Federal University of São Carlos– UFSCar
Center for Exact Sciences and Technology– CCET
–
Postgraduate Program in Computer Science– PPGCC

Maylon Pires Macedo

Towards a UX Data-Centric Approach: Providing Tools to Support Software Development

São Carlos
2025

Maylon Pires Macedo

Towards a UX Data-Centric Approach: Providing Tools to Support Software Development

Thesis presented to the Postgraduate Program in Computer Science of the Center for Exact Sciences and Technology of the Federal University of São Carlos, as part of the requirements for obtaining the PhD. in Computer Science.

Supervisor: Prof. PhD. Luciana Aparecida Martinez Zaina

São Carlos
2025

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

Folha de Aprovação

Defesa de Tese de Doutorado do candidato Maylon Pires Macedo, realizada em 11/08/2025.

Comissão Julgadora:

Profa. Dra. Luciana Aparecida Martinez Zaina (UFSCar)

Profa. Dra. Carla Maria Dal Sasso Freitas (UFRGS)

Prof. Dr. Marco Antonio Alba Winckler (UCA)

Prof. Dr. Tayana Uchoa Conte (UFAM)

Prof. Dr. Daniel Lucrédio (UFSCar)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Ciência da Computação.

Acknowledgements

I thank God for allowing, blessing and coordinating all the opportunities I have received.

To my wife Letícia de Paula Monteiro who supports me in all my endeavors and is essential to continue. To my parents Adriana Aparecida Pires Macedo and Mauro Moreira Macedo who have always supported my studies. Thank you very much for your strength, understanding and affection.

To my PhD supervisor Luciana Zaina who welcomed me as a child of science and guides me along the paths of scientific thought. Thank you for your trust and patience.

To the evaluation committee for accepting to participate, contribute and enrich this research.

To all the friends of the User eXperience – Laboratory of Studies in Networks, Innovation and Software (UXLeris) research group and everyone who directly or indirectly is part of this journey.

Give me a sharp sense of understanding, a retentive memory,
and the ability to grasp things correctly and fundamentally.
Grant me the talent of being exact in my explanations,
and the ability to express myself with thoroughness and charm.
Point out the beginning, direct the progress, and help in completion,
through Christ our Lord. Amen.
(St Thomas Aquinas)

Abstract

The frequently collected data regarding user interaction with software (i.e., UX data) holds potential to enhance User Experience (UX), which encompasses users' emotional, practical, and perceived value interactions with products. However, a challenge exists between the collection of UX data and its effective analysis. Software practitioners frequently encounter difficulties in selecting relevant data and applying appropriate techniques to gain insights for software improvement. The field of Information Visualization (InfoVis) offers approaches to assist users in analyzing emerging information from data. While initial efforts explored how InfoVis principles could be applied, this research revealed a more fundamental problem. A Grey Literature Review on UX data not only identified diverse formats, definitions, and suggested visual representations but, exposed a pervasive lack of formal UX data definitions in practice. Two initial exploratory studies further reinforced these challenges, demonstrating that practitioners struggled to derive insights from unstructured UX data without a clear conceptual foundation. Building upon these findings, this research strategically pivoted its focus to address the core issue for effective UX data analysis. The central artifact developed is a UX data taxonomy, designed to organize and clarify UX data understanding by categorizing objectives, purposes, and data characteristics. This taxonomy itself represents a theoretical contribution, providing a structured model for the UX data domain. To ensure its practical utility and accessibility, this taxonomy was operationalized through two distinct tools: an interactive web-based system for exploration and a Generative AI (GenAI) RAG (Retrieval-Augmented Generation) chat tool for natural language interaction. The utility and acceptance of this taxonomy and its related tools were then empirically validated with software professionals. Finally, the research also explored emerging opportunities for UX data analysis, particularly concerning the role of GenAI, assessing how these advancements intersect with the need for structured data understanding, thus providing theoretical insights into the evolving landscape of UX data analysis. This thesis provides both practical tools to empower software professionals in leveraging UX data for enhancing interactive system design, and lays a theoretical foundation for understanding and working with UX data.

Keywords: UX Data. User Experience. Information Visualization. UX Data Taxonomy.

List of figures

List of tables

List of abbreviations and acronyms

DSR Design Science Research

GLR Grey Literature Review

GenAI Generative Artificial Intelligence

InfoVis Information Visualization

HCI Human-Computer Interaction

LLM large language model

MCP Model Context Protocol

RAG Retrieval-Augmented Generation

SML Systematic Mapping of the Literature

TAM3 Technology Acceptance Model Version 3

TLX NASA Task Load Index

UX User eXperience

UXLeris User eXperience – Laboratory of Studies in Networks, Innovation and Software

Contents

Chapter 1

Introduction

This chapter presents the context and motivation, the research problem, the objectives of this doctoral research, the overview of the methodological approach, the researcher's trajectory and the organization of the thesis.

1.1 Context and motivation

User eXperience (UX) encapsulates the entirety of a user's engagement with a product, system, or service. This includes not only its functional performance and ease of use but also the crucial emotional and perceptual responses that occur before, during, and after interaction (HASSENZAHL, 2018). This comprehensive view is formally recognized by the ISO 9241-210 standard, which defines UX as "a person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service" (STANDARD; ISO, 2019). The ISO 9241-210 standard further emphasizes that UX is influenced by the system's characteristics, the user's state (i.e., including prior experiences, attitudes, skills, and personality), and the context of use, thereby highlighting its subjective and comprehensive nature.

UX is an important aspect of software quality that affects user acceptance of the product (KASHFI; FELDT; NILSSON, 2019), encompassing both software functionalities and perceived quality characteristics (HASSENZAHL, 2018). Discussions about including UX from software conceptualization stages, such as requirement elicitation and ideation, are increasingly important for software improvement (KASHFI; FELDT; NILSSON, 2019; CHOMA et al., 2022). To sustain the development of UX in products, UX design activities centers on creating interfaces that cater to user needs and preferences (NORMAN; NIELSEN, 2018). It encompasses the entire product acquisition and integration process, from branding to functionality, all informed by a rigorous research process, often referred to as UX Research (NORMAN; NIELSEN, 2018; HASSENZAHL, 2018).

UX research is crucial for promoting an iterative design approach, frequently employing diverse methods such as, interviews and usability testing to refine products based on user feedback and transform theoretical insights into practical solutions (MILLET, 2018; CHRISTENSEN et al., 2020). Professionals performing UX research rely not only on technical knowledge but also on interpretive and context-aware judgment regarding user behavior, business goals, and technological constraints (HASSENZAHL, 2018; MARTINELLI; LOPES; ZAINA, 2022).

UX research can significantly leverage UX data, which refers to information gathered from users' interactions and interpretations of digital products, to create more usable designs that meet expectations and avoid superficial insights (CONVERTINO; FRISHBERG, 2017; OLDENBURG; HILLENBRAND, 2024; MARTINELLI; LOPES; ZAINA, 2022). UX data is naturally generated throughout the UX research process from the application of quantitative and qualitative techniques, and offers significant information for companies to increase profitability and gain a competitive edge when effectively analyzed by UX professionals (ZAINA; SHARP; BARROCA, 2021; FRITZ; BERGER, 2015).

Although data collection is considered a common task by software practitioners in a UX research phase, the scientific literature has presented some challenges, such as: difficulty in defining appropriate analysis methods, how to choose the suitable data from users, and how to interpret the data to obtain relevant information in scenarios where there is no specific professional to deal with UX (HOKKANEN; VÄÄNÄNEN-VAINIO-MATTILA, 2015; UNTERKALMSTEINER et al., 2016).

UX data maps data from users' interactions and decisions with interactive systems, enabling measurements and statements to understand users' reactions to software elements (LUTHER; TIBERIUS; BREM, 2020; TONG et al., 2022). It can be sourced from formal and informal collection methods, including spontaneous conversations with end users (FABIJAN; OLSSON; BOSCH, 2016; RIVERO; CONTE, 2017; KIEFFER. et al., 2019). The collected data serves various purposes from validating initial product ideas during pre-development phases (HOKKANEN; KUUSINEN; VÄÄNÄNEN, 2016) to assessing overall UX post-development (FABIJAN; OLSSON; BOSCH, 2016).

Although the concept of UX is well-established, the literature indicates a lack of a practical definition of UX data that can effectively guide software professionals in leveraging its power to improve UX throughout the development cycle (KASHFI; FELDT; NILSSON, 2019; FABIJAN; OLSSON; BOSCH, 2016). This gap is especially critical for software teams without dedicated UX professionals, as they often lack the practical knowledge needed to effectively leverage user interaction data. This underscores their need for tools and methods to gain insights into user behavior and preferences for creating user-centric solutions (KOESTEN; SIMPERL, 2021). While the literature advocates for more

UX data visibility to engage software practitioners with UX design (ZAINA; SHARP; BARROCA, 2021; KASHFI; FELDT; NILSSON, 2019), UX data is rarely explored by software practitioners (KASHFI; FELDT; NILSSON, 2019).

Visualizations are a suitable approach to help software professionals overcome the barrier of UX data abstraction and obtain relevant insights (LACHNER et al., 2016a; BUONO et al., 2020a). The Information Visualization (InfoVis) area covers aspects related to the design and functionalities of visualizations, assisting in developing visualizations more aligned with user needs (MUNZNER, 2014). Applying Human-Computer Interaction (HCI) premises, InfoVis aims to provide users with visualizations that facilitate the understanding of complex, abstract, and large datasets (WARE, 2012). The use of visualizations expands human perception capabilities, facilitating interpretation and decision-making (CARD; MACKINLAY; SHNEIDERMAN, 1999). Visual data representations can guide designers and developers through the exploration of UX data, reducing the effort to obtain information based on patterns and emergent data (KASHFI; FELDT; NILSSON, 2019; DESAI, 2019).

Studies highlight a growing interest in **understanding how UX data is handled** in design processes and communicated among software practitioners (KASHFI; FELDT; NILSSON, 2019; KOESTEN; SIMPERL, 2021; ZAINA; SHARP; BARROCA, 2021). However, much of the practical knowledge acquired daily by software and UX practitioners from UX data exploration is **not typically published in formal scientific literature** due to the laborious nature of academic publishing. Moreover, the business environment is primarily driven by practical outcomes, innovation speed, and market demands rather than by the development of academic research, which further contributes to the gap between industrial practice and academic dissemination. (GLASS; DEMARCO, 2006; RAINER; WILLIAMS, 2019).

Software professionals, especially those dealing with qualitative data in textual format, struggle to identify and extract meaningful UX insights, making the **extraction of UX-related information a complex task** (RHIE; YUN, 2017). This process requires defining meanings to provide different values to raw data, as analysis without clear definitions depends on individual researcher judgment, leading to variable results based on their knowledge and experience about UX (RHIE; YUN, 2017; DESAI, 2019).

While there are existing papers about UX data visualizations, the **literature on UX data definition is sparse**. The lack of an overview of the literature limits a precise understanding of its gaps and potential for further investigations. This highlights a critical need for comprehensive reviews that examine UX data visualizations, covering purposes, data sources, chart formats, interaction functionalities, and their objectives. To date, no other work has examined UX data visualization through as many diverse lenses as ours, providing a broad overview of objectives, data characteristcs, and gaps for HCI researchers and practitioners.

Finally, Generative Artificial Intelligence (GenAI)[1] is reshaping practices across digital product development cycles (KHALIFA; ALBADAWY, 2024). Among these transformations, GenAI presents an emerging opportunity to support user experience (UX) research, particularly in tasks involving the interpretation of user interaction data (e.g., navigation logs or clickstream events) (CHENG; LI; BING, 2023; HOU et al., 2024). This indicates a potential avenue for future research to bridge the existing gaps in UX data analysis and GenAI utilization.

1.2 Research problem

The challenge in exploring UX data to enhance interactive systems stems from: a lack of clear definition for what constitutes UX data and the subsequent difficulty in its practical implementation. While existing academic literature has explored various UX data visualization techniques, a significant gap persists in providing comprehensive overview or addressing diverse data sources beyond traditional logs. This ambiguity is further compounded by the realities of industry practice, where professionals often grapple with inconsistent data definitions and an over-reliance on empirical methods. The literature reveals a fragmented understanding and absence of standardized UX data classification and operationalization methods, which ultimately constrains its potential to truly inform and enhance software development. Considering the motivations and identified gaps in the literature, the following research problem is defined:

> The unclear definition of what constitutes relevant UX data and the complexities of its implementation, which prevents software professionals from adequately applying insights from user interaction data to guide and support the enhancement of interactive systems.

Based on the research problem, this doctoral research answers the following research questions:

- RQ1 - What is the current state of UX data visualization in literature, and what are the practical challenges of utilizing UX data in software development?
- RQ2 - What kind of user interaction information can be explored to support software practitioners in discovering opportunities to improve interactive systems?
- RQ3 - How can software professionals improve their understanding of UX data, and effectively integrate and operationalize this knowledge into their software development practices?

[1] GenAI, short for Generative Artificial Intelligence, encompasses machine learning models capable of generating high-fidelity outputs; Chapter 2 provides further details.

- RQ4 - What are the emerging opportunities for UX data analysis, particularly concerning the role of GenAI?

1.3 Objectives

This doctoral research addresses the critical deficiencies and emergent demands within software development concerning the exploration of UX data through visualizations. The general objective has two perspectives: (i) The criteria applied to select the suitable detail of UX data to support improvements in interactive systems?; and (ii) The application of visualization techniques to improve the interpretation and understanding of the information emerging from UX data.

Considering the two perspectives above, this doctoral research provides a UX data taxonomy as a main contribution. This taxonomy supports software professionals to recognize the potential of available UX data to then organize it into a visualization form, promoting a better communication of UX-related information embedded into this UX data. This solution aims to provide support and autonomy to software practitioners and encourage research on techniques related to UX data analysis to enhance interactive system design based on HCI principles (e.g., ease of interaction, data-driven design, User centered design), supported by interaction data to foster better user acceptance and experience.

In order to achieve the general objective, the following specific objectives were defined:

- Define the "state of the practice" by investigating scientific and grey literature[2] to see how software practitioners have explored UX data.

- Define the characteristics of UX data to help software practitioners classify available data and understand existing opportunities for interactive system improvements.

- Define of visualization practices that can be applied to UX data with information on *how* to use visualizations, *what* can be extracted from UX data, *when* to use visualizations, and *what* are the advantages of using certain visualizations.

- Design of a taxonomy to support software practitioners in: (i) recognize meaningful information embedded in user interaction data; (ii) integrate UX tasks into the development lifecycle to enhance user experience based on existing UX data; and (iii) using visual representations to obtain relevant information emerging from UX data;

[2] Grey literature is a term used to identify information that is widely disseminated and has not necessarily been peer-reviewed (ADAMS; SMART; HUFF, 2017).

- ❏ Evaluate the taxonomy with software professionals through field studies, interviews and questionnaires to validate and expand the proposal based on real usage scenarios.

- ❏ Explore how GenAI can facilitate software practitioners' interaction with the UX data taxonomy through natural language, enabling the retrieval of all relevant information from its knowledge base to support the discovery of opportunities for UX data analysis.

1.4 Methodological approach

The general methodology is based on Design Science Research (DSR) which implies that the steps have cyclical characteristics and can be executed again whenever necessary (HEVNER, 2007). The DSR process employed throughout this work is characterized by its iterative and adaptive nature, combining cycles of relevance and rigor with continuous design and evaluation.

This flexibility led to a significant pivot in the primary artifact of this research. Initially, the focus was on developing a visualization framework for UX data. However, during exploratory studies, it became evident that the definitions of UX data were insufficiently clear, thereby compromising a fundamental basis of the proposed visualization framework. The cycle of relevance and rigor of the DSR facilitated the identification of this critical issue and redirected efforts towards constructing a taxonomy for defining UX data. This strategic shift, rather than a deviation, underscores the DSR's inherent rigor, demonstrating the commitment to addressing the most fundamental problem identified.

Figure 1 offers a concise overview of the methodological approach employed in this doctoral research. While the project's execution was inherently non-linear and cyclical (i.e., a characteristic of DSR) the main activities have been organized sequentially into comprehensive steps (A to H) to facilitate understanding and highlight their contributions. These steps are further grouped into three distinct phases (G1 to G3) to enhance clarity regarding the project's evolution. This synthesized representation does not explicitly link these steps to the iterative cycles inherent in the DSR methodology (i.e., which will be fully explained in Section 8.3). The complete methodology and the definitions about our cycles of DSR will be presented in the Chapter 3.

The **G1** joined activities that focused on the constitution of a solid knowledge base and diagnosing the nature of the problem in UX data (i.e., Steps A to C). The goal was to understand both the state of the art and state of the practice to identify gaps and real needs. This involved empirical studies that not only explored visualization applications but also corroborated and deepened our understanding of the difficulties in working with unstructured UX data, reinforcing the necessity of the taxonomy as a foundational element.

The **G2** represents steps about the materialization of the thesis's main contribution (i.e., the taxonomy) and its operationalization into practical tools (i.e., Steps D to F). This demonstrates the applicability and practical impact of the taxonomy, transforming a conceptual artifact into tangible solutions.

The **G3** completes the scope of this doctoral investigation, evaluating, exploring new frontiers and consolidating the doctorate's findings and contributions. It demonstrates the breadth of the research and its connection with emerging technological advancements, showing how AI can complement UX data analysis and, implicitly, how a solid conceptual basis (i.e., like the taxonomy) remains crucial.

