data, architecture, and optimization methods.

## 7. Future Work

This study provides a starting point toward understanding how LLMs can contribute to UX research. As the field evolves rapidly, several avenues remain open for deeper exploration.

One promising direction is refining prompt engineering techniques to simulate a broader range of user experiences - particularly those involving emotional depth or less common scenarios. Future work could focus on generating more varied and lifelike responses that capture edge cases, such as frustration, confusion, or highly specific user needs.

Expanding this research beyond the online food delivery service context would also be valuable. Exploring additional domains could help clarify where LLM-generated insights are most useful and where direct engagement with real users remains essential.

Another potential direction is examining how LLMs can function as part of a collaborative UX workflow. Rather than replacing human researchers, AI could be integrated into early-stage activities - such as drafting interview guides, generating initial personas, or brainstorming feature concepts - while leaving interpretation, validation, and final decision-making firmly in human hands.

Although this study focused primarily on ChatGPT, future research could compare multiple LLMs - such as Claude, Gemini, or open-source models - performing the same tasks. Cross-model comparisons could help identify which systems excel at replicating user behaviour, emotional tone, or realistic task flows in different UX domains.

Given that new versions of LLMs are released frequently, longitudinal studies could track whether newer models offer more accurate, emotionally aware, or contextually nuanced outputs, and whether they reduce the need for manual refinement.

Finally, further research should examine the ethical considerations of using synthetic data in UX. This includes ensuring transparency about which insights originate from AI, avoiding the creation of stereotypical personas, and recognising when simulating certain user groups or topics may be inappropriate. Establishing clear guidelines for how LLM-generated personas and scenarios are applied - especially in sensitive contexts - will be critical for responsible adoption.

## 8. Conclusion

This thesis investigated the potential of LLMs to generate synthetic user data for UX research, using an online food delivery service as a case study. By comparing results from the five real-world interviews with five LLM-generated interviews using ChatGPT, the study examined whether synthetic outputs could provide insights similar to those produced through traditional qualitative methods.

The results show that LLMs can generate coherent, structured responses that align with many common user concerns, such as navigation difficulties, delivery timing, and order modification issues. These outputs were particularly useful for creating plausible personas and scenarios in early-stage design or ideation. Advantages included speed, consistency, and broad thematic coverage, making LLMs suitable as supportive tools within UX workflows.

However, the analysis also revealed significant limitations. LLM-generated responses lacked the emotional nuance, unpredictability, and lived experience present in the real-world interviews. While real-world participants shared personal stories, frustrations, and context-specific needs, LLM outputs were more generalised and occasionally repetitive. This indicates that while LLMs can assist UX research, they cannot replace direct user engagement, particularly in emotionally sensitive, high-risk, or highly contextual design projects.

The study followed a structured process: Chapter 2 introduced the theoretical and technical background; Chapter 3 reviewed existing studies on synthetic data in UX contexts; Chapter 4 detailed the methodology for collecting and analysing real-world and synthetic data; Chapter 5 presented the comparative findings; and Chapter 6 discussed their implications, limitations, and recommendations.

Overall, the findings suggest that LLMs should be positioned as complementary research tools. They can accelerate certain aspects of the process and extend the scope of early exploration when time, budget, or participant access is limited. Yet, their outputs require human validation and refinement to ensure that design decisions remain grounded in genuine human needs. As AI technology evolves, ongoing research will be essential to refine these tools, improve their contextual sensitivity, and define best practices for their integration into UX research.

## References

- [1] L. Floridi and M. Chiriatti, “GPT-3: Its nature, scope, limits, and consequences. Minds and Machines”, 2020. [Online]. Available: <https://doi.org/10.1007/s11023-020-09548-1>
- [2] M. Takaffoli, S. Li and V. Mäkelä, “Generative AI in User Experience Design and Research: How Do UX Practitioners, Teams, and Companies Use GenAI in Industry?”, 2024. [Online]. Available: <https://doi.org/10.1145/3643834.3660720>
- [3] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, ... and P. , “On the opportunities and risks of foundation models“, 2021. [Online]. Available: <https://doi.org/10.48550/arXiv.2108.07258>
- [4] L. Weidinger, J. Mellor, M. Rauh, C. Griffin, J. Uesato, P. S. Huang, et al., “Ethical and social risks of harm from language models”, 2021. [Online]. Available: <https://doi.org/10.48550/arXiv.2112.04359>
- [5] D. A. Norman, *The design of everyday things*, (revised ed.). MIT Press, 2013.
- [6] B. Shneiderman, “Human-centered artificial intelligence: Reliable, safe & trustworthy”, 2020. [Online]. Available: <https://doi.org/10.48550/arXiv.2002.04087>
- [7] Xipeng Qiu, Tianxiang Sun, Yige Xu, Yunfan Shao, Ning Dai, and Xuanjing Huang, “Pre-trained Models for Natural Language Processing: A Survey”, 2020. [Online]. Available: <https://doi.org/10.48550/arXiv.2003.08271>
- [8] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria, “Recent Trends in Deep Learning Based Natural Language Processing”, 2018. [Online]. Available: <https://doi.org/10.48550/arXiv.1708.02709>
- [9] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, et al., ”Language Models are Few-Shot Learners”, 2020. [Online]. Available: <https://doi.org/10.48550/arXiv.2005.14165>
- [10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin, “Attention Is All You Need”, 2017. [Online]. Available: <https://doi.org/10.48550/arXiv.1706.03762>