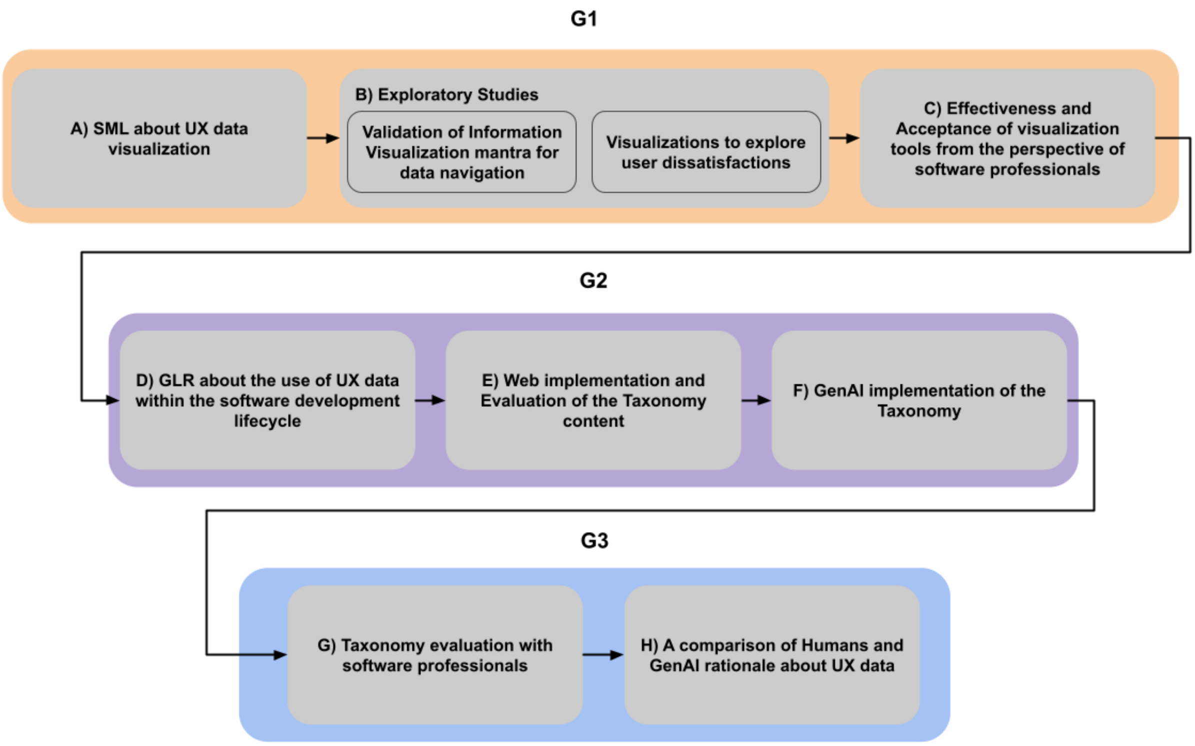


Figure 1 – Summary of the methodological approach of the doctoral research - Colors were used only for organization. Source: By the author.

The following paragraphs describe each step of the methodological approach overview, specifying whether it resulted in a scientific publication, and its contributions to the overall methodological progression (i.e., presented in the Chapter 3).

Step A: Problem Identification and Foundational Understanding. We started with a Systematic Mapping of the Literature (SML), detailed in **article**: "What is the State of the Art on UX Data Visualization? A Systematic Mapping of the Literature". This initial phase served to establish the state-of-the-art in UX data visualization, systematically investigating existing approaches to extract relevant information from UX data. This article **contribution** was to formally define UX data based on academic literature and pinpoint critical knowledge gaps. These insights were crucial for determining the application domain and identifying shortcomings in the scientific knowledge base, thereby fulfilling the initial premises of the relevance and rigor cycles.

Step B: Exploratory Studies and Problem Reinforcement. Two subsequent exploratory studies served as early design iterations, providing empirical feedback and further reinforcing the need for conceptual clarity in UX data. The **article** "Supporting User-Centered Requirements Elicitation from Lean Personas: A UX Data Visualization-Based Approach" is focused on validating initial knowledge by developing and evaluating visualizations for Lean Personas data. This article **contribution** provides initial feedback on visualization utility and suitability, confirming the applicability of Information Visualization mantras for data navigation. The **article** "Mind the Gap Between UX Data and Visualization Proposals: An Approach Emerged From the Grey Literature to Support the Analysis of User Dissatisfaction" brings **contribution** by validating visualization approaches and data sources in a more comprehensive scenario, expanding the understanding of conditions necessary for effective UX data usage. Both studies, while exploring visualization design, consistently highlighted the challenges users faced due to the ambiguity and unstructured nature of raw UX data, thereby indirectly validating the necessity of the proposed taxonomy as a foundational element for more effective analysis and visualization.

Step C: Practical Context. The **article** "Handling UX Data through visualization tools: an investigation of technology acceptance with tech-users"[3] provides crucial context on technology acceptance within the UX data domain. This study investigated the effectiveness and acceptance of visualization tools from the perspective of software development professionals. This article **contribution** was to offer insights into the practical conditions and potential benefits of integrating visualization tools into software development processes, setting the stage for understanding the broader challenges of adopting new data analysis technologies.

Step D: Exploring Practice and the Emergence of the Core Artifact. To bridge the gap between academic understanding and practical application, we conducted a Grey Literature Review (GLR), the findings are presented in the **article**: "Beyond Raw Data: A Taxonomy for UX Data Exploration to Support the Design of Interactive Systems". This investigation explored the utilization of UX data within the software development lifecycle. This article primary **contribution** was to reveal the "state of the practice", highlighting that UX data definitions and their use were often unclear and highly dependent on practitioners' empiricism. This crucial discovery directly led to the strategic pivot of our research. Instead of developing a generic visualization framework, it became evident that a more fundamental artifact was needed: a comprehensive taxonomy for UX data exploration. This taxonomy, which systematically organizes UX data by

[3] Under preparation at the time of thesis construction

objectives, purposes, and characteristics, emerged as our primary contribution.

Step E: Web implementation and Evaluation of the Taxonomy. To concretely demonstrate the practical applicability of the taxonomy, we implemented a web-based tool to facilitate interactive exploration and navigation of the taxonomy, allowing users to understand the relationships between objectives, purposes, and UX data. This implementation **contribution** showcase how our conceptual artifact can be operationalized to empower practitioners.

Step F: GenAI implementation. An AI Chat Tool utilizing Retrieval-Augmented Generation (RAG)[4] was developed, demonstrating an innovative approach to querying the taxonomy conversationally, effectively using the taxonomy's structured data as its knowledge base.

Step G: Artifact Evaluation. The **article**[5] focusing on the *evaluation of the UX data taxonomy by software professionals*, will represent the formal assessment of our primary artifact. This **contributions** directly evaluates the relevance, utility, and acceptance of our taxonomy by its target users. The results of this evaluation demonstrated the real impact and validation of our DSR solution.

Step H: Future Perspectives and Broader Impact. Finally, our research extends to emerging frontiers, showcasing the breadth and forward-looking nature of our work. **Article** "Can GenAI Provide Insights Like a UX Professional? A comparison of Humans and GenAI rationale about UX data" explores the role of Generative AI in UX data analysis. This article **contribution** comparing human and GenAI capabilities in interpreting UX data, offering insights into where AI aligns with expert perspectives and identifying areas requiring human insight. This discussion not only broadens the scope but also highlights how structured knowledge (i.e., like our taxonomy) remains vital even with advanced AI, providing the necessary foundation for more reliable and contextually relevant insights. This step updates the literature and informs the application domain about new possibilities and limitations in the use of UX data.

1.5 Researcher's trajectory

Throughout the doctoral journey, from February 2021 to July 2025, the activities undertaken were strategically structured around the pillars of teaching, research, and outreach. This organization reflects the comprehensive demands of an academic career within Brazilian public universities. This integrated approach allowed the development of theoretical knowledge, practical application and engagement with the community. Each experience, whether directly linked to the thesis or conducted in parallel, contributed

[4] Retrieval-Augmented Generation (RAG) is an AI framework that enhances the output of Large Language Models (LLMs) by retrieving relevant information from an external knowledge base to ground its responses. A more detailed definition is available in Chapter 2

[5] Under evaluation at the time of thesis construction

significantly to the researcher's development. Figure 2 presents columns indicating the years of the doctoral period and a line for each pillar (teaching, research, and outreach), with markings along the timeline illustrating when each activity occurred.

The initial phase of the doctoral experience included a technical training scholarship from October 2020 to July 2021 (see R1 in Figure 2), funded by FAPESP (process number 20/12343-3). This period was dedicated to the research and design of interfaces within immersive environments and user data collection for product enhancement. The insights gained from real-world application (i.e, outreach activity) directly informed the research trajectory and illuminated critical gaps in the existing literature, thus establishing a practical grounding for the initial phases of the doctoral research project.

The researcher's dedication to student development started with Gabriel Viana Teixeira, a scientific initiation scholar. From 2021 to 2023 (see R2 in Figure 2), the researcher provided tutoring in areas such as Information Visualization, Interaction data, and the planning and execution of controlled experiments. This productive partnership continued even after the conclusion of Gabriel's scientific initiation scholarship. This effort culminated in co-authorship on an article published at ICEIS 2024 and the software registration of a system designed for visualizing Lean Personas data, marking a tangible contribution to the field and a significant milestone in Gabriel's early academic career.

From August 2021 to February 2023, serving as a temporary lecturer at the Federal Institute of São Paulo (see T1 in Figure 2). Courses were delivered in programming, software engineering, and scientific research methodology for technical and postgraduate programs. This role include student tutoring and co-supervision, applying knowledge acquired during doctoral studies to guide aspiring professionals.

From July 2021 to July 2023, the researcher, in collaboration with Luciana Zaina, helped supervise Ariel Emanuel Sebastian Campos during their first year of master's research titled "UX-RIVIS: Information Visualization of UX Data Based on App Reviews" (see T2 in Figure 2). Collaboration was established in Information Visualization, Interaction Data, and Systematic Literature Mappings, with contributions in suggesting literature and tracking article writing. The student's subsequent honorable mention for the Junia Coutinho Anacleto Award for Best Master's Dissertation underscored the impact of this collaborative research and teaching effort.

The mentoring task continued with the co-supervision of Lucas Katib do Amaral, an undergraduate student FAPESP scientific initiation scholar (process number 22/11272-0), from October 2022 to October 2023 (see R3 in Figure 2). This role in teaching and research enabled guidance from the initial engagement with scientific inquiry through the successful execution of a systematic literature mapping on user interaction data visualizations. The outcomes, including a poster presentation at IHC 2023 and an article publication in the Journal on Interactive Systems (JIS).

A pivotal period for research and outreach development was from March to December 2023, when the researcher served as the lead researcher (process number 23/03414-2) on a FAPESP-funded Innovative Research in Small Businesses (PIPE FAPESP) project (process number 22/13247-3), culminating in the creation of a software startup (see O1 in Figure 2). This role involved managing a team of four scholarship holders and three associate researchers, focusing on software development, agile methodologies, documentation, innovation, data protection laws, user data collection, user research, data-driven improvements, and the application of GenAI tools in the software lifecycle. This experience provided a unique platform for conducting real-world user research and enriching the startup's product with doctoral insights, while simultaneously gaining practical experience that significantly matured the research perspective.

In July 2023, the researcher received the Gary Marsden Travel Awards (see A2 in Figure 2), which allowed him to participate in the Engineering Interactive Computer Systems (EICS 2023) in Swansea, UK (see O2 in Figure 2). In this event, particularly the Doctoral Consortium, the researcher talked to supporter Kris Luyten[6] (i.e., clarifying the positive points of the work) and with the opponent Alan Dix[7] (i.e., exploring the negative aspects with the intention of refining the direction of the work). Thanks to this participation, an article was published in the EICS 2023 Workshops Extended Proceedings.

Erik Gabriel Rodrigo da Silva, another scientific initiation scholar, also received co-supervision from October 2023 to October 2024 (see R4 in Figure 2). This initiative, blending teaching and research, involved a literature review on UX, InfoVis, and user interaction data, alongside field studies to assess software developers' acceptance of interaction data visualization tools. This collaborative effort resulted in two field studies, quantitative and qualitative analyses, the application of TAM3, linear regression analysis, and a forthcoming article in the Journal Interacting with Computers (IwC).

Dedicated research efforts intensified from January to December 2024 as a postgraduate scholarship recipient, allowing for full-time dedication to the doctoral work, particularly between January and July (see R5 in Figure 2). This period was instrumental for conducting rigorous quantitative and qualitative analyses and disseminating research findings. Contributions to the scientific community were further demonstrated through presentations at the International Conference on Enterprise Information Systems (ICEIS 2024) in Angers, France, in April 2024 (see O3 in Figure 2), and at the Brazilian Symposium on Human Factors in Computing Systems (IHC 2024) in October 2024 (see O4 in Figure 2), where an honorable mention was received for an innovative ideas paper. These engagements served as crucial platforms for research dissemination and expanding the collaborative outreach network, leading to an invitation for an extended version of the

[6] Kris Luyten is a Professor at Hasselt University and the Deputy Managing Director of the Digital Future Lab, a Flanders Make core lab. He's known for his expertise in human-computer interaction, and user interfaces.

[7] Alan Dix is the Director of the Computational Foundry at Swansea University and a highly influential figure in human-computer interaction, with extensive work on HCI theory, and pervasive computing.

work in the Journal of the Brazilian Computer Society (JBCS).

Recent engagements as a temporary lecturer at the State University of Technology of São Paulo (August 2024 to July 2025) in data structures, competitive programming, and IoT (see T3 in Figure 2), and again at the Federal Institute of São Paulo (April 2025 to July 2025) in programming (see T4 in Figure 2), further solidified the teaching and outreach proficiencies. These roles encompassed guiding students in Developer eXperience (DX), advising across diverse themes, and preparing teams for competitive programming marathons, thus broadening their professional networks and immersing them in challenging environments.

The researcher engagement in research and outreach activities was impactful. The recognition of his master's dissertation with title "Information Visualization for monitoring students in electronic learning environments" with first place in the Junia Coutinho Anacleto Award in October 2021 (see A1 in Figure 2), while not directly part of his doctoral process, served as a significant outreach validation of his expertise in part of the doctoral research themes (i.e., HCI and InfoVis), fostering deeper connections with the community.

Beyond formal academic contributions, the researcher actively engaged with the broader research community through various outreach initiatives. This included serving as a reviewer for the UFSCar Scientific Initiation Congress in 2024, demonstrating a commitment to fostering early-career researchers. Furthermore, the researcher delivered several lectures, sharing expertise on topics such as the evolution and application of the Technology Acceptance Model (TAM), distinguishing between User Experience (UX) and User Interface (UI), and elucidating the practical application of user-facing parameters in current market AIs. This dedication to knowledge dissemination was also extended through the recording of a video lesson on personas for the "STEAMulando Futuros" group.

Expertise in InfoVis was formally recognized in March 2025, when an invitation was extended to serve as a reviewer for the International Journal of Human-Computer Interaction (see A3 in Figure 2). This continuous engagement with cutting-edge research culminated in the approval of an article for the Workshop on Engineering Interactive Systems Embedding AI Technologies at EICS 2025 in Trier, Germany, in June 2025 (see O5 in Figure 2). This latest outreach endeavor showcases exploratory work on applying the core doctoral contributions with GenAI tools, outlining future research in Human-GenAI Interfaces.

The researcher's experience and studies in leveraging GenAI as a research support tool directly led to two additional significant contributions. The article, "GenAI Embedded in Activities of Academic Research: Experiences and Lessons from HCI Studies", was approved at IHC 2025 (see A4 in Figure 2). Furthermore, "HCI Research with GenAI: A Guide for Students and Researchers improving their academic work", a course focusing on the use of AI tools in academic research, was approved for INTERACT 2025 (see A5 in Figure 2). These contributions, while mentioned as part of the researcher's broader trajectory rather than being central to this thesis's dedicated chapters, underscore the researcher's continued focus on advancing the integration of GenAI within the research landscape, providing valuable insights and practical guidance for the academic community.

Finally, in July 2025, a core research contribution of the doctorate was accepted for publication in the International Journal of Human-Computer Interaction, marking a significant milestone in disseminating the work to the broader academic community (see A6 in Figure 2). Collectively, these experiences across teaching, research, and outreach have not only provided a comprehensive understanding of academic life but have also implicitly guided the evolution of the researcher. Each activity, from mentoring students and leading innovative projects to disseminating research and engaging with the global scientific community, has contributed equipping the researcher with the skills, knowledge, and network essential for a sustained career in academia.

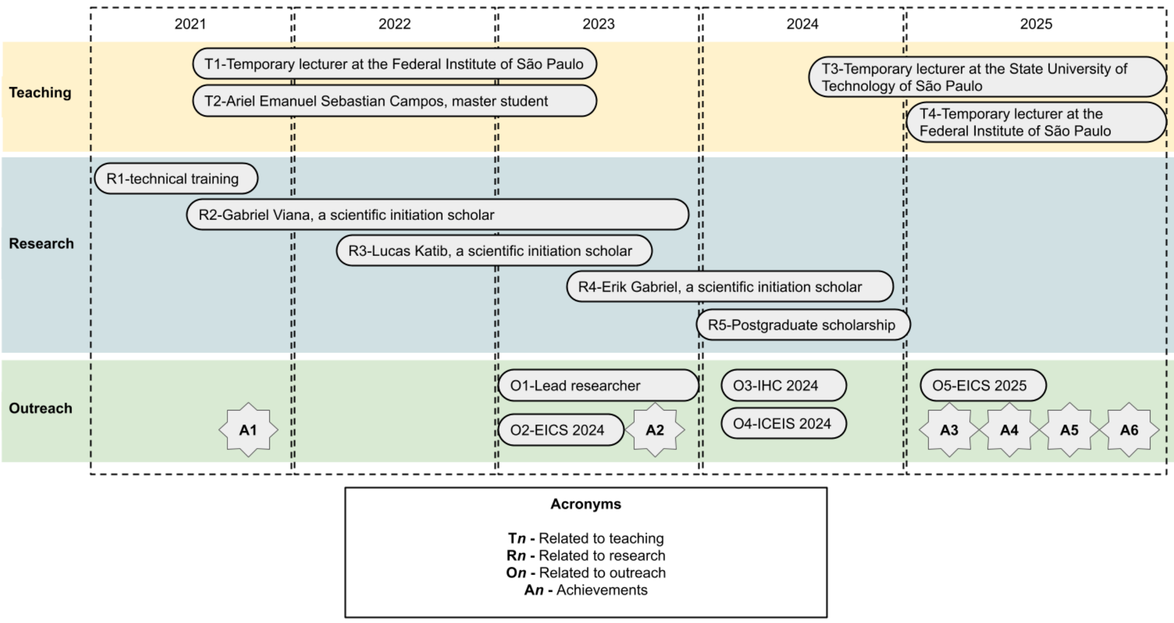


Figure 2 – Timeline of the researcher's doctoral activities across teaching, research, and outreach. Source: By the author.