- [11] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurélien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. “LLaMA: Open and Efficient Foundation Language Models”, 2023. [Online]. Available: <https://doi.org/10.48550/arXiv.2302.13971>
- [12] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe, “Training Language Models to Follow Instructions with Human Feedback”, 2022. [Online]. Available: <https://doi.org/10.48550/arXiv.2203.02155>
- [13] Elizabeth Goodman, Mike Kuniavsky, and Andrea Moed, *Observing the User Experience: A Practitioner’s Guide to User Research*, 2nd edition. Elsevier, Waltham, MA, 2012.
- [14] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser, *Research Methods in Human–Computer Interaction*, 2nd edition. Morgan Kaufmann, Cambridge, MA, 2017.
- [15] Kim Goodwin, *Designing for the Digital Age: How to Create Human-Centered Products and Services*. Wiley, Indianapolis, IN, 2009.
- [16] Jenny Preece, Helen Sharp, and Yvonne Rogers, *Interaction Design: Beyond Human-Computer Interaction*, 3rd edition. Wiley, Chichester, UK, 2012.
- [17] Bella Martin and Bruce Hanington, *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. Rockport Publishers, Beverly, MA, 2012.
- [18] Jeff Sauro and James R. Lewis, *Quantifying the User Experience: Practical Statistics for User Research*, 2nd edition. Morgan Kaufmann, Cambridge, MA, 2016.
- [19] Andre Gonçalves, Priyadip Ray, Braden Soper, Jennifer Stevens, Linda Coyle, Ana Paula Sales, et al., “Generation and evaluation of synthetic patient data”, 2020. Accessed: Aug. 05, 2025. [Online]. Available: <https://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/s12874-020-00977-1>

- [20] Neha Patki, Roy Wedge, and Kalyan Veeramachaneni, “The Synthetic Data Vault”, 2016. Accessed: Aug. 05, 2025. [Online]. Available: <https://dai.lids.mit.edu/wp-content/uploads/2018/03/SDV.pdf>
- [21] Tao Ge, Xin Chan, Xiaoyang Wang, Dian Yu, Haitao Mi, and Dong Yu, “Scaling Synthetic Data Creation with 1,000,000,000 Personas”, 2025. [Online]. Available: <https://doi.org/10.48550/arXiv.2406.20094>
- [22] OpenAI, “GPT-4 Technical Report”, 2023 (revised 2024). [Online]. Available: <https://doi.org/10.48550/arXiv.2303.08774>
- [23] X. Zhang et al., “Auto-Generated Personas: Enhancing User-centered Design Practices among University Students”, 2024. [Online]. Available: <https://doi.org/10.1145/3613905.3651043>
- [24] A. Baki Kocaballi, “Conversational AI-Powered Design: ChatGPT as Designer, User, and Product”, 2023. [Online]. Available: <https://doi.org/10.48550/arXiv.2302.07406>
- [25] Ammar Ahmed and Ali Shariq Imran, “The Role of Large Language Models in UI/UX Design: A Systematic Literature Review”, 2025. [Online]. Available: <https://doi.org/10.48550/arXiv.2507.04469>
- [26] Jihong Zhang, Xinya Liang, Anqi Deng, Nicole Bonge, Lin Tan, Ling Zhang, and Nicole Zarrett, “Leveraging Interview-Informed LLMs to Model Survey Responses: Comparative Insights from AI-Generated and Human Data”, 2025. [Online]. Available: <https://doi.org/10.48550/arXiv.2505.21997>
- [27] Mohammadmehdi Attaei, Hyunmin Cheong, Daniele Grandi, Ye Wang, Nigel Morris, and Alexander Tessier, “Elictron: An LLM Agent-Based Simulation Framework for Design Requirements Elicitation”, 2024. [Online]. Available: <https://doi.org/10.48550/arXiv.2404.16045>
- [28] Fernando Ayach, Vitor Lameirão, Raul Leão, Jerfferson Felizardo, Rafael Sobrinho, Vanessa Borges, Patrícia Matsubara, and Awdren Fontão, “Generating Proto-Personas through Prompt Engineering: A Case Study on Efficiency, Effectiveness and Empathy”, 2025. [Online]. Available: <https://doi.org/10.48550/arXiv.2507.08594>

- [29] Jonas Eschrich and Sam Sterman, “A Framework for Discussing LLMs as Tools for Qualitative Analysis”, 2024. [Online]. Available: <https://doi.org/10.48550/arXiv.2407.11198>
- [30] L. Hasanah and B. Yuniarto, “Early study of LLM implementation in survey interviews”, 2025. [Online]. Available: <https://doi.org/10.34123/jurnalasks.v17i1.792>
- [31] P. Hämäläinen, J. Oulasvirta, and T. Jokinen, “Evaluating large language models in generating synthetic HCI research data: a case study,” 2023. [Online]. Available: <https://doi.org/10.1145/3544548.3580688>
- [32] Laura Weidinger et al., “Taxonomy of Risks Posed by Language Models”, 2022. [Online]. Available: <https://doi.org/10.1145/3531146.3533088>
- [33] Selina Meyer, David Elsweiler, Bernd Ludwig, Marcos Fernández-Pichel, and David E. Losada, “Do We Still Need Human Assessors? Prompt-Based GPT-3 User Simulation in Conversational AI”, 2022. [Online]. Available: <https://dl.acm.org/doi/10.1145/3543829.3544529>
- [34] J. Roberts, M. Baker, and J. Andrew, “Artificial intelligence and qualitative research: The promise and perils of large language model ‘assistance’”, 2024. [Online]. Available: <https://doi.org/10.1016/j.cpa.2024.102722>
- [35] M. Liu, C. Xue, and Y. Zhai, “From Lab to Virtual: Comparing Real and AI-Generated User Interviews in Home Appliances Evaluation”, 2024. [Online]. Available: <https://doi.org/10.1145/3629606.3629672>
- [36] S. Tschepe, “AI-Enhanced Design Thinking,” 2024. [Online]. Available: <http://dx.doi.org/10.13140/RG.2.2.35306.12484>
- [37] M. Liu, T. Wang, C. A. Cohen, S. Li, and C. Xiong, “Understand user opinions of large language models via LLM-powered in-the-moment user experience interviews”, 2025. [Online]. Available: <https://doi.org/10.48550/arXiv.2502.15226>
- [38] A. Wuttke, M. Aßenmacher, C. Klamm, M. M. Lang, Q. Würschinger, and F. Kreuter, “AI conversational interviewing: Transforming surveys with LLMs as adaptive interviewers”, 2024. [Online]. Available: <https://doi.org/10.48550/arXiv.2410.01824>

- [39] A. Schuller, D. Janssen, J. Blumenröther, T. M. Probst, M. Schmidt, and C. Kumar, “Generating Personas Using LLMs and Assessing Their Viability”, 2024. [Online]. Available: <https://doi.org/10.1145/3613905.3650860>
- [40] J. Buchmüller, S. Schneider, and A. Kintz, “Generating user experience based on personas with AI assistants”, 2024. [Online]. Available: <https://doi.org/10.1145/3639478.3639810>
- [41] X. Zhang et al., “Auto-Generated Personas: Enhancing User-centered Design Practices among University Students”, 2024. [Online]. Available: <https://doi.org/10.1145/3613905.3651043>
- [42] Eaman Jahani, Benjamin S. Manning, Joe Zhang, Hong-Yi TuYe, Mohammed Alsobay, Christos Nicolaides, Siddharth Suri, and David Holtz, “Prompt Adaptation as a Dynamic Complement in Generative AI Systems”, 2024. [Online]. Available: <https://doi.org/10.48550/arXiv.2407.14333>
- [43] Michael Hewing and Vincent Leinhos, “The Prompt Canvas: A Literature-Based Practitioner Guide for Creating Effective Prompts in Large Language Models”, 2024. [Online]. Available: <https://doi.org/10.48550/arXiv.2412.05127>
- [44] Barbara Kieslinger and Rosa Wippler, “Simulating the Human in HCD with ChatGPT: Redesigning Interaction Design with AI”, 2024. [Online]. Available: <https://doi.org/10.1145/3637436>
- [45] Nielsen Norman Group, “AI as a UX Assistant”, 2024. Accessed: May 07, 2025. [Online]. Available: <https://www.nngroup.com/articles/ai-roles-ux/>
- [46] Virginia Braun and Victoria Clarke, *Thematic Analysis: A Practical Guide*. SAGE Publications, 2021.
- [47] Nielsen Norman Group, “Personas: Definition and Uses”, 2014. Accessed: May 07, 2025. [Online]. Available: <https://www.nngroup.com/articles/persona/>
- [48] Eaman Jahani, Benjamin S. Manning, Joe Zhang, Hong-Yi TuYe, Mohammed Alsobay, Christos Nicolaides, Siddharth Suri, and David Holtz, “As Generative Models Improve,

People Adapt Their Prompts”, 2024. [Online]. Available:  
<https://doi.org/10.48550/arXiv.2407.14333>

# Appendices

## Appendix I: Interview guideline

**Introduction:** This study is part of a Master's Thesis conducted at the University of Tartu (Estonia). The objective of the research is to explore how synthetic user data generated by Large Language Models (LLMs) compares with real-world user insights, particularly in the context of online food delivery applications.

The study aims to understand user experiences, behaviors, and pain points when interacting with food delivery apps and to evaluate whether LLM-simulated personas and scenarios can replicate similar qualitative insights. During the study, the researcher will conduct in-person interviews with selected participants. Conversations will be audio recorded with the participant's consent (collected in writing prior to the interview).

### **Materials, software and environment:**

- Printed consent form
- Printed interview guideline
- Printed interview question sheet
- Voice recorder (smartphone-based)

### **Overall procedure:**

#### *Before the interview:*

1. Contact the participant and confirm availability for an in-person interview.
2. Print and prepare the consent form and the interview question sheet.
3. Bring all necessary materials listed above.
4. Assign an anonymous index code to each participant in advance (e.g., H1–H5).
5. Ensure a quiet and comfortable setting for conducting the interview.

#### *At the beginning of the interview:*

1. Welcome the participant and thank them for participating.
2. Introduce yourself and your affiliation with the University of Tartu.
3. Explain the purpose and scope of the study (see Introduction above).
4. Provide the printed consent form and ask the participant to sign it.
5. Ask if the participant has any questions before proceeding.

6. Start the audio recording (no video).
7. Confirm verbal consent on the record.

*During the interview:*

1. Conduct the interview using a semi-structured format, covering the interview themes:
  - a) Useful and difficult app features
  - b) Experiences with order modification
  - c) Order abandonment
  - d) Navigation issues
  - e) Feature suggestions and general feedback
2. Allow for open-ended elaboration and clarification where needed.

*After the interview:*

1. Stop the audio recording and thank the participant again.
2. Assign the corresponding index number to the audio file.
3. Store the audio file in a designated, secure folder (separate from the consent forms).
4. Transcribe the recording for qualitative analysis.
5. Note down initial impressions or themes that stood out.