1.6 Thesis organization

This work is organized into 12 chapters, with the contents organized as follows:

Chapter 1 outlines the research problem, its context, and the thesis objectives. The **objective** is to highlight challenge of extracting relevant insights from UX data (i.e., user interaction data) due to a clear lack of standardized methods for its classification, exploration, and visualization (i.e., Step A). The **contributions** include providing a foundational understanding of the research landscape and offering an overview of the methodological approach, giving a glimpse into the researcher's journey.

Chapter 2 outlines the foundational concepts for this thesis. The **objective** is to establish essential background knowledge in key areas such as UX, UX data, Information Visualization, and GenAI (i.e., Step A). The **contributions** include providing a understanding of these core concepts and presenting a review of related work, highlighting existing contributions pertinent to the challenges addressed in this thesis.

Chapter 3 outlines the research methodology employed in this doctoral work. The **objective** is to detail the DSR cycles that guided the project, provide a explanation of the artifacts developed, and describe the diverse methodologies applied in their creation and evaluation (i.e., Step A). The **contributions** include explicitly tracing the origin of the published contributions back to their respective DSR cycles, illustrating how each publication emerged from the iterative development and refinement process.

Chapter 4 presents the article "What is the State of the Art on UX Data Visualization? A Systematic Mapping of the Literature" (AMARAL; MACEDO; ZAINA, 2025). The **objective** is to thoroughly investigate the application of Information Visualization techniques to UX Data, specifically mapping chart formats, data sources, and visualization functions (i.e., Step A). The **contributions** include establishing foundational knowledge by showcasing academic accomplishments, identifying underexplored data sources and the absence of overview functions in existing visualizations, underscoring the need for further research, and serving as the initial indication of UX data's inherent complexity, which paved the way for grey literature exploration and a subsequent pivot in the research direction.

Chapter 5 presents the article "Supporting User-Centered Requirements Elicitation from Lean Personas: A UX Data Visualization-Based Approach" (MACEDO et al., 2024). The **objective** is to detail the development and evaluation of visualizations designed to explore Lean Personas data (i.e., Step B). The **contributions** include demonstrating initial attempts at applying visualizations, validating existing knowledge related to user interaction data, and simultaneously revealing the practical complexities and challenges inherent in working with unstructured UX data. The limitations and insights derived from this study strongly reinforced the necessity of a structured conceptual framework to guide the selection of visualizations and their underlying data.

Chapter 6 presents the article "Mind the Gap Between UX Data and Visualization Proposals: An Approach Emerged From the Grey Literature to Support the Analysis of User Dissatisfaction" (MACEDO; ZAINA, 2024). The **objective** is to present a more contextualized study, evaluating visualizations designed to gain insights into problems within a mobile application (i.e., Step B). The **contributions** include deepening the understanding of challenges involved in interpreting UX data through visualizations and validating the premises that eventually led to the creation of the taxonomy, especially regarding the difficulty in correlating data without a clear context or predefined structure.

Chapter 7 presents the article “Handling UX Data through visualization tools: an investigation of technology acceptance with tech-users”[8]. The **objective** is to delve into the challenges and factors influencing technology acceptance within the UX data context, specifically exploring the effectiveness and acceptance of data visualization tools from the perspective of software development professionals (i.e., Step C). The **contributions** include providing a broad context for the acceptance challenges that tools in the UX Data area might face, with its conclusions offering insights into the practical barriers to adopting artifacts designed to assist with UX data.

Chapter 8 presents the article “Beyond Raw Data: A Taxonomy for UX Data Exploration to Support the Design of Interactive Systems” (MACEDO; ZAINA, 2025). The **objective** is to describe how the taxonomy emerged as a central artifact of the research, specifically detailing the grey literature investigation into UX Data usage in practice and proposing the taxonomy based on its three legs: objectives, purposes, and UX data characteristics (i.e., Step D). The **contributions** include practically confirming the emerging conclusions from previous chapters regarding the unclear definitions and empirical reliance of UX data, and demonstrating that a visualization framework alone would be less effective without first resolving the fundamental problem of conceptual clarity. The taxonomy, therefore, becomes an approach to bridge this gap, offering a structured way to organize and navigate these definitions and practices to meet the real needs of professionals.

Chapter 9 outlines the implementation of the taxonomy as an interactive web system to make it operational. The **objective** is to demonstrate how the system enables professionals to interactively navigate and understand the taxonomy, connecting analysis objectives, purposes, and UX data types, with its online deployment ensuring accessibility from any device (i.e., Step E). The **contributions** include providing a functional tool for exploring the taxonomy and detailing its evaluation where software professionals used the system to assess potential insights derived from a given dataset.

Chapter 10 presents the article “Can GenAI Provide Insights Like a UX Professional? A comparison of Humans and GenAI rationale about UX data”[9]. The **objective** is to investigate the future perspectives of using GenAI in UX data analysis, specifically comparing the capabilities of GenAI with those of UX professionals in data analysis (i.e., Step H). The **contributions** include highlighting the continued relevance of the research in a practical application scenario for GenAI and underscoring the importance of structured knowledge bases, even for advanced AI systems.

Chapter 11 outlines the development and evaluation of a conversational interface. The **objective** is to empower professionals to explore the taxonomy and extract insights

[8] Under preparation at the time of thesis construction

[9] Article accepted to presentation at the Workshop on Engineering Interactive Systems Embedding AI Technologies (EISEAIT) on the ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS) in 2025

from UX data, even without specialized knowledge in visualization or databases. This tool, developed using a RAG approach, exclusively leverages the taxonomy as its knowledge base to answer user questions through natural language (i.e., Steps F and G). The **contributions** include significantly enhancing the taxonomy's usability by making it more accessible via AI through a DSR cycle, and presenting an evaluation by software professionals that highlights its relevance and perceived utility. The evaluation results strongly indicate that the taxonomy, along with its associated tools, can effectively assist software professionals in managing UX data.

Chapter 12 outlines the conclusions, contributions, and future work of the thesis. The **objective** is to synthesize the main contributions, reiterate the achieved objectives, discuss the DSR cycles undertaken, and elaborate on how the pivot towards the taxonomy was a decision driven by research rigor, resulting in a more impactful contribution (i.e., Step H). The **contributions** include providing a comprehensive overview of the research's accomplishments, acknowledging its limitations, and proposing avenues for future work.

Chapter 2

Fundamentals

This chapter presents the definitions, explanations, and literature on the main topics related to this thesis, namely: UX, UX data, InfoVis, and GenAI. Related works are also presented.

2.1 User eXperience (UX)

UX is a broad concept that refers to a user's entire interaction with a product or service (STANDARD; ISO, 2019). Hassenzahl (2018) provides a content-driven definition of UX organized into three levels based on user interaction: *why*, *what*, and *how*. The *why* level encompasses the motivations and needs that lead an individual to use the product. The *what* level refers to the product features that satisfy users' needs. The *how* level is about the user's activity in the product to interact with its functionalities. When the user achieves their goal (*why*) through the successful use of the concrete actions (*how*) of the product's functions (*what*), a feeling of well-being is aroused (HASSENZAHL, 2018). The Figure 3 represents the relationship between the levels proposed by Hassenzahl (2018).

UX research aims to enhance user interactions with interactive systems and optimize products (LUTHER; TIBERIUS; BREM, 2020; TONG et al., 2022). This involves analyzing user data to avoid superficial insights (CONVERTINO; FRISHBERG, 2017) and to inform the creation of successful, user-centered design solutions (KOESTEN; SIMPERL, 2021). UX professionals rely on both technical knowledge and interpretive judgment to consider user behavior, business goals, and technological constraints (HASSENZAHL, 2018; MARTINELLI; LOPES; ZAINA, 2022). Applying various techniques for UX data analysis can yield insights into user behavior and preferences, ultimately leading to enhanced user satisfaction (FRITZ; BERGER, 2015).

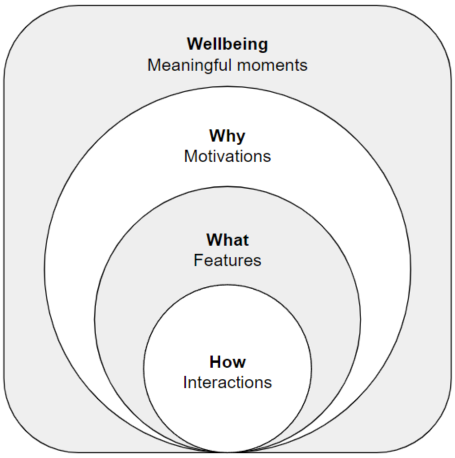


Figure 3 – Levels of the content-oriented UX model proposed by Hassenzahl (2018). Source: By the author.

2.2 UX data

UX data encompasses information mapping user interactions with interactive systems. It can be categorized as quantitative, including measurable metrics like success rates, page visits, and return rates (ALBERT; TULLIS, 2013), or qualitative, involving user-generated information such as opinions, preferences, and system experiences (RHIE; YUN, 2017).

Topics about UX and UX data are not new in software development discussions (OHASHI et al., 2018; KASHFI; FELDT; NILSSON, 2019; BAKIU; GUZMAN, 2017; YANG; CHOU; CHEN, 2020; ZAINA; SHARP; BARROCA, 2021). However, the literature shows that software teams have difficulty using usability-related information in the software development cycle (KASHFI; FELDT; NILSSON, 2019), especially without the assistance of a UX specialist (FABIJAN; OLSSON; BOSCH, 2016). The literature argues that software professionals have a clear understanding of the importance of UX-related information in designing better products (ZAINA; SHARP; BARROCA, 2021; KASHFI; FELDT; NILSSON, 2019).

While essential for enhancing user experiences and optimizing product performance (LUTHER; TIBERIUS; BREM, 2020; TONG et al., 2022), analyzing UX data is complex. Visualizations are pivotal in facilitating this interpretation (CARD; MACKINLAY; SHNEIDERMAN, 1999). By creating visualizations, UX and software professionals can overcome the abstraction inherent in raw UX data, simplifying the understanding of complex user interactions, feedback, and large-scale behavioral datasets to extract meaningful insights (WARE, 2012; BUONO et al., 2020a).

2.3 Information Visualization (InfoVis)

Using visual representations to explore human abilities to obtain information from visual pattern recognition is one of the premises of InfoVis (SOUSA; BARBOSA, 2013). InfoVis addresses the design and functionalities of visualizations, assisting in the development of visualizations aligned with user needs (MUNZNER, 2014). The abstraction promoted by visualization reduces the effort to obtain information through the perception of emergent properties of the data, such as patterns; deviations; groupings; and trends. (WARE, 2012; CARD; JACKO, 2012; SOUSA; BARBOSA, 2013).

The main resource that enables a visualization to achieve the objectives of InfoVis is the interactions made available to the user (WARE, 2012; CARD; JACKO, 2012). In conjunction with HCI, InfoVis aims to provide users with visualizations that facilitate the understanding of complex, abstract, and large datasets (WARE, 2012). Visualizations expand human perception capabilities, thereby facilitating interpretation and decision-making (CARD; MACKINLAY; SHNEIDERMAN, 1999).

Given the importance of visualization for UX data interpretation, understanding the specific visualization techniques employed in practice by professionals is essential. The literature also discusses the challenge of building visualizations from qualitative data, due to its non-numerical nature, to bring benefits to software teams (GORLENKO; ENGLEFIED, 2006; SLONE, 2009; CARD; JACKO, 2012)

2.4 Generative Artificial Intelligence (GenAI)

GenAI, particularly large language model (LLM), is transforming practices across digital product development cycles (KHALIFA; ALBADAWY, 2024). GenAI offers an emerging opportunity to support UX research, especially in tasks involving the interpretation of user interaction data like navigation logs or clickstream events (CHENG; LI; BING, 2023; HOU et al., 2024). GenAI models are now widely used, ranging from idea generation to qualitative data analysis through natural language processing (RAHMAN et al., 2023). They assist in tasks like source code generation and data analysis (HOSSEINI; HORBACH, 2023).

LLM, like ChatGPT[1] and Gemini[2], are GenAI models trained on vast text data, enabling them to comprehend, generate, and interact with human language (CHANG et al., 2024). These models predict the probability of word sequences and generate coherent text based on a specific input (BROWN et al., 2020), known as a prompt (SCHMIDT et al., 2024b). A prompt serves as the primary interface between humans and LLM (CHEN et al., 2024). Prompt Engineering is an essential practice for enhancing LLM performance

[1] https://chatgpt.com/
[2] http://gemini.google.com/

and reliability, involving the strategic design and refinement of prompts to achieve more accurate, relevant, or creative outputs (LIU; CHILTON, 2022).

LLM process language by breaking text into tokens (i.e., words, sub-words, or characters) (VASWANI et al., 2017). A set of tokens forms a context, which is what the model considers when formulating a response, encompassing the initial prompt and ongoing conversation. The quality and relevance of an LLM output depend significantly on clarity of contextual information provided by the prompt (DONG; STRATOPOULOS; WANG, 2024). Internal parameters, adjusted during training, determine a model's capability to represent complex linguistic structures (DONG; STRATOPOULOS; WANG, 2024). Temperature is a crucial parameter influencing response creativity: lower values (e.g., 0.2) yield focused and deterministic answers, while higher values (e.g., 0.8 or above) produce more creative and sometimes less predictable outputs (GILARDI; ALIZADEH; KUBLI, 2023).

To overcome limitations related to hallucination and accessing up-to-date or private information, RAG has emerged as a significant advancement in GenAI (GOOGLE, 2025). RAG combines the generative capabilities of LLMs with a retrieval mechanism that pulls relevant information from an external knowledge base (AWS, 2025). This allows the LLM to ground its responses in verified data, thereby increasing accuracy and relevance, and reducing the likelihood of generating incorrect or unsupported facts (BELCIC, 2025). This integration enables LLMs to access and synthesize information beyond their initial training data, making them more robust for tasks requiring specific, current, or proprietary knowledge (AWS, 2025; GOOGLE, 2025; BELCIC, 2025).

2.5 Related work

The literature presents various studies exploring the the UX Data, with a common focus on presenting data collected during usability tests (FRANCO et al., 2019b; MÓRO; DARÁZ; BIELIKOVÁ, 2014; DITTRICH; HOF; WIETHOFF, 2019a; BERNHAUPT et al., 2020b). Other objectives observed in the literature include visualizing user navigation through websites (BUONO et al., 2020a; RODDEN, 2014), displaying user evaluations of software (BAKIU; GUZMAN, 2017) and mobile applications (YANG; CHOU; CHEN, 2020), and reconstructing user journeys with specific products (KARAPANOS; MARTENS; HASSENZAHL, 2012a). These works collectively highlight the role of making UX data more accessible and interpretable for practitioners and researchers.

Franco et al. (2019b) developed UXmood, an interactive dashboard featuring visualizations of users' emotional states. These visualizations were generated from data (i.e., video, audio, interaction logs, sentiment, and eye-tracking) collected via sensors. The authors evaluated UXmood with 122 participants to assess the efficiency of using visualizations for communicating UX data. The results indicated that the visualizations improved the communication of UX information. This demonstrates the potential of visual interfaces to convey complex emotional and behavioral data from user interactions.

Considering website navigation data, Buono et al. (2020a) developed four visualizations to assist software teams with limited UX experience in analyzing usability data. An evaluation involving fifteen participants aimed to ascertain team satisfaction and whether the visualizations could aid in identifying usability issues. The study's findings revealed that the provided visualizations helped participants analyze data quickly and objectively, suggesting their utility in bridging the gap for teams less familiar with dedicated UX analysis methods.

(DITTRICH; HOF; WIETHOFF, 2019a) created InteracDiff, a prototype visualization tool designed to aid in the interpretation of UX data collected from user tests. The visualization displays participant responses in a predefined cyclical grid, grouping them into categories related to UX quality, such as product attractiveness and user sentiment. Twelve participants performed tasks interacting with the visualization tool during its evaluation. The results indicated that InteracDiff offered a direct approach to explore UX data; however, it demanded a high cognitive load from participants, indicating a trade-off between directness of representation and cognitive demands on the user.

To support the analysis of usability evaluation results based on user tasks, Bernhaupt et al. (2020b) developed three visualizations. These visualizations adhered to Shneiderman (1996)'s mantra, "Overview first, zoom and filter, then details-on-demand", and incorporated two levels of granularity: overview and details. A case study showed that employing visualizations is a viable approach for extracting insights from UX data, emphasizing a structured approach to data exploration.

Karapanos, Martens e Hassenzahl (2012a) designed the iScale tool to assist UX researchers in retrospective user studies. The tool presented two versions of a timeline chart, differing in how chronological data was loaded into the visualization (i.e., either automatically by the tool or via user interaction). The effectiveness of both tool versions in eliciting user experience accounts with a product was evaluated by 48 participants. The results indicated that using visualizations to organize collected data aided participants in analyzing their perceived user experience, reinforcing the value of understanding longitudinal user data.