## Appendix II: Interview consent form

### Consent to Act as a Participant in a Research Study

**Study title:** User Experience and Usability of Food Delivery Apps

**Principal Investigator:** Viktorija Suslova, University of Tartu (Estonia)

**Introduction:** You are invited to participate in a user research interview regarding food delivery applications. This study aims to understand user experiences, preferences, and challenges when using such platforms. Your participation is voluntary, and you may withdraw at any time without consequence.

**Content of the study:** This study is conducted as part of my master's thesis research at the University of Tartu. During this interview, you will be asked questions about your experiences with food delivery apps, including ease of use, features you find helpful, and any difficulties you may have encountered. The interview will take approximately 20-30 minutes and may be recorded for analysis. Your responses will be anonymized and used solely for research purposes.

**Participation requirements:** Any person 21 or older who is using food delivery applications is eligible to participate.

**Risks and Benefits:** The risks that are associated with this research are no greater than those ordinarily encountered in daily life. There are no direct benefits to participants but the researchers anticipate potential societal benefits being derived from their research.

**Privacy and Confidentiality:** Your responses will be anonymized and used solely for research purposes. The researchers will follow strict procedures to protect participants' identities. The original audio files will remain on the original recording device, which is only accessible to the Principal Investigator. The audio files will be transcribed using 3rd-party tools, and any potential identifiers will be removed or aggregated. The original audio files will be deleted after transcription. Your data and consent form will be stored separately. Your consent form will be securely stored and will not be disclosed to third parties.

By participating, you understand and agree that the data and information gathered during this study may be used by the University of Tartu for publication purposes. However, no identifiable information will be mentioned in any publication or dissemination of research data or results. The University of Tartu requires all research records to be maintained for at least five years following the final reporting or publication of a project. Aggregated data will be archived by the Principal Investigator for that timespan.

**Questions about the Study:** If you have any questions, comments, or concerns about the study before, during, or after participation, please contact Viktorija Suslova at [vsuslova@ut.ee](mailto:vsuslova@ut.ee).

**Voluntary Participation:** Your participation in this research is voluntary. You may discontinue participation at any time during the research activity. Your decision regarding whether to participate in this study will not result in any loss of benefits to which you are otherwise entitled.

**Participant:** The above information has been explained to me, and all of my questions have been answered. I understand that I am encouraged to ask questions, voice concerns, or make

complaints about any aspect of this research study. I understand that future questions, concerns, or complaints will be answered by the investigator(s) listed on this document.

I am 21 or older. I have read and understood the information above. I consent to participate in this research.

Yes  No

**Participant's Name:** \_\_\_\_\_

**Participant's Signature:** \_\_\_\_\_

**Date:** \_\_\_\_\_

**Researcher's Name:** \_\_\_\_\_

**Researcher's Signature:** \_\_\_\_\_

**Date:** \_\_\_\_\_

### **Appendix III: Interview questions**

1. Are there any specific app features that you find particularly useful?
2. Are there any specific app features that you find particularly difficult?
3. Have you had any issues with modifying your order when using food delivery services?
4. Have you ever abandoned an order before completing it? If so, why?
5. Have you had any issues with navigation, for example finding specific items within the app?
6. What improvements or new features would you like to see in food delivery apps?
7. Is there anything you would like to add that we haven't discussed?

## Appendix IV: Interview responses

*H#* denotes a real-world interview and its participant number, while *LLM#* denotes a synthetic interview generated by the LLM and its corresponding number.

	<b>Q1. Are there any specific app features that you find particularly useful?</b>	<b>Q2. Are there any specific app features that you find particularly difficult?</b>	<b>Q3. Have you had any issues with modifying your order when using food delivery services?</b>	<b>Q4. Have you ever abandoned an order before completing it? If so, why?</b>	<b>Q5. Have you had any issues with navigation, for example finding specific items within the app?</b>	<b>Q6. What improvements or new features would you like to see in food delivery apps?</b>	<b>Q7. Is there anything you would like to add that we haven't discussed?</b>
<b>H1</b>	A feature that allows me to set the food delivery time myself and schedule it for a specific time. Also, having different payment options - local apps mainly offer card payments, while in England, cash payment upon delivery is often available. Additionally, the ability to search not by restaurant or dish but	There were no particularly complex features, and the interface was quite user-friendly. I didn't have any issues with Estonian apps, but in London, there is a huge selection, and food delivery apps have too many options - "Are you sure?", "Would you like something else?",	There were times when I had to call the restaurant to make changes because online support in the chat didn't handle it. I'd like to have the option to resolve issues without calling the restaurant - I prefer not to talk to people. I didn't see such an	No, only with modifying the orders.	In one app, the cart resets when returning to the restaurant/food search, which makes using the app inconvenient. I ended up deleting it.	The app could have options for customizing dishes - right now, the choices are limited compared to ordering in person at a restaurant. It would be great to have the same selection as in the restaurant. A chat with the restaurant would also be helpful to avoid calling or dealing with online support. Additionally, an option	Adding calorie and nutrient information for each dish would be a great improvement. Additionally, allowing the combination of different orders into one courier would be beneficial, so you don't have to place 10 separate orders for a team-building event.

	by meal type, such as breakfast, lunch, etc.	"Do you want to add this or that?" It's confusing.	option in the navigation. Chat support is available for resolving issues if there's a problem with the delivered order.			to add packaging for hot food would be useful. As a mom, I find there aren't enough options for kids - there should be dedicated choices. I'd also like mandatory labelling for kid-friendly meals (e.g., non-spicy).	
<b>H2</b>	Nothing special- I usually order only for dinner and not very often, so I don't remember anything particularly noteworthy.	Ordering and payments are simple. However, there are issues with ordering from certain restaurants that don't deliver to my location.	No issues - I've done this before but don't remember any negative experiences. I used online support for it.	No, I didn't abandon the order, but the app canceled it in the evening after we had already paid. App support informed me that the restaurant couldn't process the order and they will return the payment.	Nothing special, it seems complicated.	There could be more choices available for my location. The search could also be simplified - right now, there are too many categories and options. Maybe for someone who uses the app daily, it's easy to navigate, but for me, it feels too complicated.	I would like to have McDonald's options available for Nömme.