2.6 Final considerations

This chapter clarified the core concepts tangential to this thesis: UX, UX data, InfoVis, and GenAI. We also presented related works, illustrating scientific efforts in exploring UX data. While existing research addresses the visualization of UX Data, most studies focus on developing specific tools for particular contexts. They often lack adaptable of their applicability in broader scenarios (FRANCO et al., 2019b; MÓRO; DARÁZ; BIELIKOVÁ, 2014; DITTRICH; HOF; WIETHOFF, 2019a; BERNHAUPT et al., 2020b; BUONO et al., 2020a; RODDEN, 2014). Instead of directly constructing a data exploration tool, this thesis establishes a knowledge base designed to equip software practitioners with a understanding of the available possibilities for UX data analysis.

Chapter 3

Methodology

This chapter details the research methodology underpinning this doctoral thesis. The primary objective is to describe the iterative cycles of DSR that guided the entire project, from initial conception to final contributions. Each step of the general methodology has a cyclical nature (i.e., DSR feature). A key contribution of this chapter is to explicitly connect the published articles of this thesis back to their specific DSR cycles, illustrating the iterative process through which each step output was conceived, developed, and validated.

3.1 Introduction

The development of this doctoral research was based on the three DSR cycles defined by Hevner (2007). DSR is an approach to direct the development and evaluation of artifacts built to meet business needs identified in specific contexts (VENABLE, 2006; BASKERVILLE, 2008). DSR has been widely used in the computer science field (IIVARI; VENABLE, 2009). Among the possible artifacts resulting from a DSR process, highlights the creation of algorithms, prototypes, requirements, tools, frameworks, taxonomies, or models (OFFERMANN et al., 2010).

Relevance, **Rigor**, and **Design** are the DSR cycles defined by Hevner (2007). In the **Relevance** cycle, the researcher investigates the problem to include the application domain in the project scope, ensuring that the developed artifact is relevant to real needs. The **Rigor** cycle is responsible for exploring resources to provide knowledge about the project scope, ensuring that the developed artifact is innovative in relation to the state of the art. Finally, the **Design** cycle involves the construction and evaluation of artifacts, considering all the information captured by the **relevance** and **rigor** cycles.

The organization of this doctoral research using DSR cycles enabled an adaptive approach by combining emerging information from the literature (i.e., scientific and grey) with findings from practical activities on the use of UX data and data visualizations. The continuous data collection throughout the DSR cycles was crucial for the iterative maturation of the proposal. This process allowed the solution to be progressively adapted and enhanced, guided by emerging insights from the practical needs of software professionals. In this doctoral research, the **Relevance** and **Rigor** cycles were always carried out together to provide evidence and guide analysis processes based on data collected in grey literature (i.e., relevance), scientific literature (i.e., rigor) and investigations with the target audience (i.e., relevance).

While the initial trajectory of this research was set towards visualization techniques for UX data, the adaptive nature of DSR proved crucial in pivot the direction of this thesis. Through iterative engagement with design and evaluation cycles, a more fundamental and pressing need emerged: the critical lack of clarity in the definitions and consistent use of UX Data within the field. This insight led to a strategic pivot, shifting the primary focus from visualization to the development of a taxonomy for UX Data. It is worth noting that the new research path followed was not a deviation. It was a feature of the adaptability of DSR, demonstrating its capacity to respond to emergent research needs. Figure 4 provides an overview of DSR cycles and the features each cycle interacts with.

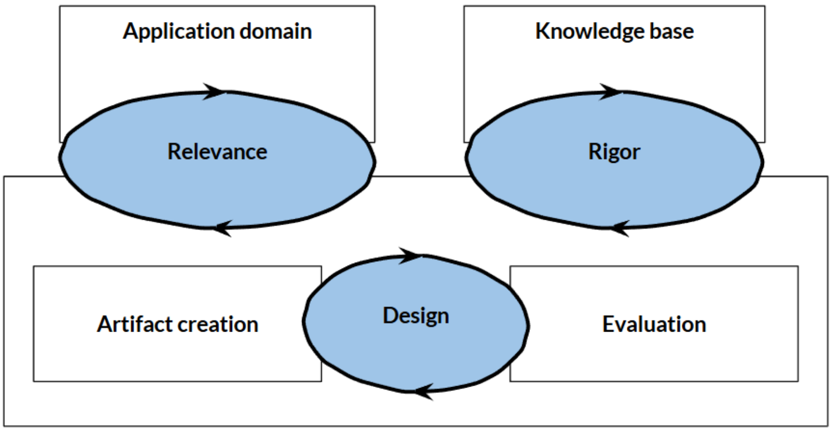


Figure 4 – Representation of DSR cycles according to Hevner (2007). Source: By the author.

3.2 DSR in action

Although the DSR does not follow a linear path in this doctoral project, this section presents the DSR in sequential steps. The sequential explanation highlight each step building upon the insights and artifacts generated in the previous one, ultimately leading to the development and validation of a UX data taxonomy. The research's development is first explained through the lens of the executed DSR cycles (numbered in Figure 5). Subsequently, the distinct steps undertaken during their execution (indicated by the grey rectangles in Figure 5) are detailed.

Our **design** cycles were focused on developing and evaluating artifacts, both conceptual (e.g., relationship of the taxonomy data) and tangible (e.g., web interfaces), to address the complexities of UX data. Considering that grey literature brings together the "state of practice" (i.e., related to the **relevance** cycle) while scientific literature represents the state of the art (i.e., related to the **rigor** cycle), it was defined that the **rigor** and **relevance** cycles would be unified so that the information from both sources would be complementary. In total, six complete DSR cycles (see the numbers in Figure 5) were performed:

Cycle 1: Problem Identification and Motivation. This initial cycle focused on understanding the problem domain and establishing the research's relevance by identifying critical gaps and justifying the need for a new artifact. **Rigor and Relevance** were addressed through a SML (detailed in Chapter 4) that scientifically mapped the state of the art in UX data visualization, highlighting limitations, and a GLR (addressed in Chapter 8) that investigated the practical use of UX data in industry, revealing critical ambiguities. The **Design** aspect of this cycle involved structuring these investigative activities, including defining search protocols and analysis frameworks, to effectively map the problem space. As a key **Artifact**, this cycle yielded the identification of a significant lack of conceptual clarity and structured approaches for exploring and utilizing UX data, a core insight that framed the subsequent DSR cycles by highlighting how this absence hindered effective visualization and analysis.

Cycle 2: Initial Exploration and Iterative Refinement of Visualization Concepts. It involved early exploratory studies with visualizations, which subtly reinforced the need for the structured understanding provided by the taxonomy. **Rigor and Relevance** were addressed by conducting two exploratory studies to test initial visualization concepts. An initial study explored visualizations for Lean Personas data (detailed in Chapter 5), providing feedback on early visualization attempts. Subsequently, a more contextualized study evaluated visualizations for analyzing user dissatisfaction in a mobile application context (covered in Chapter 6). The **Design** aspect focused on creating the preliminary visualization concepts and setting up the empirical studies to gather user feedback. As a key **Artifact**, insights were gained into the practical challenges of effectively visualizing unstructured UX data, further validating the underlying problem that the taxonomy was designed to address; specifically, the difficulties encountered by participants strongly reinforced the necessity of clear data definitions.

Cycle 3: Development of the Core Artifact. It focused on the conceptualization and initial design of the thesis's primary artifact, driven by insights from the problem identification stages (Cycles 1 and 2 in Figure 5). **Rigor and Relevance** were maintained as we strategically pivoted, recognizing the critical need for a fundamental conceptual artifact: the UX data taxonomy, before developing the initially proposed comprehensive visualization framework. This adaptation showcases DSR's capacity to adjust based on emergent knowledge, ensuring the research addressed the most pressing need. The **Design** aspect involved structuring the taxonomy's categories, definitions, and relationships based on the ambiguities identified earlier (presented in Chapter 8). As a key **Artifact**, this cycle culminated in the creation of the UX data taxonomy, specifically aimed at bringing conceptual clarity and structure to the ambiguous domain of UX data.

Cycle 4: First Implementation and Evaluation of the Taxonomy (Web version). This cycle focused on translating the conceptual taxonomy into a tangible, usable tool, thereby demonstrating its practical applicability. **Rigor and Relevance** were addressed through the empirical evaluation of the UX data taxonomy's content with software professionals, directly assessing its perceived relevance and usefulness in managing UX data (detailed in Chapter 9). The **Design** aspect involved implementing an interactive web system that allowed professionals to effectively navigate and understand the taxonomy. As a key **Artifact**, this cycle yielded an interactive web system showcasing the taxonomy, alongside empirical evidence validating the UX data taxonomy's ability to help professionals manage UX data more effectively, confirming its practical relevance and acceptance.

Cycle 5: Second Implementation and Evaluation of the Taxonomy (GenAI version). It involved systematically evaluating the main artifact (i.e, the taxonomy) within a novel context of use. **Rigor and Relevance** were addressed through a study directly measuring the UX data taxonomy's ability to help professionals manage UX data more effectively, thereby confirming its practical relevance and acceptance (detailed in Chapter 11). This cycle also included a technology acceptance study that illuminated the real-world environment for artifact adoption. For **Design**, we developed a conversational AI interface RAG, which utilizes the taxonomy as its exclusive knowledge base to answer natural language queries about UX data. As a key **Artifact**, this cycle produced a GenAI-powered conversational interface demonstrating the taxonomy's utility, along with empirical evidence of its practical value and insights into its real-world adoptability.

Cycle 6: Broadening Perspectives and Future Directions. This final cycle encompassed looking beyond the immediate solution, exploring new frontiers, and synthesizing the overall impact of the research. **Rigor and Relevance** were maintained as we explored Generative AI's capabilities in analyzing UX data, comparing its insights with human UX professionals (detailed in Chapter 10), demonstrating how evolving technologies intersect with the need for structured UX knowledge. The entire research journey, its contributions, limitations, and future directions were then consolidated (see Chapter 12), thereby updating the domain knowledge and completing the final iteration of the DSR's rigor and relevance cycle. The **Design** aspect involved structuring the comparative study of AI and human analysis, along with the framework for synthesizing the thesis's overall impact and charting future research avenues. As a key **Artifact**, this cycle produced a comprehensive overview of the research contribution to the field, acknowledging new technological advancements and setting the agenda for future investigations.

Figure 5 presents an overview of the steps of the doctoral project methodology and its relationships with the DSR cycles.

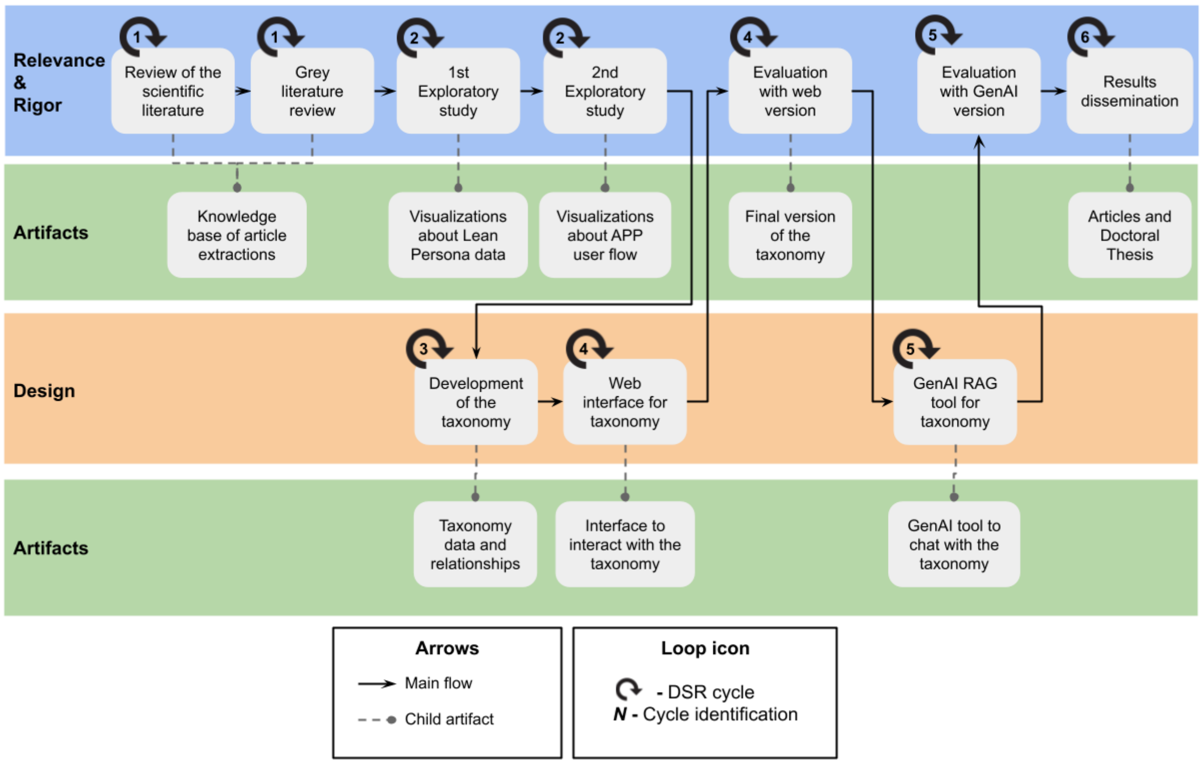


Figure 5 – Steps of development of doctoral research. The two artifact sections serve only to organize the figure. Source: By the author.

Below is a description of the conduct steps and their contribution of this doctoral research:

- **Review of the scientific literature:** The objective of this step was to investigate the work related to the application of InfoVis techniques to obtain relevant information from UX data. The investigation was carried out following the research protocol of a SML (PETERSEN et al., 2008). **Contribution**: The formal definitions found for UX data and the works related to its use. It was possible to determine the application domain and the gaps in the scientific knowledge base.

- **Grey literature review:** The research and data collection carried out in this step were carried out in accordance with the guidelines for the execution of a GLR (GAROUSI; FELDERER; MäNTYLä, 2019). The objective of the research was to investigate the use of UX data within the software development cycle. **Contribution:** Based on the information gathered in this step, it was possible to determine the "state of the practice" and extract information about which UX data are considered relevant and which practices and tools are used to analyze them.

- **1st Exploratory study:** The objective was to validate the knowledge acquired in the first DSR cycle using Lean Personas data. Visualizations were developed that allowed navigation through the quantitative data from the Lean Personas to facilitate the selection of relevant qualitative data. The visualizations were submitted to a developer evaluation to verify their acceptance and suitability for use in requirements gathering. **Contribution:** This study contributed to analyzing software professionals feedback on visualizations developed based on the extractions from the grey literature (first DSR cycle). The results shed light on the usefulness of the visualizations used, the necessary conditions for leveraging the visualization, and the suitability of Shneiderman (1996)'s mantra for navigating the data.

- **2nd Exploratory study:** The 1st Exploratory study received some feedback indicating that it would be appropriate to provide more information about the context in which visualizations are used. Therefore, for 2nd Exploratory study, a complete scenario was developed for a mobile application for airline ticket sales. The scenario included the application prototype, the company's objectives and business rules, and reports of dissatisfaction (extracted from grey literature) with the application's performance from the management team. Based on these reports of dissatisfaction, five visualizations were developed to analyze the mobile application's usage flow. The objective was to verify the use of visualizations to gain insights into the application's problems. **Contribution:** This exploratory study established the applicability and broader potential of the data extracted from grey literature for create UX data visualization for software professionals, demonstrating its capacity to deepen user understanding, enhance self-knowledge, and optimize development resources.

- **Development of the taxonomy:** Building upon the insights from DSR cycles 1 and 2 we analyze and collected lessons learned from the grey literature, which encompassed formal definitions of UX data, the objectives that could be achieved through its analysis, and the purposes it could serve within the software development lifecycle. A formal grammar and a relational database were subsequently employed to establish clear relationships between these data, enabling a structured navigation from a specific objective to its related purposes and the necessary UX data. **Contribution:** This development provided a structured way for understanding and classifying UX data, directly addressing the core problem of conceptual ambiguity. It offers a navigable structure for professionals better identify relevant data, understand its purpose, and apply UX data exploration, thus advancing the theoretical and practical understanding of UX data in interactive systems.

- **Web interface for taxonomy:** Recognizing that the developed UX data taxonomy, while conceptually robust, required practical accessibility for software professionals, an interactive web system was designed and implemented to operationalize it. This system was crafted to allow users to intuitively navigate and comprehend the taxonomy's intricate relationships, linking analysis objectives, their corresponding purposes, and the specific types of UX data required. Built as an online platform, it ensures broad accessibility from any device, translating the formal grammar and relational database that define the taxonomy into a user-friendly interface. **Contribution:** This development provided a tangible and interactive artifact that directly demonstrates the practical application of the conceptual taxonomy, making its structured knowledge readily available and usable for practitioners, thereby significantly enhancing its utility and potential for real-world impact.

- **Evaluation with web version:** To assess the practical utility and effectiveness of the operationalized UX data taxonomy, an empirical evaluation was conducted utilizing its interactive web interface. This study aimed to determine whether access to the taxonomy via the web system (i.e., UX Data Wiki) would enhance the accuracy with which software professionals could identify achievable UX objectives from a given dataset and concurrently reduce the perceived cognitive workload of this task. Through a comparative study design, participants navigated a clickstream dataset, either with or without the aid of the UX Data Wiki, to identify relevant objectives. **Contribution:** This evaluation provided empirical evidence of the web system's efficacy in facilitating a more precise and less burdensome analysis process, directly validating the practical benefits of operationalizing the conceptual taxonomy for real-world application by practitioners.

- **GenAI RAG tool for taxonomy:** To further democratize access to the insights embedded within the UX data taxonomy, a conversational interface powered by GenAI and a RAG approach was developed. This tool was specifically designed to enable software professionals to consult the taxonomy and obtain insights about UX data through natural language queries, eliminating the need for deep knowledge of data visualization principles or database structures. By exclusively utilizing the taxonomy as its knowledge base, the RAG-based system ensures that responses are grounded in the structured information provided by the research. **Contribution:** This development represents a significant innovation in making the taxonomy accessible, leveraging AI to provide an intuitive means of exploring UX data concepts, thereby greatly enhancing the taxonomy's usability and potential for practical application.

- **Evaluation with GenAI version:** The evaluation was structured into two distinct studies. The first, a pilot evaluation, compared the UX Data Wiki AI performance against general-purpose LLMs in providing relevant and clear answers to UX-related questions. The second, a evaluation involving 23 software professionals assessed the impact of the taxonomy on their ability to explore real UX datasets and derive design insights, captured through standardized questionnaires and qualitative feedback. **Contribution:** This development represents a innovation in making the taxonomy accessible, leveraging AI to provide an interaction way using natural language to exploring UX data concepts, thereby enhancing the taxonomy's usability and potential for practical application.

- **Results dissemination:** As a final and crucial component of this doctoral research, efforts were dedicated to broadly disseminating the findings and contributions. This involved actively sharing the knowledge generated, including the developed UX data taxonomy, its supporting tools (i.e., Web and GenAI), and all empirical insights, with both the scientific community and industry practitioners. Activities encompassed publishing scientific articles (those integrated into this thesis and others later), presenting at conferences, delivering lectures, and creating supplementary materials like posters or guides. **Contribution:** This ongoing process of dissemination is vital for ensuring the research's impact, fostering knowledge transfer, and updating both the academic literature and the application domain about the possibilities and limitations of working with UX data.

3.3 Final considerations

This chapter presented the general methodology for conducting this doctoral research and its relationship to the DSR cycles. Details were also provided on the approaches used within each stage, as each stage has its own systematic process. Each chapter of this thesis specifies the steps followed to achieve the presented result. This chapter briefly presents how each chapter relates to the development of the proposal.

Chapter 4

Review of the scientific literature

This chapter presents the article "What is the State of the Art on UX Data Visualization? A Systematic Mapping of the Literature" (AMARAL; MACEDO; ZAINA, 2025), authored by Lucas Katib, Maylon Macedo, and Luciana Zaina. Published in the Journal on Interactive Systems (JIS) in 2025. This article serves as the initial step (i.e., Step A) within the overall thesis methodology. Its primary objective is to thoroughly investigate the application of Information Visualization techniques to UX data, specifically by mapping chart formats, data sources, and visualization functions. The contributions of this work include establishing foundational knowledge by showcasing academic accomplishments, identifying underexplored data sources, and highlighting the absence of overview functions in existing visualizations. This chapter underscores the need for further research in the field and provides the initial indication of UX data's inherent complexity, which consequently informed the subsequent exploration of grey literature and a shift in the research direction.

4.1 Introduction

User experience (UX - *User eXperience*) arises from the interaction between the user's internal state, the software, and the context of use (HASSENZAHL; TRACTINSKY, 2006). The practices associated with UX go beyond establishing a product design. They also involve an experiential perspective of the users. This perspective arises from the combination of user needs and emotions before, during, and after the interaction (GIBBONS, 2020).

The literature defines data resulting from user interaction with a product as UX data (ALBERT; TULLIS, 2022). UX data can be quantitative or qualitative. Its sources include formal and informal collection, such as spontaneous conversations with end users (FABIJAN; OLSSON; BOSCH, 2016; RIVERO; CONTE, 2017; KIEFFER. et al., 2019). This data can serve various purposes, from validating initial product ideas during the pre-development phases (HOKKANEN; KUUSINEN; VÄÄNÄNEN, 2016) to assessing the overall UX post-development (FABIJAN; OLSSON; BOSCH, 2016).

Data obtained during different moments of use shed light on how the UX evolves in the long term. For instance, it can help explain how the UX reflects the learning curve of users (KUJALA et al., 2011). Furthermore, UX data contribute to understanding the relationship between pragmatic and hedonic aspects of UX, such as usability and user needs (MOELLENDORFF; HASSENZAHL; PLATZ, 2006).

UX data can inform company decisions about improving a product or developing a new one. The literature advocates for more UX data visibility to engage a team with the UX design (ZAINA; SHARP; BARROCA, 2021; KASHFI; FELDT; NILSSON, 2019). However, UX data is rarely explored to improve and create software (KASHFI; FELDT; NILSSON, 2019).

Data visualizations can facilitate the interpretation of UX data. Munzner (2014) defines data visualizations as visual representations of data that assist users in carrying out tasks. The author proposes a visualization analysis framework with four nested levels: domain, which refers to the context of use and the target audience; abstraction, which encompasses the data and task abstraction; idiom, which defines the visual encoding and interaction idioms; algorithm, which refers to the computational implementation.

While there are existing papers about UX data visualizations, the literature on this subject is sparse. The lack of an overview of the literature limits a precise understanding of its gaps and potential for further investigations. This paper aims to answer the research question: *What is the state of the art on UX data visualization?* The results of this investigation provide Human-Computer Interaction (HCI) researchers and practitioners with a broad overview of categories, trends, and gaps related to UX data visualizations — covering purposes, data sources, chart formats, interaction functionalities, technologies, and more. These findings may inspire further investigations into UX data visualization from various perspectives, as our review addresses multiple aspects of the literature. To the best of our knowledge, no other literature review has examined UX data visualization through as many diverse lenses as ours.

To achieve the paper's goal, we conducted a Systematic Mapping of the Literature (SML) that selected 57 papers. To support the consolidation of the findings, we derived ten guiding questions from the four nested levels of Munzner (2014)'s framework while also considering the Visual Information Seeking Mantra of Shneiderman (1996). For the qualitative analysis of the SML, we adopted the open coding technique (SALDAÑA, 2021). The papers reported successful cases of visualizations that helped users analyse software from a UX point of view. The investigation revealed several chart formats as well as the data sources combined with each one of them. We also produced codes for the research methods, purposes, technologies, and interaction functions related to UX data visualizations.

This paper is structured as follows: Section 4.2 describes the related work; the process followed for the SML is detailed in Section 4.3; Section 4.4 presents the findings of the coding stage organized according to the ten guiding questions; the results are discussed in Section 4.5 and compared with the framework of Munzner (2014) in Section 4.6; finally, the limitations of the SML are presented in Section 4.7 and the conclusion in Section 4.8.

4.2 Related Work

The papers related to UX data visualization usually focus on particular points about users and propose the adoption of well-known chart formats. Considering charts for timeline visualization, Karapanos, Martens e Hassenzahl (2012b) built a line chart to show the evolution of the user's perceived experience. Alternatively, Da Silva Franco et al. (2019) displayed user sentiments through time in a Gantt chart and animated scatterplot. Combined with user sentiment, Da Silva Franco et al. (2019) also used eye-tracking data in the scatterplot and in a scanpath. On the other hand, Móro, Daráž e Bieliková (2014) employed tables and bar charts to display eye-tracking data.

Büschel, Lehmann e Dachselt (2021) and Kepplinger et al. (2020) featured custom chart formats. Through a mixed reality toolkit, Büschel, Lehmann e Dachselt (2021) contextualized user interaction and movement in its original environment. The authors provided scatterplots, heatmaps, and point plots in the toolkit. In addition, it contained a 3D tube format to encode the movement of devices. In the context of Games User Research, Kepplinger et al. (2020) developed a visualization of triangulated data. The data consisted of user sentiment, gaze information, and character movement. In this visualization, a trajectory map encoded movement, outlines of in-game objects depicted gaze information, and particle clouds represented user sentiment.

Hussain et al. (2018) also made use of triangulation with data from audio devices, video equipment, biometric devices, surveys, and user interaction logs. Influenced by Semiotic Engineering, Ferreira et al. (2017) presented an approach focused on user interaction logs to guide UX design. The approach consisted of exploring data collected at the strategic, tactical, and operational levels. To communicate the UX of a product between interdisciplinary teams, Lachner et al. (2016a) created a radar plot that displays nine UX dimensions of a product. These dimensions measured the impact of the UX of a product in distinct areas of a company.

There are studies that investigate the literature on eye tracking data and sentiment visualization. However, these studies do not focus on visualizations derived from data sources that describe user interactions with software applications. For instance, Blascheck et al. (2017) classified 110 papers on eye-tracking data visualization, categorizing them based on multiple criteria about data sources and chart formats. The authors further segmented the visualizations according to their suitability for point-based or AOI-based analysis. Similarly, in a survey of 132 papers on sentiment visualization, Kucher, Paradis e Kerren (2018) analyzed visualizations across several categories, including data domain, data source, analytic tasks, visualization tasks, visual encoding of variables, and chart formats. To explore the uncertainty in sentiment and stance visualizations, Ramalho, Jorge e Gama (2023) conducted a survey examining 35 papers. This study classified the visualizations based on evaluation criteria, purpose, data source, data encoding, chart format, and the representation of uncertainty. In a Systematic Literature Review focused on visualizations in opinion mining systems, Shamim, Balakrishnan e Tahir (2014) identified visualizations that encode the polarity of reviews. However, these visualizations were primarily targeted at domains such as phones, vehicles, printers, and books, rather than software applications.

Some secondary studies have explored specific scopes of the literature on UX data visualization. For example, Davila, Paz e Moquillaza (2023) performed a Systematic Literature Review to identify purposes, data sources, and challenges on heatmap visualizations of user interaction data captured during usability tests. In addition, Wallner e Kriglstein (2013) conducted a literature review on gameplay data visualization. They focus on visualizations for game development purposes. From their findings, the visualizations were categorized according to target audience (player or game developer), field of application (specific game or across games or genres), data sources, and chart formats.

Although papers exploring UX data visualization are present in the literature, its knowledge is still sparse. Therefore, it is difficult to form a comprehensive assessment of its gaps and potential. In this context, this paper aims to investigate and synthesize knowledge about the domain, abstraction, idiom, and algorithms of UX data visualizations in line with Munzner (2014)'s framework.

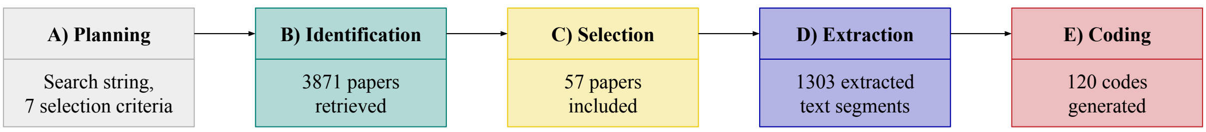


Figure 6 – Stages of the SML.

4.3 Methodology

Systematic Mapping of the Literature (SML) is a method for investigating primary studies that provides an overview of the literature within a research field (KITCHENHAM; CHARTERS, 2007; PETERSEN et al., 2008). SMLs aim to offer a high-level classification scheme for a research field rather than an in-depth analysis, typical of a Systematic Literature Review (SLR) (PETERSEN et al., 2008). Thus, an SML is more appropriate than an SLR to establish evidence and structure research literature that is still underexplored (KITCHENHAM; CHARTERS, 2007; PETERSEN et al., 2008). To achieve a comprehensive overview that supports researchers and practitioners in managing UX data visualizations, this paper adopts the SML process outlined by Petersen et al. (2008). Figure 6 illustrates the process with the outcomes of each stage, as described in the following subsections. All authors participated in Stage A. Under the supervision of the second and third authors, who have experience in SMLs, the first author was responsible for Stages B to E.

4.3.1 Stage A - Planning

As a first step, we elaborated the search string. Through an iterative process, three researchers refined the search string by including, excluding, and testing different combinations of keywords. For each test, the three researchers independently analyzed the metadata of distinct random samples of retrieved papers. Based on words found in seminal works of the UX data visualization literature (DITTRICH; HOF; WIETHOFF, 2019b; BÜSCHEL; LEHMANN; DACHSELT, 2021; LACHNER et al., 2016a; KEPPLINGER et al., 2020), the first tested string was (“information visualization” OR “information visualisation” OR “infovis”) AND (“user experience” OR “UX” OR “UX data” OR “UX-data” OR “UX data driven” OR “UX data-driven”). Due to the limited number of retrieved papers, the researchers tested the inclusion of three new keywords separately: “emotion”, “sentiment”, and “sentiment analysis”. The metadata analysis revealed that the new keywords retrieved no additional relevant papers. Subsequently, by excluding the keywords “information visualisation”, “UX-data”, and “UX data driven”, the researchers discovered that these terms did not influence the relevance of the search results. Following this insight, the researchers modified the original search string by excluding these three keywords and testing the inclusion of “graph” and “UX measure”. They found that “UX measure” improved the relevance of the retrieved papers while “graph” significantly increased the number of papers unrelated to UX data visualization.

After these refinements, we defined the SML’s search string as ***(“user experience” OR “UX” OR “UX data” OR “UX measure”) AND (“information visualization” OR “InfoVis”)***. This string combines the terms “UX” and “user experience” from the User Experience field (HASSENZAHL, 2010) with the terms “information visualization” and “InfoVis” from the Information Visualization field (MUNZNER, 2014). Furthermore, we included the terms “UX data” and “UX measure”. The term “UX data” denotes the data generated during the user interaction with a software application (ALBERT; TULLIS, 2022), and “UX measure” refers to the measured characteristics of the interaction (HARTSON; PYLA, 2018). Except for “InfoVis”, these terms were drawn from seminal works of the UX data visualization literature (DITTRICH; HOF; WIETHOFF, 2019b; BÜSCHEL; LEHMANN; DACHSELT, 2021; LACHNER et al., 2016a; KEPPLINGER et al., 2020). Despite this, the term "InfoVis" was kept as the abbreviation of "information visualization".

Drawing from the digital libraries recommended by Buchinger, Cavalcanti e Hounsell (2014) for scientific research, we applied the search string in four digital libraries of relevance to Software Engineering (KITCHENHAM; CHARTERS, 2007): *ACM Digital Library*, *IEEE Digital Library*, *ScienceDirect*, and *Scopus*. From the search results, we selected papers that addressed all inclusion criteria and did not match any exclusion criteria. The exclusion criteria were: (CE1) not published in a journal or conference proceedings, (CE2) have up to four pages, (CE3) be a secondary study, (CE4) be a duplicate paper, and (CE5) be unavailable in English. As described by their motivations, we adopted the following inclusion criteria:

- (CI1) acts on the user interaction with a software application through data visualization.
  - Motivation: to ensure that the visualization presented contributes to improving the user interaction with a software application. This contribution may arise from users analyzing their own interaction or a development team analyzing the user interaction.
- (CI2) focuses on visualizing data related to user interaction with a software application.
  - Motivation: to ensure that the paper presents a visualization with data about UX as its source. For instance, the selection should exclude papers proposing business data or software codebase visualization.

4.3.2 Stage B - Identification

In January 2023, we applied the search string across all fields in each digital library. Table 1 shows the number of papers returned from each digital library. In total, we retrieved 3871 papers. Instead of using the search string builder provided by each database, we applied the same search string across all databases using the simple search function available in each. Furthermore, we did not use automatic filters from the digital libraries to reduce the sample of papers.

4.3.3 Stage C - Selection

First, we applied the exclusion criteria based on the papers' metadata made available by the digital libraries. After that, we considered the inclusion criteria based on their

	Returned	CE1 to CE5	Accepted (CI1 and CI2)
ACM Digital Library	1017	612	11
IEEE Digital Library	41	18	1
ScienceDirect	466	307	2
Scopus	2347	1829	42
Total	**3871**	**2766**	**57**

Table 1 – Number of papers returned and accepted per search base after applying the selection criteria. The search did not return the seminal work of Lachner et al. (2016a), but we included it in the SML as justified in Subsection 4.3.3.

titles and abstracts. Finally, we revised the selection after reading the remaining papers. Table 1 details the number of selected papers in each step.

The seminal work of Lachner et al. (2016a) describes its proposed visualization as a UX tool and focuses on presenting the characteristics of the displayed data rather than the developed visualization. Thus, the paper does not mention terms that characterize the Information Visualization field, such as "information visualization" or "InfoVis". As a consequence, the execution of the search string did not return it. However, we included it in the SML as it fits the selection criteria. As a result, we accepted 57 papers. Appendix A identifies the selected papers from P01 to P57.

4.3.4 Stage D - Extraction

To form a comprehensive overview of the literature, we derived ten questions from the analysis framework of Munzner (2014) to guide Stages D and E (see Figure 6). The framework divides visualization analysis into four nested levels: domain, abstraction, idiom, and algorithm. These questions cover the four levels of the framework, as discerned in Table 2. For each selected paper, we extracted text segments relevant to the guiding questions and linked each segment to its corresponding source and questions. All extractions are centralized on an electronic spreadsheet page.

Considering the idiom level questions, questions Q7 to Q9 refer to the Visual Information Seeking Mantra: "Overview first, zoom and filter, then details on demand" (SHNEIDERMAN, 1996). The mantra structures the user navigation flow and guides design decisions concerning visualization interaction functions. According to the mantra, the user should initially receive an overview of the data. Then, the visualization should progressively reveal details in response to user interactions. According to Wall et al. (2019) and Stasko (2014), the value of a visualization is based on its ability to reveal the essence of the data through an overview. The usability benefits of combining the overview and focused views outweigh their costs, for example, by increasing the user task completion rate (COCKBURN; KARLSON; BEDERSON, 2009). In addition, overviews reduce the cognitive effort for interpretation and aid user navigation through the data space (HORNBÆK; HERTZUM, 2011). The Human-Data Interaction Design Guidelines of Victorelli e Reis (2020) recommend reducing the information density and gradually revealing the data.

Level	ID	Question
Domain	Q1	What are the described objectives of the papers?
	Q2	What research methods were used, and how many participants were included?
	Q3	What were the achieved results in evaluations?
Abstraction	Q4	What are the purposes of use of the visualizations?
	Q5	What data sources were explored by the visualizations?
Idiom	Q6	What chart formats were used?
	Q7	How was the function of “having an overview of the data” included in the visualizations?
	Q8	How was the function of “applying zoom and filters to the displayed data” included in the visualizations?
	Q9	How was the function of “viewing the data at the most detailed level” included in the visualizations?
Algorithm	Q10	What technologies were used to develop the visualizations?