<b>H3</b>	Yes, there is a feature where you can use one courier for deliveries from different places. It's also useful that you can filter by food type, like sushi or burgers.	There are no complex functions.	I haven't needed to modify an order.	I never canceled an order, but I had orders canceled on me. One time, no one responded for over 10 minutes, and then I got a message saying the restaurant wasn't responding, so the order was canceled. Another time, after placing an order, the restaurant called and suggested a replacement.	No issues, easy to use.	I'd like to have statistics for my profile - what I eat, order frequency, and expenses. It would be useful for tracking my budget.	Too many features would make the app overly complicated. People value simplicity.
<b>H4</b>	The most useful feature is the ability to quickly select the type of food you want, such as "Thai" or "Asian."	There are no complex functions - the challenges are mostly related to the restaurants themselves.	I haven't used this option, but there is a new feature that allows adding items from a nearby store.	I have canceled an order because the estimated delivery time was too long and didn't match the initially promised time. I contacted app	Navigation isn't as detailed as I'd like at times, but overall, it's clear enough.	There could be a feature to choose the search language - currently, search results vary depending on whether you search in Estonian or English, showing different lists.	Additionally, orders could be standardized across all restaurants, ensuring a consistent process and a clear breakdown of included and additional items.

				support and canceled it through an automated chat option.			
<b>H5</b>	A very convenient app - you can place an order in just a few clicks and choose a convenient delivery time.	I would like all free items to be automatically included in the order, just like in a restaurant, without the need for additional confirmation	There were no issues with modifying the order.	I had to cancel an order once because the delivery was delayed by 1.5 hours.	There were no issues with navigation.	A suggestion - automatic inclusion of free items, so you don't need to do a lot of clicks.	No other comments.
<b>LLM1</b>	As a busy mom and professional, I appreciate features that save me time. The "save for later" or "favorite orders" feature is super useful because it allows me to quickly reorder meals that my family likes. I also really like real-time order tracking - it gives	I find it frustrating when the app doesn't let me customize my order easily, like when I can't add or remove ingredients. The process of modifying an order sometimes feels clunky, and I've had times where I have to	Yes, occasionally. If I want to change something after I've placed my order, it's not always clear how to do that. Sometimes, the "modify order" button isn't visible or doesn't work correctly. This is	Yes, I've abandoned orders a couple of times, usually because the app was too slow or confusing. If the checkout process takes too long or there are too many steps, I'll just give up and find another	Definitely! Sometimes, the app is not intuitive enough, and I struggle to find specific items. Categories don't always match what I'm looking for, and I end up scrolling through endless lists of food. A better	I would love to see a feature that allows me to quickly see meal options that fit my family's preferences or dietary needs. For instance, if the app could recognize if we're ordering more kid-friendly meals or healthier options, that	I think having more options for flexible delivery time slots would make a big difference, especially since my schedule as a working mom can change. It would also be helpful to have a "recommended for you" section based on

	<p>me peace of mind to know exactly when my food will arrive. Payment options like card save or mobile payment integrations are very convenient for me too.</p>	<p>restart the app if something wasn't processed correctly. Also, some apps have poor filtering options, so it's hard to quickly find food that matches my dietary preferences (like vegetarian or gluten-free).</p>	<p>frustrating, especially if I've noticed something wrong or want to make a quick adjustment.</p>	<p>way to get food. Also, if the delivery time feels uncertain or too long, I'll abandon the order because I need to know that food will arrive when expected.</p>	<p>search function with more filters (like dietary preferences, meal types, or price ranges) would make things much easier.</p>	<p>would be super helpful. Also, an option for bundled family meals or meal deals would be great. And perhaps an automatic loyalty program where I earn points for each order could be an added bonus.</p>	<p>past orders, but without feeling like the app is bombarding me with too many suggestions. Personalization is great, but I don't want it to feel overwhelming.</p>
<b>LLM2</b>	<p>I really appreciate the ability to save my favorite orders or restaurants in the app. It saves me time when I'm too tired or busy to scroll through the options. Also, the estimated delivery time is always helpful. I can plan my meals better, especially when juggling work and kids.</p>	<p>One thing I don't like is when the app doesn't update the availability of certain items in real time. Sometimes I'll add something to my cart, only to find out it's out of stock when I go to pay, which can be frustrating. Another issue is the menu layout - it's</p>	<p>Yes, I have encountered problems modifying orders, especially when I try to make changes after I've already placed the order but before it's been confirmed. Some apps are really slow to update the changes, or they don't allow any</p>	<p>Yes, I have abandoned an order once. It was mostly due to confusion in the checkout process. I wasn't sure if I had selected all the items I wanted, and the app's interface wasn't clear about whether I had added the right extras or if</p>	<p>Sometimes, yes. I find that the navigation can be cluttered, especially when there are too many categories. It's not always intuitive. For example, I'll often struggle to find kid-friendly meals or specific dietary options like gluten-free, because the</p>	<p>I would love to see a "family-friendly" filter where I can quickly find meals suitable for kids. Also, more detailed nutrition information would be helpful, so I can make healthier choices for my family. It would also be nice if the app could track rewards or loyalty points more clearly,</p>	<p>One thing that would really improve my experience is better customer support. Sometimes when there's an issue with the delivery or an incorrect order, it's hard to get in touch with someone who can fix it. A live chat feature or a quicker response time would</p>