Table 2 – Guiding questions of the SML’s extraction and coding stages, divided according to the analysis framework of Munzner (2014).

4.3.5 Stage E - Coding

To consolidate the findings, we used the open coding technique (SALDAÑA, 2021). We examined and compared the extracted text segments from Stage D (see Figure 6) to find patterns within the data. Based on the identified similarities and differences, we grouped the text segments and assigned a code to each group. Each code encapsulated the essence and meaning behind each group. As the analysis progressed, the codes were refined and merged.

We documented the coding for each guiding question in a separate electronic spreadsheet page. Each page contains the corresponding codes generated and the associations between extractions and codes. Figure 7 shows two examples of extractions with their sources and codes.

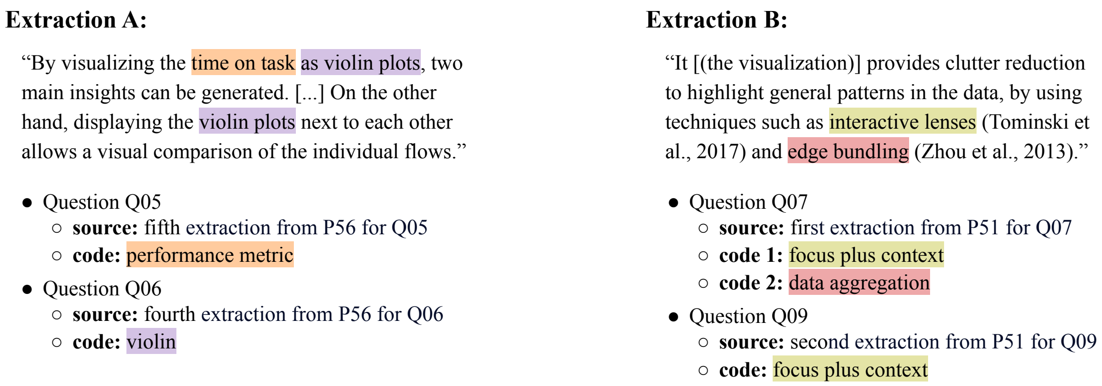


Figure 7 – Extraction examples with correlated codes.

4.4 Results

Figure 8 – Number of papers by source.

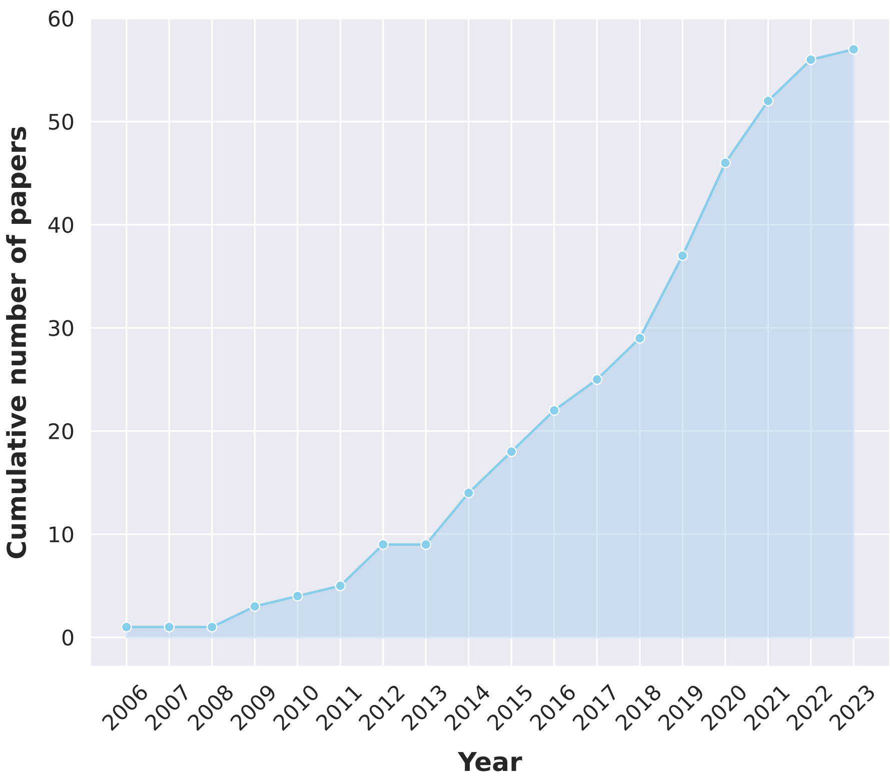


Figure 9 – Cumulative number of papers by year.

Figures 8 and 10 show the number of selected papers by source and author, respectively. Meanwhile, Figure 9 displays the cumulative number of papers over the years.

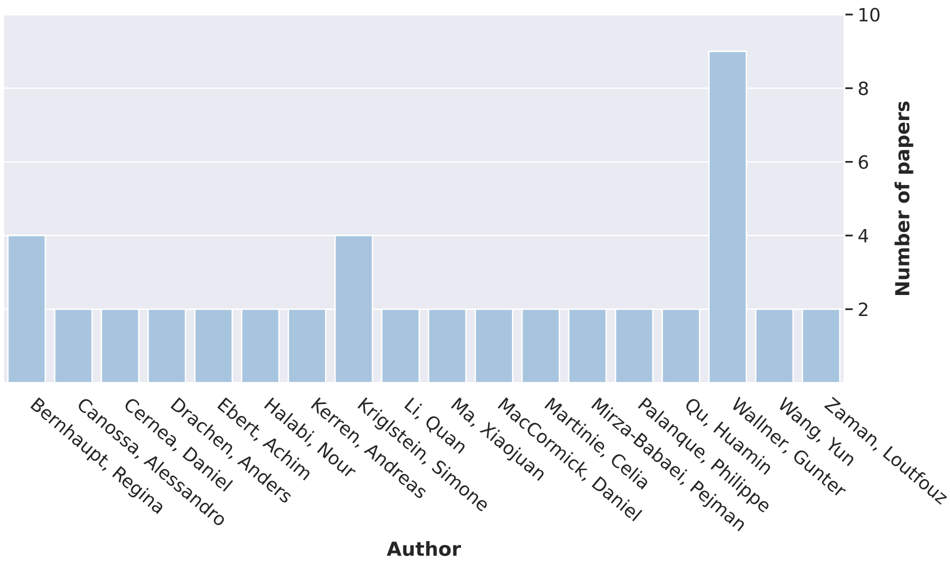


Figure 10 – Number of papers by author with at least two papers.

Considering the SML's selection, the first publication on UX data visualization was in 2006. Since 2013, the number of publications has quintupled. The papers are distributed across 39 sources, including conferences, journals, and symposiums, with the Conference on Human Factors in Computing Systems (CHI) having the highest number of papers. Among 177 authors, 18 have published more than two papers. In the following subsections, we organized the results of the SML in response to the predefined questions and highlighted the generated codes in *italics*.

4.4.1 What are the described objectives of the papers?

Most (32) of the papers aim to *assist in the UX evaluation of a software application*. Some of these papers focused on displaying attitudinal data, such as user feedback [P06, P19, P24, P54], user sentiment [P24, P25, P50], results of card-sorting sessions [P09], and answers to questionnaires and surveys [P04, P31, P42, P43]. Others focused on the analysis of behavioral data, such as eye tracking [P28, P44, P51, P55], user trajectory [P08, P16, P35, P48, P49, P53, P56], and user actions and software events [P01, P02, P10, P17, P20, P21, P26, P36, P52, P57]. In particular, one paper focused on elaborating collaborative discussion and analysis features for UX evaluators [P17]. A quarter (16) of the papers aimed to facilitate the analysis of player behavior. Their visualizations either aim to *inform the player* about their performance [P30, P34, P38] or *assist in the PX evaluation of a game* (PX — Player eXperience) [P03, P05, P07, P11, P13, P15, P18, P22, P23, P29, P38, P40, P45, P47]. Four papers about UX and PX evaluation also intended to create visualizations accessible to multidisciplinary development teams [P18, P20, P31, P42].

Some papers aim to *raise user awareness* about the interaction with the software. With visualizations of their sentiments, users assessed how the interaction influenced their emotional state [P25, P37]. Through visualizations of web navigation [P27, P39, P41] and music playback histories [P32], users analysed their behavior and preferences. Two papers intended to *guide exploratory searches in text collections* [P33, P46]. Other papers aim to *facilitate the communication of data discoveries* [P12, P14, P42]. For instance, a paper included a visual narrative creation feature on a visualization to enhance the communication of steps preceding a discovery [P14]. In another scenario, a paper proposed a visualization to communicate a product's UX among interdisciplinary development teams [P42].

4.4.2 What research methods were used, and how many participants were included?

Some authors employed research methods to explore the visualizations' domain, such as the domain's users, tasks, and tools. Examples of these research methods are: *field studies* [P52], semi-structured *interviews* with experts [P17, P19, P20, P34, P42] and other interview types [P06], *questionnaires* with users [P34], *literature reviews* [P36, P42], and *contextual inquiries* with experts [P19]. During design, *concept tests* were applied [P20]. In some cases, the prototyping stage included continuous user [P05, P18] and expert feedback [P12].

For these methods, when specified, the number of involved participants was less than 9 in three papers [P05, P17, P20], between 10 and 30 in two [P19, P42], and above 100 in one [P34]. The participants included designers [P19], UX professionals [P05, P17, P20, P42], software and game developers [P05, P18], data analysts and scientists [P05, P12, P52], and product and customer success managers [P20]. In addition, some participants were students and professionals from other areas [P06, P34].

The prototyping stage also incorporated multiple research methods. *Usability tests* were conducted according to different experimental formats. These formats involved: a group of participants to test a prototype [P01, P07, P14, P17, P20, P23, P32, P36, P37, P45, P55, P56], a group to test and compare multiple tools, such as the proposed visualization and its variants [P19, P22, P31, P33, P39, P41, P43, P44, P48, P49], and a group for each tool to be tested [P08, P12, P25, P33, P34, P46].

In parallel to usability tests, users [P01, P12, P14, P17, P19, P20, P22, P23, P31, P32, P34, P37, P39, P41, P43, P57] and experts [P01, P07, P14, P19, P20, P34, P44, P45, P56] participated on *interviews*. These interviews were unstructured [P32], semi-structured [P07, P12, P17, P22, P23, P31, P34, P41] or structured by the repertoire grid technique [P43]. Similarly, usability tests were complemented by *questionnaires*, such as NASA-TLX [P22, P23, P48, P49], USE [P37], UEQ [P34], INTUI [P31], SUS (System Usability Scale) [P33, P48, P49], ResQue framework's questionnaire [P33], and adaptations of TAM (Technology Acceptance Model) and TAM2 [P14][1]. Authors also adopted custom questionnaires [P01, P07, P12, P17, P19, P20, P22, P23, P25, P32]. For instance, some of these custom questionnaires aim to evaluate user engagement [P12] and satisfaction [P32] and measure quality metrics of the visualization, such as clarity, ease of use, and informativeness [P07, P17, P22].

In some cases, custom questionnaires [P13, P51] and user [P51] and expert [P13] interviews validated the prototype independently of usability tests. In addition, validations were also conducted with *unmoderated remote tests* [P38, P46], *algorithmic performance analyses* [P35], *expert inspections* [P48], and analyses of *eye tracking* [P31] and *customer feedback* data [P42].

Among the research methods adopted during validation, when mentioned, the number of participants was less than 7 in eleven papers [P14, P20, P32, P34, P36, P39, P45, P48, P55, P56, P57], ranged from 9 to 16 in fifteen [P07, P12, P13, P17, P22, P25, P31, P32, P34, P37, P41, P43, P48, P49, P51] and between 20 and 50 in seven [P01, P08, P19, P23, P33, P43, P44], and was at least 100 in two [P38, P46]. The participants were designers [P19, P43], visualization experts [P36], UX professionals [P17, P20, P56], HCI experts [P36, P48], software and game developers [P07, P13, P14, P18, P22, P23, P31, P32, P45], data analysts and scientists [P41, P49], and product and customer success managers [P20]. Furthermore, some participants were students and professionals from other areas [P01, P08, P12, P25, P31, P32, P33, P34, P37, P38, P39, P43, P46, P48, P51, P55, P57].

4.4.3 What were the achieved results in evaluations?

Results of the evaluations indicate that *papers accomplished their described objectives.* Their visualizations reduced the workload and improved the analysis capacity of the participants during UX [P14, P17, P19, P36, P42, P43, P46, P51, P52, P56] and PX evaluations [P05, P22, P45]. In addition, some visualizations promoted emotional [P25, P37] and behavioral awareness as intended [P32, P39, P41].

Participants appreciated the presence of *complementary data*, such as contextual data [P36] and data about the software of the interaction [P44]. When these data were absent, the participants demanded more data sources and details [P20, P22, P34, P41, P42, P45]. On the other hand, *uncommon data* for domain tasks decreased the visualization's clarity [P17, P44, P45]. The same consequence arose from *ambiguous data* with an uncertain collection method [P13, P45]. Meanwhile, participants felt displeased and discouraged by *ambiguous or negative labels*, such as “miscellaneous cluster” [P39] and “negative user sentiment” [P19]. Participants also raised questions about user interaction *data privacy and ownership* [P25]. In addition, participants recognized how *data value depends on the*

[1] The glossary of Appendix B describes some questionnaires.

analysis context, such as the role of the analyst [P20] and the analysis timing (before, during, or after the interaction) [P34].

The *obstruction* of visualizations' items resulted from the high quantity, large size, and low transparency of items [P01, P07, P13, P23, P45, P48]. The *use of VR/AR headsets* limited the field of vision and induced participant fatigue [P36]. Violin [P56], stream [P39], and 3D bubble [P01] charts were *unfamiliar chart formats* for some participants. Consequentially, some participants had difficulty interpreting the encoding [P39, P56] and using the interaction idiom [P01]. Furthermore, participants mentioned *confusing and enlightening chart formats* across different data sources. Participants appreciated analysing user trajectories through Sankey and node-link diagrams [P48]. On the other hand, node-link tree diagrams posed challenges in this regard due to their limitations on representing cycles without repeating nodes [P48]. In another case, tables aided inexperienced evaluators in structuring their analysis [P19]. Meanwhile, heatmaps highlighted regions displaying atypical behavior [P13].

Regarding the interaction idiom, participants appreciated interactions promoted by the Visual Information Seeking Mantra (SHNEIDERMAN, 1996). The following were *valued functionalities* by the participants divided according to the functions of the mantra:

(1) Overview of the data: small multiples [P51] and data aggregation [P07, P51].

(2) Zoom and filters: spatial filters [P51] and filters by user [P31], keyword [P46], category [P23], and user sentiment [P46].

(3) Details on demand: revelation of aggregated data [P52] and control of the visualization's temporal state [P22, P23, P41], camera view [P22, P23], and spatial arrangement of items [P52].

Participants also noted some *issues with the functionalities*. Data aggregation overlooked subtle or unusual user behavior patterns [P45]. Participants disregarded pre-established views of the visualizations [P22]. Furthermore, the lack of a time indicator for lengthy data filtering operations confused participants [P52].

4.4.4 What are the purposes of use of the visualizations?

Analyse user behavior is a common objective among the visualizations [P01, P03, P05, P07, P08, P10, P11, P12, P13, P14, P16, P17, P18, P20, P22, P23, P26, P27, P29, P31, P32, P35, P36, P38, P39, P40, P41, P42, P44, P45, P46, P47, P48, P49, P51, P52, P53, P55, P56, P57]. Some focus on visually comparing user behavior based on usage trajectory [P47] and performance metrics [P30]. Others aim to categorize users by task outcome (e.g., success or failure) to observe commonly followed trajectories [P35, P48].

Format	Citation
Adjacency Matrix	[P10, P40]
Affinity	[P06]
Arc	[P48]
Area	[P05, P12, P39]
Bar	[P02, P12, P32, P39, P46, P50, P52]
Box	[P02]
Bubble	[P12, P41, P51]
Choropleth Map	[P07, P13, P15]
Gantt	[P50]
Heatmap	[P11, P25, P28, P35, P36, P38, P40]
Histogram	[P25, P34]

Format	Citation
Icicle	[P02]
Line	[P05, P10, P17, P43, P56]
Node-link	[P03, P09, P14, P16, P24, P26, P27, P29, P40, P48, P49, P54, P57]
Plain Text	[P46]
Point	[P32, P36]
Radar	[P05, P30, P33, P42]
Sankey	[P35, P48, P56, P57]
Scanpath	[P28, P50, P51]

Format	Citation
Scatter	[P04, P28, P36, P44, P50]
Sliders	[P34]
Stream	[P39]
Stripe	[P10, P17, P39]
Sunburst	[P41]
Table	[P02, P04, P14, P17, P19, P49, P53]
Trajectory Map	[P05, P07, P13, P15, P40, P45]
Violin	[P56]
Word Cloud	[P39, P50]

Table 3 – Well-known chart formats present in the literature.

Reconstruct the interaction to *explore the perceived experience* [P02, P06, P19, P24, P31, P42, P43, P46] or *explore the internal state* of the users [P04, P09, P15, P17, P24, P25, P33, P37, P50, P54] is another described purpose. In some cases, it is possible to compare the perceived experience of a product over time or between different products [P02, P31, P42]. Furthermore, some visualizations aim to *analyse the context* of the interaction (e.g., user movement and nearby objects) [P36, P56] and *identify the user profile* based on demographic data and user behavior [P35, P53]. Finally, users are also able to produce annotations to *encode events and UX issues* [P06, P14, P16, P17, P19].

4.4.5 What data sources were explored by the visualizations?

It is possible to explore *demographic* data [P03, P20, P31, P35, P40] and *acoustic* data from think-aloud sessions, such as tone, volume, and speech rate [P17]. Regarding *user sentiment*, four data types emerged: excitement level (excited or calm) [P07, P13, P25], polarity (positive, negative, or neutral) [P15, P17, P19, P24, P46, P50], valence level (pleasant or unpleasant) [P25], and emotion (happy, sad, surprised, among other emotions) [P24, P37, P45, P50].