		sometimes hard to find what I'm looking for when I'm in a rush.	modifications after you've clicked "confirm," which feels limiting.	it was charging me the correct amount. I didn't want to risk paying more than I expected, so I just gave up.	filters don't work as well as they could. I have to scroll through so much before finding what I need.	because I forget to use them sometimes.	make me feel more comfortable using the service more often.
<b>LLM3</b>	Yes, I really appreciate the order tracking feature that shows you when your food is being prepared, picked up, and when it's about to arrive. It's especially helpful when I'm juggling my work and my child at home. Also, I like the personalized recommendations based on my past orders - it saves time when I don't want to decide on something new.	I struggle with the search functionality sometimes. It's hard to filter out specific preferences like dietary restrictions (e.g., gluten-free, vegan) without getting irrelevant results. Also, when I try to add extra items or modify a dish, I often find that the options aren't clear or the modifications don't show up correctly in the cart.	Yes, modifying my order can be frustrating. Sometimes when I try to make changes, like adding extra toppings or swapping an ingredient, the app either doesn't update correctly, or the restaurant says the change isn't possible even though I selected it. I've also faced issues where the	I've abandoned an order once or twice, usually when I couldn't find what I was looking for or the process was taking too long. For example, when I couldn't filter options by what I needed, or when the app kept loading slowly, I just gave up. It's especially frustrating when I'm on a tight schedule with my	Yes, navigating can be a bit of a hassle. I often find myself scrolling through many categories to find a dish or restaurant that fits my needs. The app tends to have a lot of options, but it's not always intuitive. Sometimes I have to go back to the home screen to start over or search again.	It would be great if the app had a better filtering system for specific dietary needs, like gluten-free, low-carb, etc. A family-friendly section would also be useful, where I can quickly find kid-friendly meals. I'd love an option to save custom meal preferences (for example, no dairy or no spicy food) that automatically apply to my orders. Another	One thing I think would improve the experience is a better loyalty or reward system. I feel like if I order frequently, there should be more incentives, like discounts or special offers. Also, a clearer delivery window would help me plan better since sometimes the wait feels unpredictable.

			price doesn't adjust after modification, which creates confusion.	kid, and I can't afford to waste time.		thing would be multi-person orders, where you can easily order for the whole family and customize each person's meal.	
<b>LLM4</b>	I find the real-time tracking feature really useful. It helps me know exactly when my food will arrive, and I can plan around that. The ability to save favorite orders is also great since it saves me time when I'm ordering the same thing. I like having a clear overview of my order and the delivery status, so I'm not left guessing when my food will arrive.	I often find the payment process a bit cumbersome, especially when I need to switch between different payment methods or add a new card. It's also frustrating when the app doesn't remember my payment details correctly or when I have to go through several screens just to apply a discount or promotion code. It's	Yes, sometimes it can be a bit tricky to modify an order, especially when it's already been confirmed. If I realize I forgot to add something or want to remove an ingredient, there's often no clear way to do it once the order is in the system. The app sometimes doesn't allow for changes, or it requires me to	Yes, I've abandoned an order a couple of times. Sometimes, if the app is taking too long to load or if there are too many errors while adding my delivery address, I just give up. If there's a glitch or if I feel like the process is dragging on too much, I just don't have the patience to go through with it.	Occasionally, yes. While the general layout is okay, I've had trouble finding specific items or categories, especially when there's a lot of variety. The search function isn't always as precise as I'd like. For example, if I'm looking for a vegetarian dish, sometimes the results can be mixed in with all other categories, making it harder to	I'd love to see a more intuitive recommendation system that suggests meals based on my past orders or preferences. Also, a feature to see how the meal is being prepared (e.g., an option to see if it's being cooked or has already been packed) would make the process more transparent. Improved filtering options would also be great, so I can quickly filter by dietary	One thing that's really important to me is the customer service experience. If I ever have an issue with my order, I want to know that there's a quick way to resolve it. Sometimes, it feels like reaching out for support can be slow, or I don't get an immediate solution. Better communication from the app's side would go a long way

		not the worst experience, but it definitely feels like an unnecessary hassle at times.	cancel the entire order and reorder, which is a bit inconvenient.	Also, if the delivery options seem overly complicated or there are too many unnecessary steps to finalize the order, I tend to walk away from it.	narrow down my options. It can get overwhelming when I just want to order something quickly.	preferences (e.g., vegetarian, gluten-free) or even by what's available for delivery at the moment.	in improving my experience.
<b>LLM5</b>	Yes, I really appreciate when apps allow me to easily filter food based on dietary preferences or restrictions, like gluten-free or vegetarian options. The estimated delivery time is super helpful too, especially when I'm home with my kids and trying to plan my day. I also love when the app offers an easy way to repeat previous orders. It saves me time,	Sometimes, the payment options can be a hassle. I prefer using my bank card, but some apps require me to re-enter payment details every time, and it feels like an unnecessary step. Also, the notifications can be overwhelming at times. I don't need constant updates about my order if everything's on track,	Yes, I've had issues in the past when trying to change something after placing the order. The app sometimes doesn't update my modifications in real-time, and I end up getting the wrong item. It's a bit frustrating, especially when I'm ordering for my	Yes, I've abandoned orders before. Mostly when the app was too slow, or I couldn't find the item I was looking for. Once, I was ordering a pizza, and the app crashed, so I just gave up. Other times, if the menu wasn't clear or didn't show the prices upfront, I felt	Yes, navigation can sometimes be tricky. If I'm trying to find a specific dish or restaurant, it's not always intuitive, and I end up scrolling through categories or searching multiple times. It would be great if there were clearer labels or a better search function	It would be amazing if there were a "family meal" feature that combines meals for kids and adults in one easy package. Also, I'd love more integration with local businesses. I want to support smaller, local restaurants, but sometimes they're harder to find on the apps. A better loyalty program would also be nice, especially for	I would just say that having better customer service options is really important. If there's a problem with my order, I want to be able to contact someone quickly. Sometimes it feels like it takes too long to get a response, and when I'm juggling kids and a busy day, it makes it even more stressful.

	especially when I'm running on little sleep.	especially when I'm multitasking with my kids.	children and need something specific.	like it was too much hassle to keep going.	to find exactly what I want.	regular customers like me.	
--	--	--	---------------------------------------	--	------------------------------	----------------------------	--

## **Appendix V: List of Activities Performed to Choose AI Tool for UX Interview**

To identify the most suitable AI tool for generating synthetic interview responses, a structured search and evaluation process was followed.

1. **AI-Assisted Search** - An exploratory query was conducted using ChatGPT with the following prompt:  
*“Provide a list of AI tools that can simulate user research interviews and respond to pre-defined interview questions instead of real users.”*
2. **Search Engine Research** - Google was used to identify additional applications relevant to UX interview simulation.
3. **Preliminary Screening** - All tools identified through both methods were reviewed and briefly tested against two basic criteria:
  - The ability to simulate user research interviews responses
  - The ability to use pre-defined interview questions
4. **Functionality Evaluation** - Tools that passed the initial screening were compiled into a summary table and assessed against three key functional requirements:
  - Generation of semi-structured interview responses
  - Support for pre-defined parameters (interview questions)
  - Availability of a free version

This process ensured that the final tool selection was based on both broad market scanning and targeted functionality testing, aligning with the study’s methodological requirements.

## Appendix VI: List of AI tools for simulation UX interviews

Name	Website	Simulated interview responses	Pre-defined questions	Free version available	Use case example	Verdict	Best for
ChatGPT (OpenAI)	<a href="https://openai.com/chatgpt">https://openai.com/chatgpt</a>	Yes	Yes	Partial	A UX researcher uses ChatGPT to simulate user feedback before conducting real interviews, refining questions and responses accordingly.	Provides flexible, text-based simulated interviews that adapt dynamically.	General-purpose, quick-turnaround idea testing, refining interview scripts, early-stage concept validation.
User Interview Simulator	<a href="https://www.yeschat.ai/">https://www.yeschat.ai/</a>	Yes	Yes	Partial	Product managers, UX/UI designers, researchers, and developers looking to understand their user base better and refine their products based on simulated user feedback can benefit from this tool.	Allows you to conduct simulated interviews by asking direct questions in a chat interface and interacting dynamically with the AI persona.	Simulating realistic user personas for early product feedback and question refinement.

Delve AI's Synthetic Research	<a href="https://www.delve.ai/">https://www.delve.ai/</a>	Yes	Yes	No	UX researchers, product managers, designers, marketers, startups, and academics to quickly run simulated user interviews without recruiting real participants. It helps test ideas, interfaces, and messaging with AI-generated personas, providing fast, cost-effective, and directional insights early in the process.	Allows realistic, interactive interview simulations with AI personas based on your prompts, but no free version for interviews.	Scalable, multi-persona simulated interviews for directional insights before live testing.
UXPressia	<a href="https://uxpressia.com/ai-persona-chat?utm_source=chatgpt.com">https://uxpressia.com/ai-persona-chat?utm_source=chatgpt.com</a>	Yes	Yes	No	UX researchers, designers, marketers, product managers, and educators who want to test ideas, explore user perspectives, and refine customer journeys through live AI persona conversations.	Enables live AI persona conversations for UX research and design, but only available from the Pro plan upward.	Testing and refining customer journeys, messaging, and UX flows through live persona interactions.

Heysho	<a href="https://heysho.com/en/tool/user-interview.html">https://heysho.com/en/tool/user-interview.html</a>	Yes	Yes	Partial	UX researchers, product managers, designers, marketers, startups, and educators who need to quickly test research questions or hypotheses with multiple AI personas. It's especially useful for gathering diverse perspectives early in a project, validating ideas before real user testing, and training teams in interview techniques.	Useful for UX researchers, product managers, designers, and marketers to quickly test ideas with multiple AI personas and gather diverse early-stage insights.	Gathering diverse persona perspectives quickly for early-stage concept validation and research training.
--------	---	-----	-----	---------	---	--	--

## **Appendix VII: List of Activities Performed to Choose AI Tool for Creating UX Persona**

To identify the most suitable AI tool for creating user personas, a structured search and evaluation process was followed.

1. **AI-Assisted Search** - An exploratory query was conducted using ChatGPT with the following prompt:  
*“Provide a list of AI tools that can create user persona.”*
2. **Search Engine Research** - Google was used to identify additional applications relevant to UX persona creation.
3. **Preliminary Screening** - All tools identified through both methods were reviewed and briefly tested against two basic criteria:
  - The ability to create user personas
  - The ability to use pre-defined parameters (data)
4. **Functionality Evaluation** - Tools that met the initial screening were compiled into a summary table and assessed against three key functional requirements:
  - Creation of personas with a customizable template
  - Support for pre-defined parameters (data)
  - Availability of a free version

This approach ensured that the final selection was based on both broad market scanning and targeted functionality testing, aligning with the methodological requirements of the study.