Following the use of a software application, it is possible to explore data about *user actions and software events* [P04, P05, P07, P10, P11, P12, P13, P15, P17, P18, P20, P22, P23, P26, P27, P28, P32, P35, P36, P38, P39, P40, P41, P45, P46, P47, P53, P56]. This includes an event's frequency [P12, P27, P35], intensity (e.g., scroll speed) [P17] or location on a screen [P04, P28, P36, P53]. The exploration of *navigation trajectory* through web pages [P08, P10, P27, P39, P48, P49, P52, P53] or software states is also possible [P03, P14, P16, P29, P35, P40, P56, P57]. The navigation can be associated with *eye tracking* data, such as fixation points [P28, P44, P50, P51, P55], eye movement type [P51], field of view [P44, P56], fixation duration and frequency [P28, P44, P45, P50, P51, P55], and gaze trajectory combined with the fixated regions' intensity, color, and orientation [P44].

User interaction can produce *interaction-related text*, such as user reviews [P06, P19, P24, P31, P46], transcribed comments [P07, P15, P16, P17, P36, P50], annotations [P16], metadata of visited pages [P39, P54], and words used in search fields [P33, P39] or card-sorting sessions [P09]. *UX, satisfaction, and efficiency metrics* were collected using software usability and aesthetic attractiveness questionnaires [P02, P26, P31, P42, P43]. To explore the interaction through a *performance metric*, task completion time [P04, P08, P18, P20, P26, P28, P35, P52, P53], user action frequency [P04, P27], and task success rate were considered [P18, P50, P52, P53].

4.4.6 What chart formats were used?

Table 3 lists *well-known chart formats* present in the papers. The visualizations also adopted custom chart formats, such as a *particle cloud system* to point out user sentiment locations [P45], a *multi-line timeline* to display concurrent user navigations [P08], and a *3D bubble* chart that compose a configurable framework [P01]. Other custom formats include a *tangible magnetic force* format [P37] and the *overlapping of points and lines* to display continuous events (lines) concomitant with discrete events (points) [P56]. Through augmented reality, a *3D tube* format depicted users' and objects' movement and dispersion of positions in their original environment [P36]. In addition, a *3D scene simulation* reproduced gameplay in its original virtual environment [P22, P23].

4.4.7 How was the function of "having an overview of the data" included in the visualizations?

Visualizations employed the *focus plus context* technique to provide details while maintaining an overview of the data [P12, P25, P26, P33, P51]. Through the *small multiples* and *superimpose* techniques, analysts could simultaneously view data of multiple users [P02, P05, P51, P56], timeframes [P26, P51], software interface regions [P51], and evaluation scenarios [P02, P26]. Chart formats that used these two techniques included the scanpath [P51], histogram [P26], table [P02], violin [P56], icicle [P02], radar [P05], and bar formats [P02, P26]. Regarding summary charts, bar [P20, P32, P39, P50] and stream charts [P39] depicted the *overall frequency and duration* of events and user sentiments by using accumulated and partitioned data over time. Numeric indicators in pop-up windows [P08] and a device color in tangible visualizations [P37] served the same objective.

To decrease the level of detail, visualizations also applied techniques to *aggregate* the data. A force-directed edge clustering algorithm aggregated data depicting gaze direction [P51]. In a trajectory map, a spatial subdivision algorithm grouped lines representing a player's character movement [P45]. In a particular case, an analyst could customize the algorithm that would aggregate the bubbles of a bubble chart [P01]. Furthermore, two node-link diagrams aggregated navigation trajectory data in multiple ways [P03, P40]. Their vertices encoded aggregations of software states resulting from customizable algorithms. Each of their edges depicted the aggregation of transitions between two states. Finally, these diagrams had symbols that represented the aggregation of users. In this case, users were aggregated based on the software state they currently were on.

Code	Technologies
Algorithms and techniques	Algorithm of Andrienko and Andrienko (2011) [P07], Artificial neural networks [P15, P50], Bubble Sets [P44], CMDS [P03, P29], DBSCAN [P07, P40], Google speech recognition [P50], I-VT [P44], isomorphism check of Sun et al. (2012) [P40], K-means [P35, P39], Lexicon-based classifier [P15], LinRel [P33], algorithm of Newman (2004) [P54], QT [P03, P40], SVM [P50], and TF-IDF [P54].
APIs	Google speech recognition API [P17], and OpenGL API [P03].
Chart development software	ArcGIS [P11, P28, P47], Graphviz5 [P49], NVivo [P21], OGAMA [P28], SWISH DataLab [P49], and Universal Visualization Platform [P16].
Computer languages and associated technologies	Adobe Flash ActionScript [P08], AJAX [P18], C++ [P44], CSS [P06, P19, P32, P46], C# [P03, P28, P40], HTML [P06, P19, P30, P32, P39, P46, P50], Java [P28, P29, P32, P51, P52] (with Hadoop [P52], MDSJ [P03, P29, P40], OpenCL9 [P51], and Swing [P51]), JavaScript [P04, P06, P12, P15, P18, P19, P24, P30, P32, P35, P39, P46, P50, P55] (with D3.js [P12, P24, P35, P50, P55], jQuery [P18, P19], NodeJS [P15], Protovis [P18, P32], React [P15, P35], and RequireJS [P30]), PHP [P18, P39], Processing [P10, P51], Python [P15, P17, P19, P35, P37, P46, P50] (with Auditok [P17], Django [P19], Gensim [P46], OpenCV [P17], Praat-Parselmouth [P17], and Valence Aware Dictionary and Sentiment Reasoner [P17]), SQL [P15, P18, P19, P28, P39, P47, P55] (with MySQL [P18, P39], NoSQL MongoDB [P15], PostgreSQL [P19], RavenDB [P55], and SQLite [P28]), Swift [P31], and Typescript [P15].
Data formats	AVI [P50], CSV [P36, P50, P51], geoJSON [P28], JSON [P24, P27, P32], MOV [P50], MP3 [P50], MP4 [P50], WAV [P50], and XML [P01].
General purpose software	Box2DWeb [P12], HAMSTERS-XLE [P02], Qt5 [P44], Unity [P15, P22, P36, P45], and Webstrates [P06].
Physical devices	Arduino microcontroller [P37], Emotiv EPOC [P50], Microsoft HoloLens v2 [P36], Tobii pro X2-30 [P44], and Tobii TX300 [P28].

Table 4 – Technologies used to develop the visualizations divided by generated code.

4.4.8 How was the function of "applying zoom and filters to the displayed data" included in the visualizations?

Analysts could control the *zoom* of the visualizations [P08, P18, P24, P31, P40] and direct their view of the data through *panning* [P08, P31, P40]. Some visualizations offered the functionality of *individually hiding items on demand* [P01, P36]. Through *attribute filters*, it was possible to select which attributes to display [P08, P11, P26]. Filters based on the values of an attribute were also available but in abundance. Those filters either selected items to be displayed or highlighted. The most adopted filters were the *temporal filters* [P05, P12, P15, P17, P18, P24, P25, P27, P32, P36, P40, P41, P50, P51]. Analysts could use these temporal filters through controls [P05, P15, P32, P51], slider bars [P25], calendars [P32], and animations [P05, P12, P17, P24, P40, P41, P51].

Through *spatial filters*, analysts could pinpoint items in a region of a software interface [P51] or virtual environment [P18]. The spatial filters could also limit the data to items starting/ending in an area or around/outside a region [P51]. Assisting the analysis of movement data, *direction filters* constrained data to the movements that followed a specific direction [P51]. In node-link diagrams, *cycle and path filters* selected paths based on length [P40] and number of vertices and edges of a category [P40]. Through these filters, analysts could also pick paths that contained a vertex or edge [P24]. Furthermore, the filters could hide cycles that represented comebacks in the user navigation trajectory [P29].

Visualizations included *category filters* [P03, P05, P15, P19, P22, P23, P26, P32, P40, P51] based on, for example, categories of software events [P03, P05, P15, P32, P40]. In user review visualizations, *coverage filters* narrowed visible reviews to those unknown to the analyst [P46]. *Filters by quantitative threshold* [P03, P20, P50, P56] limited the data based on duration [P50] and frequency [P03, P56]. Furthermore, visualizations contained filters, such as *keyword filters* [P19, P39, P46], *user sentiment filters* [P15, P24, P25, P46, P50], *filters by user* [P05, P15, P29, P35, P36, P45], *filters by software* used in the interaction [P16], *annotation creator filters* [P17], and *filters by session composition*, which eliminated sessions without an event of a specified type [P52].

In addition, some visualizations provided query interactions. Unlike filters, which act on the data on display, a query starts from the absence of data being visualized (MUNZNER, 2014). In this regard, it was possible to select the items through a *query by keyword* [P08, P16, P46], and *query by category* [P20]. In node-link diagrams, a *query by subgraph* was also available. Through this query, an analyst could select graphs with a specified subgraph within it [P08, P40].

4.4.9 How was the function of "viewing the data at the most detailed level" included in the visualizations?

As mentioned in Section 4.4.7, the *focus plus context* technique preserved the context while revealing details of the data [P12, P25, P26, P33, P51]. Analysts had the *control of the visualization's temporal state* by interacting with a slider [P03, P22, P36, P40, P50], a button (e.g., play, pause, fast-forward) [P03, P40], and a timeline [P17, P41]. In 3D chart formats, analysts could explore details about multiple views through the *camera view control*. This control involved diverse interactions, such as movement [P01, P22, P23, P36], rotation [P01, P22, P23, P36], zoom [P22, P23], and the selection of predefined settings [P22, P23, P45], such as the user's camera view during the interaction [P22, P23].

The gradual *revelation of aggregated data* occurred as an analyst interacted with items, such as bubbles in bubble charts [P01], higher-level notes in affinity diagrams [P06], and edges in node-link diagrams [P40]. In node-link diagrams, it was possible to reveal underlying subgraphs by clicking on a vertex [P03, P40]. The press of a button expanded the lines of a table to uncover their aggregated counterparts [P19]. Furthermore, analysts could adjust the temporal (e.g., hours to days) [P32, P39, P41] and spatial granularity of the data [P45]. In particular, the spatial granularity could change according to the size of the chart's spatial mesh cells [P45].

Analysts could choose the *spatial arrangement of items* when necessary. The new arrangement could be calculated by a sorting algorithm, such as an attribute-based algorithm [P19, P40], the breadth-first search algorithm [P40], and the RCM algorithm for sorting adjacency matrices [P40]. The arrangement could also be a result of an alignment. A timeline could be centered on a chosen reference point [P26], and stacked bars could align starting from a specific event type [P52]. Finally, the analyst could overlay an item above others by pointing a mouse closer to the items' center [P08].

By selecting an item, the analyst could trigger the *highlighting of related data* by coloring [P06, P25, P39, P41] or linking [P08, P51]. For instance, clicking on items that encoded events generated links to similar events [P51]. Similarly, interacting with an item linked it to its corresponding occurrence time shown on a timeline [P08]. In addition, selecting a histogram bar highlighted regions of the software interface corresponding to the encoded events of the bar [P25].

Alternatively, selecting items also resulted in the *revelation of details* in text, audio, and image formats [P01, P02, P03, P05, P08, P14, P15, P19, P24, P26, P30, P40, P41, P45, P48, P49, P50, P56]. For instance, selecting a sector in a sunburst chart exposed the quantity represented by the sector's angle [P41]. Interaction with particle clouds revealed corresponding category and occurrence time [P45]. Clicking on symbols encoding users or feedback providers exposed attributes like demographic data [P03, P19, P40]. Additionally, Gantt chart timeline's bars were linked to audio transcriptions and image frames of the interaction recording [P50].

4.4.10 What technologies were used to develop the visualizations?

Table 4 details the technologies adopted during development[2]. Independently or in conjunction, papers used chart development software and computer languages to implement the interfaces and collect, process, and visualize data. The five most popular computer languages were JavaScript, HTML, Java, SQL, and Python. Figure 11 illustrates the overlap of programming language usage across the papers (SQL has been omitted to avoid overcrowding the chart, as it was used in conjunction with various programming languages). The data formats ranged from text to multimedia formats. In addition, the

[2] The glossary of Appendix B describes some technologies.

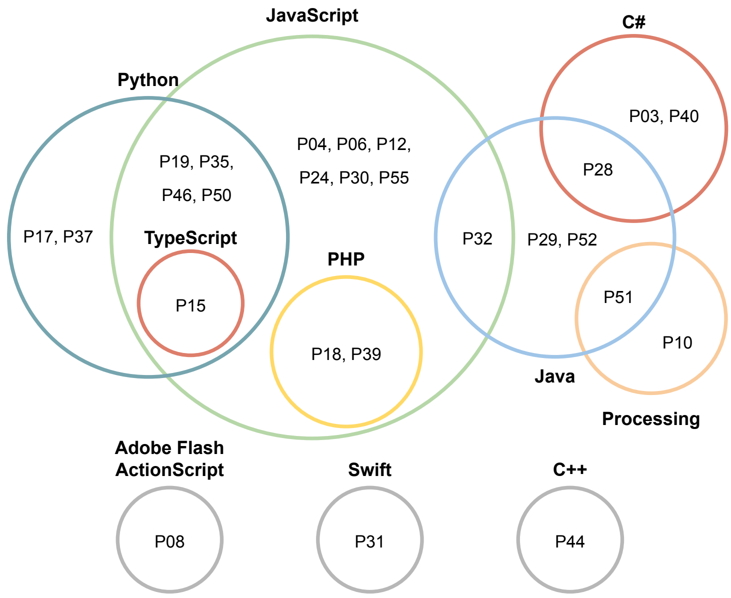


Figure 11 – Intersection between programming language use (except SQL).

papers complemented the implementations with pre-existing general-purpose software, APIs, physical devices, algorithms, and techniques. Physical devices were incorporated to build tangible visualizations [P37] and collect eye tracking [P28, P36, P44] and user sentiment data [P50]. The algorithms and techniques fulfilled various purposes, including eye tracking [P44] and user sentiment classification [P15, P50], data clustering [P03, P07, P35, P39, P40, P54], and user intent modeling [P33].

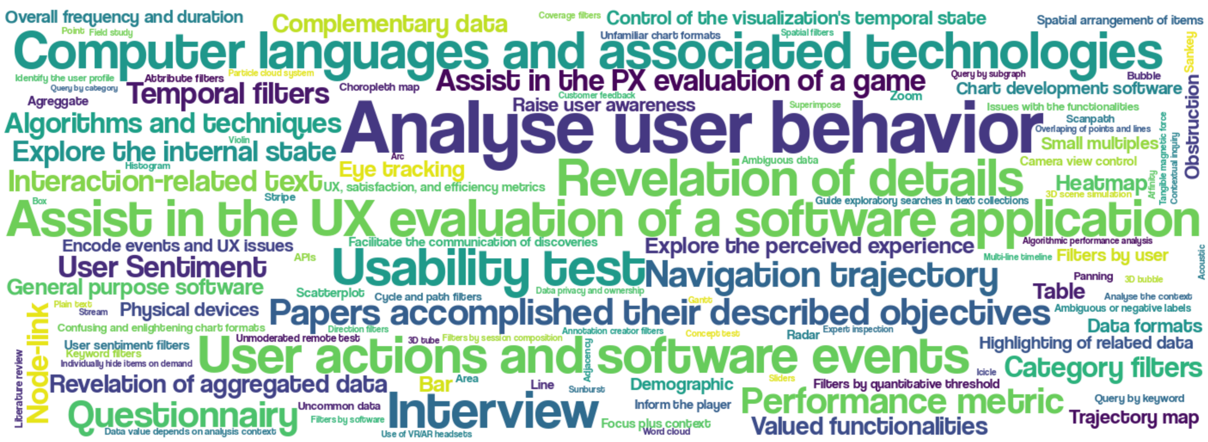


Figure 12 – Word cloud of codes with size proportional to number of associated papers.

4.5 Discussion

Figure 12 provides an overview of the codes generated through a word cloud. The size of each code is proportional to the number of papers associated with it. The word cloud

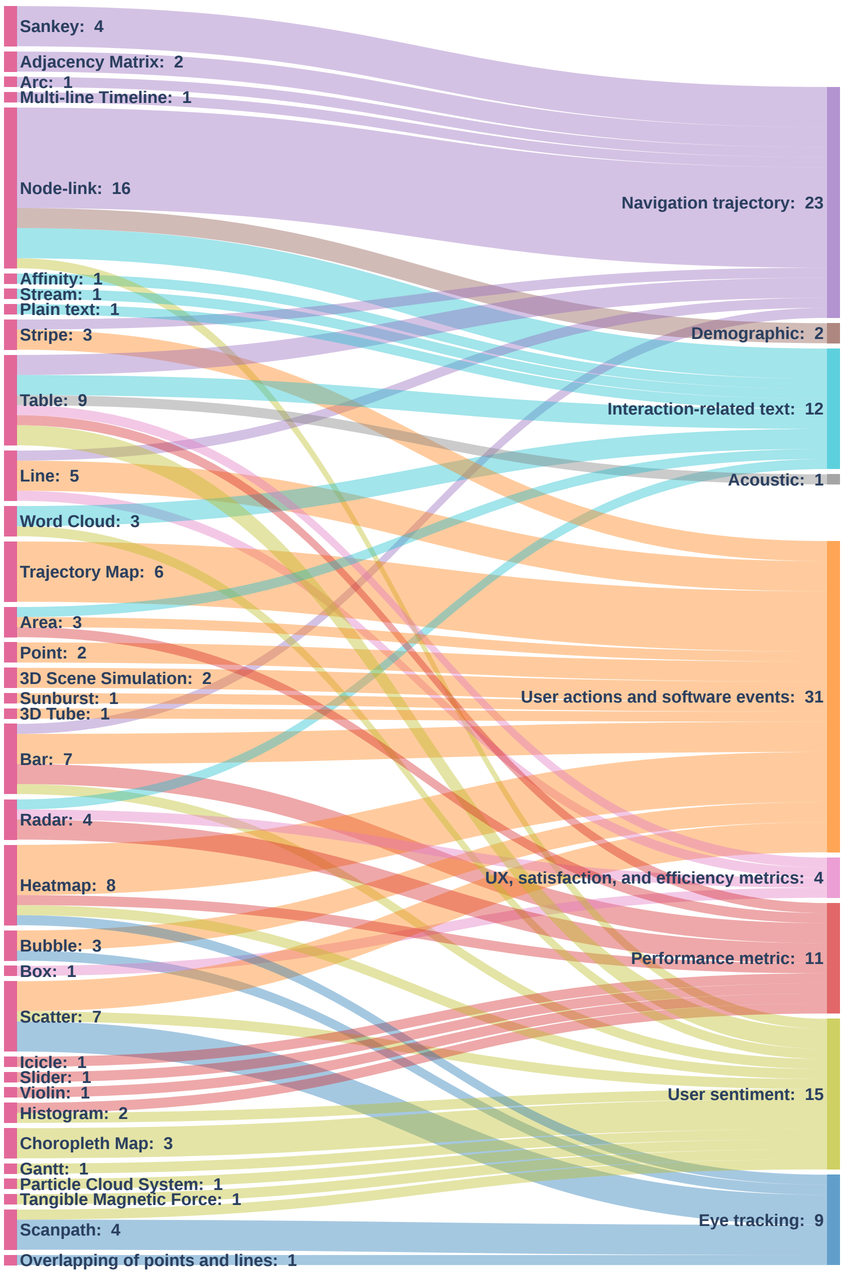


Figure 13 – Combinations of data sources and chart formats that emerged from the SML.

highlights trends, such as the use of computer languages to develop visualizations, the purpose of analyzing user behavior, and the application of usability tests and interviews to assess the visualizations. Additionally, it reveals areas for further exploration, such as experimenting with different chart formats and data sources.