## Appendix VIII: List of AI tools for creating UX personas

Name	Website	Customizable templates	Based on interview data provided	Free version	Use case example	Verdict	Best for
AI User Persona Generator by Bot Memo	<a href="https://botmemo.com/ai-user-persona-generator/?ref=taaft_feat&amp;utm_source=taaft_feat&amp;utm_medium=referral">https://botmemo.com/ai-user-persona-generator/?ref=taaft_feat&amp;utm_source=taaft_feat&amp;utm_medium=referral</a>	No	No	Yes	Entrepreneurs, marketers, product managers, researchers needing quick target audience insights.	Free AI tool that creates detailed text-based personas from simple inputs; best used as a starting point before real research.	Quick, free creation of text personas for later visual or data-backed refinement.
ChatGPT (OpenAI)	<a href="https://openai.com/chatgpt">https://openai.com/chatgpt</a>	No	Yes	Partial	A UX researcher uses ChatGPT to create several user personas, adjusting	Highly flexible, allowing iterative persona refinement based on	Flexible, conversational creation of text-based personas that can be iteratively

					them based on interview data and refining key attributes dynamically.	text-based AI generation.	refined through interactive prompts.
Writecream	<a href="https://www.writecream.com/user-persona-generator">https://www.writecream.com/user-persona-generator</a>	No	No	Yes	UX researchers, product managers, marketers, startups, and educators who need fast, AI-generated personas to guide design, product, and marketing decisions.	Quickly generates AI-based user personas from minimal input, ideal for fast audience profiling.	Quick, free generation of text-based personas that can be refined and visually designed later.
PersonaGenerator	<a href="https://www.personagenerator.com/">https://www.personagenerator.com/</a>	Built-in	No	Yes	UX researchers, marketers, product managers, and	Simple, free tool that produces detailed text-based	Quick, no-cost creation of ready-to-use text personas for

					startups wanting fast AI-generated personas without complex setup.	personas instantly from minimal input.	early design, marketing, or product planning.
Creately	<a href="https://creatly.com/lp/user-persona-creator">https://creatly.com/lp/user-persona-creator</a>	Yes	No	Yes	UX designers, marketers, product teams, and educators who need collaborative, visually rich persona creation with templates and diagrams.	Drag-and-drop persona creator with intuitive visual templates and collaborative features. Free tier available.	Team-based creation of visually designed personas using drag-and-drop templates for workshops, presentations, and strategy sessions.

## **Appendix IX: List of Activities Performed to Choose AI Tool for creating Scenario**

To identify the most suitable AI tool for generating user scenarios, a structured search and evaluation process was applied.

1. **AI-Assisted Search** - An exploratory query was conducted using ChatGPT with the following prompt:  
*“Provide a list of AI tools that can create user scenarios.”*
2. **Search Engine Research** - Additional relevant tools were identified through targeted Google searches.
3. **Preliminary Screening** - All tools identified through both methods were reviewed and briefly tested against two basic criteria:
  - The ability to create user scenarios
  - The ability to incorporate pre-defined parameters (data)
4. **Functionality Evaluation** - Tools that met the initial screening were compiled into a summary table and assessed against three key functional requirements:
  - Creation of scenario with a customizable template
  - Support for pre-defined parameters (data)
  - Availability of a free version

This approach ensured the final selection was informed by both broad market scanning and targeted functionality testing, aligning with the methodological needs of the study.

## Appendix X: List of AI tools for creating UX scenarios

Name	Website	Customizable templates	Based on persona data provided	Free version	Use case example	Verdict	Best for
ChatGPT (OpenAI)	<a href="https://openai.com/chatgpt">https://openai.com/chatgpt</a>	No	Yes	Partial	A UX researcher uses ChatGPT to generate multiple versions of a scenario for A/B testing different UI solutions.	Is flexible and conversational, useful for quick ideation and refinement.	Quick, conversational scenario ideation from persona inputs. Requires human refinement for accuracy
GenerateStory	<a href="https://generatestory.io/scenario-generator/?utm_source=chatgpt.com">https://generatestory.io/scenario-generator/?utm_source=chatgpt.com</a>	Partial	Partial	Yes	Writers, game designers, educators, and creative teams looking to quickly generate diverse, detailed scenarios for storytelling, training, or brainstorming.	Flexible AI tool that produces customizable scenarios from simple prompts, suitable for both fictional and instructional purposes.	Rapid creation of varied, text-based scenarios that can be adapted for narratives, games, or learning activities.

LeanScope	<a href="https://www.leanscope.ai/en/user-scenarios/?utm_source=chatgpt.com">https://www.leanscope.ai/en/user-scenarios/?utm_source=chatgpt.com</a>	Partial	Yes	No	UX researchers, product managers, and service designers who need to map current (“as-is”) user scenarios to identify pain points and improvement opportunities.	AI-powered tool that visualizes user interactions based on research data or expert input, making it easier to analyze and optimize user journeys.	Creating detailed, research-backed user scenario maps to inform design decisions and prioritize enhancements.
-----------	---	---------	-----	----	---	---	---

## License

### Non-exclusive licence to reproduce the thesis and make the thesis public

I, **Viktorija Suslova**,

1. grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the digital archives of the University of Tartu until the expiry of the term of copyright, my thesis *LLMs as a Tool for User Experience Research: A Comparison of synthetic and real-world data* supervised by *Yana Halas and Grace Eden*.
2. grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright;
3. am aware of the fact that the author retains the rights specified in points 1 and 2;
4. confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Viktorija Suslova

12/08/2025