Most papers aim to aid the UX evaluation of software applications by proposing visualizations. Some papers concentrated on developing visualizations to assist in game PX evaluation. Others focused on raising user awareness of their behavior and emotional state. Few papers prioritized creating (1) accessible visualizations for diverse audiences interested in UX data and (2) visualizations emphasizing communication rather than

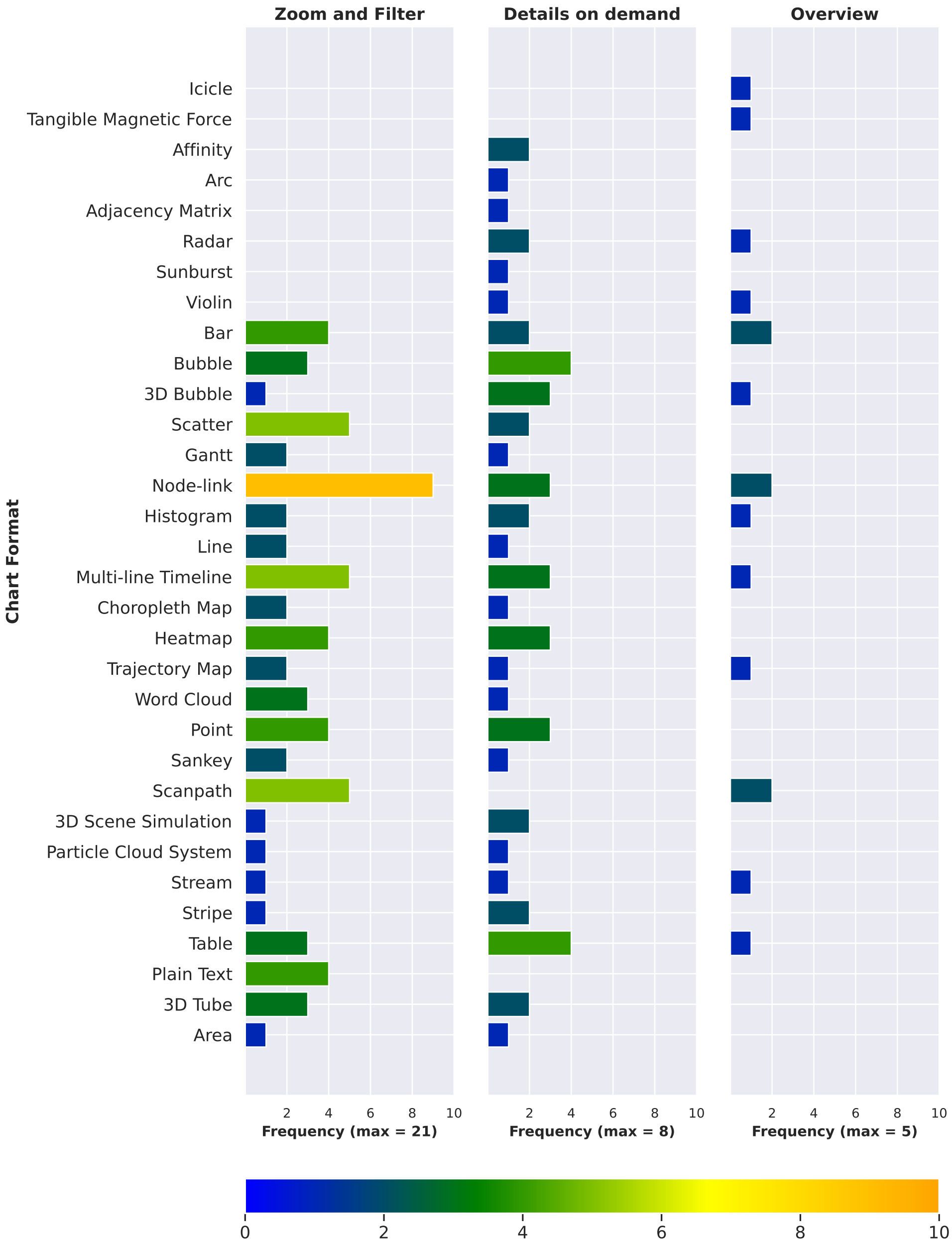


Figure 14 – Relationship between the chart formats derived from coding and the functions of the Visual Information Seeking Mantra (SHNEIDERMAN, 1996).

analysis.

The authors often applied research methods during evaluation/validation and sometimes during the conception of visualizations. Nevertheless, the SML found a diverse range of methods in both cases. Usability testing was the most prevalent, followed by questionnaires and interviews. The authors rarely used questionnaires and interviews independently of usability tests, and vice versa. Meanwhile, other methods appeared in no more than two papers.

Except for the ResQue framework questionnaire, the standardized questionnaires assessed visualizations' usability and UX. No standardized questionnaire evaluated characteristics specific to visualizations. Furthermore, the number of papers that adopted standardized questionnaires was close to the number of papers that used customized ones. During validation, UX professionals, developers, data analysts, and product managers, among others, interacted with prototypes.

The papers stated that the visualizations reduced the participants' workload and enhanced their analysis capacity during different UX data analysis tasks. Participants frequently requested additional data on the visualizations. However, issues with obstructions resulting from high information density were equally prevalent. Visualizations primarily derived data from automatically captured logs of user actions and software events. Eye tracking data and user comments, sentiments, and reviews were used to a lesser extent. On the other hand, exploration of demographic data, acoustic data, and metrics of UX, satisfaction, and efficiency remained limited.

The papers adopted a variety of chart formats. Through a Sankey diagram, Figure 13 displays the combinations of chart formats and data sources that emerged from the SML. The label of each node of the Sankey diagram is composed of a code name and the number of incident links. Each link represents, through its thickness, the number of times a chart format was combined with a data source. The most widely adopted format was the node-link diagram, often with a navigation trajectory as a data source. On the other hand, tables were the format used with the most distinct data sources. Only thirteen (40%) of the formats were combined with multiple data sources. In addition, sixteen (45%) occurred only once in the papers.

In Figure 14, a bar chart shows how many times each chart format contained each function of the Visual Information Seeking Mantra (SHNEIDERMAN, 1996). Specifically, the size and color of each bar encode the number of distinct combinations between a chart format's code and the codes belonging to a function. For reference, Figure 14 also specifies the number of codes generated for each function as the maximum possible frequency. Among the chart formats, the node-link diagram also had the most distinct applications of the functions. In general, each format contained a few different applications of each function. The incorporation of the overview function was the least mentioned in the papers. Only the node-link diagram format was combined with more than half of the possible applications of one of the functions. In addition, only 23% (8) of the formats had at least one distinct application of each function.

Most visualizations were created using computer languages. In some cases, chart development software was used, such as ArcGIS on three occasions. During data collection, physical devices were adopted to capture only eye-tracking data and user sentiment. Furthermore, the visualizations incorporated multiple algorithms and techniques involving artificial intelligence, such as classifiers and clustering algorithms.

4.6 Comparison with Munzner

As specified in Section 4.3.4, the SML's guiding questions are divided according to the analysis framework of Munzner (2014). This framework facilitates visualization analysis by separating concerns about visualizations into four levels: domain situation, task and data abstraction, visual encoding and interaction idiom, and algorithm. It structures the comparison of visualizations and enables a systematic synthesis of the knowledge about design choices that pertain to a domain. In addition to its levels, the framework also incorporates the "what-why-how" questions. They are the essential questions of the framework: (1) *what* data is presented? (2) *why* use the visualization? (3) *how* is the idiom constructed? The abstraction level of the framework addresses the "what" through data abstraction and the "why" through the task abstraction. Finally, the idiom level is responsible for the choices related to "how". Next, we analyse the SML results using the analysis framework of Munzner (2014).

Domain. The primary target users of the UX data visualizations were software and game development teams. Most papers aim to facilitate the UX evaluation of a software application and the PX evaluation of a game for these users. Visualizations also focused on helping end users of the applications. Some of these papers aim to raise user awareness about their behavior and emotional state. Others intended to guide users in the exploration of an application. Regarding the application of research methods, the articles involved participants with diverse roles beyond UX professionals. Only eleven (19.2%) papers described research methods used during visualization conception. However, more than half of the papers evaluated their visualizations. Through evaluations, it was possible to identify visualizations that reduced development teams' workload and improved their analysis capacity.

Abstraction. Most (94.7%) visualizations aim to analyse user behavior, perceived experience, and user's internal state. For this purpose, they display data sourced from, for example, software logs (e.g., user actions, software events, and navigation trajectory data), external devices (e.g., user sentiments and eye tracking data), and questionnaires (e.g., UX, satisfaction, and efficiency metrics).

Idiom. The visualizations employed 28 well-known chart formats and seven custom ones. By number of citations, the formats that stood out were the bar chart (7), heatmap (7), line chart (5), node-link diagram (13), radar plot (4), Sankey diagram (4), scatterplot (5), table (7), and trajectory map (6). Visualizations also adopted different interaction idioms, such as aggregation, superimpose, panning, filtering, and selection for revealing details. The most employed among them were the manipulation of data based on its time of occurrence and the selection to reveal details. Among the SML's codes regarding the interaction idiom, fourteen (40%) referred to filters. Furthermore, three interaction idioms stood out for being more involved with a particular domain or data source: software filters, user sentiment filters, and user filters.

Algorithm. A quarter of the papers (14) used JavaScript to implement the visualization. JavaScript was usually paired with Protovis and D3.js. In contrast, only six (10.5%) papers used chart development software. Other languages have also been adopted, such as HTML, SQL, Java, and Python. Most algorithms and techniques incorporated into the visualizations stem from artificial intelligence, such as classifiers and clustering algorithms.

4.7 Limitations and threats to validity

The papers retrieved by the search string depend on both the specific string used and the digital libraries searched. One of the limitations of this study concerns the search string applied. Although we followed best practices for string construction and tested different search strings, we recognize that the selection of terms may not have been exhaustive enough to delimit the sample of results from each scientific database. We used keywords from seminal works. However, the literature related to data visualization has a diversity of synonyms and terminological variations that this work may not have covered. In addition, the way in which the string was applied in different databases may have resulted in variations in document retrieval, either due to Boolean logic or the different functionalities and indexing of each repository. Consequently, there is a risk that we excluded potentially relevant articles or, to a lesser extent, included studies outside the intended scope. We recognize that future investigations may benefit from a broader review of search terms and additional sensitivity and precision tests to refine the results.

SMLs focus on a limited set of papers. Thus, the codes that emerged can evolve through time as the field of UX data visualization advances. For instance, new characteristics of UX data visualization may arise in future studies. In addition, a single researcher conducted the selection, extraction, and coding stages of the SML. Thus, the codes generated carry with them the subjective interpretations and associations of a single researcher. To decrease the potential bias of the researcher, two researchers with experience in conducting SMLs supervised the outcomes of each stage of the SML. Furthermore, the extractions and codes generated are available on an online electronic spreadsheet. It is also possible to reproduce the SML by following the stages described in Section 4.3.

4.8 Conclusions

This paper aims to support researchers and professionals building UX data visualizations by presenting the state of the art on UX data visualizations. To this end, we conducted an SML that included 57 papers that deal with visualization types and their characteristics. The open coding technique and ten guiding questions structured the analysis of the papers. The questions were based on the four levels of the analysis framework of Munzner (2014) and the Visual Information Seeking Mantra (SHNEIDERMAN, 1996).

Most of the included papers focused on helping development teams evaluate a software application's UX and a game's PX. On the other hand, developing visualizations accessible to interdisciplinary teams and visualizations that enhance the communication of findings was the primary objective of few papers. In the validation stage, only personalized questionnaires measured characteristics adjusted to the visualization evaluation. As a result of this stage, the papers reported that the visualizations contributed to user interaction analysis. Besides, the results allowed us to consolidate the advantages and pitfalls of the proposed visualizations.

The mapping between data sources and chart formats highlights the mainstream usage of user actions and software events, sometimes in the format of navigation trajectory data, as data sources. At the same time, it sheds light on the sparse experimentation of chart formats and the limited exploration of demographic, acoustic, and UX metrics data. A minority of the chart formats were combined with multiple data sources. Regarding the functions of the Visual Information Seeking Mantra, the overview function was adopted on a smaller scale. Furthermore, visualizations were rarely developed using existing tools rather than computer languages.

Our results provide a body of knowledge from which researchers and professionals can develop new visualizations. They also highlight multiple gaps for future research to explore. Furthermore, it would be valuable to investigate the practitioner perspectives on UX data visualizations through interviews or a gray literature review to, for instance, refine the codes generated or identify best practices.

Declarations

Funding

This paper is supported under the grant #2020/00615-9, São Paulo Research Foundation (FAPESP). It also receives the partial support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 and Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil (CNPq) grant #309497/2022-1.

Authors' Contributions

Lucas Katib do Amaral: Conceptualization, Methodology, Formal Analysis, Data curation, Writing (original draft, review and editing), Investigation.
Maylon Macedo: Conceptualization, Methodology, Writing (review and editing).
Luciana Zaina: Conceptualization, Methodology, Supervision, Writing (review and editing), Funding acquisition.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

A spreadsheet containing the extracted data and the generated codes is available at Google Drive.

4.9 Final Considerations

This chapter concluded its systematic mapping of the literature on UX data visualization by presenting the state of the art based on 57 papers. The analysis revealed a predominant focus on supporting development teams in evaluating software and game user experiences, while areas like interdisciplinary team accessibility and enhanced communication of findings received less attention. Key findings indicated a mainstream use of user actions and software events as data sources, alongside limited exploration of diverse chart formats and data types such, as demographic or UX metrics. A significant gap identified was the sparse adoption of overview functions in existing visualizations. While acknowledging limitations such as potential biases from the search string, this work highlights gaps for research, particularly regarding the need for more varied data sources, diverse chart formats, and practitioner perspectives.

Chapter 5

Exploratory study: first round

This chapter presents the article “Supporting User-Centered Requirements Elicitation from Lean Personas: A UX Data Visualization-Based Approach” (MACEDO et al., 2024), authored by Maylon Macedo, Gabriel Teixeira, Ariel Campos, and Luciana Zaina. This article, which was published in 2024 at the International Conference on Enterprise Information Systems (ICEIS), constitutes the Step B of the overall thesis methodology. Its objective is to detail the development and evaluation of visualizations designed to explore Lean Personas data. Contributions from this work include demonstrating initial attempts at applying visualizations, validating existing knowledge related to user interaction data, and simultaneously revealing the practical complexities and challenges inherent in working with unstructured UX data. The limitations and insights derived from this study reinforced the necessity of a structured conceptual framework to guide the selection of visualizations and their underlying data.

5.1 Introduction

User experience (UX) is an important aspect of software quality that affects software acceptance by users (KASHFI; FELDT; NILSSON, 2019). Although the literature provides different definitions of UX, most of them state that UX encompasses both the software functionalities and its quality characteristics that are perceived by end-users (HASSENZAHL, 2018). Studies have discussed the importance of including the discussion of UX from the software conceptualization stages (i.e., requirement elicitation, software ideation) (KASHFI; FELDT; NILSSON, 2019; CHOMA et al., 2022). Different requirements elicitation techniques have emerged in the literature (e.g., personas, user stories) (FAILY; LYLE, 2013) and the personas technique has been recognized as a more focused user-centered approach to explore data from end-users (PINHEIRO et al., 2019). By creating fictional characters that represent different groups of end-users, the personas provide designers and developers with users' data about their needs, experiences, behaviors, and goals, hereinafter called UX data (BILLESTRUP et al., 2014; PINHEIRO et al., 2019). UX data contains information that allows mapping users' decisions and interactions with a product, enabling measurements and statements about the causes and results of implementations in interactive systems (LUTHER; TIBERIUS; BREM, 2020).

Personas often provide a rich amount of qualitative UX data (GOTHELF, 2012). Nonetheless, the traditional persona demands time and effort in user data collection which becomes an expensive technique (BILLESTRUP et al., 2014). As an alternative, Gothelf proposed the proto-persona also called lean persona, an agile approach that allows designers/developers to create concise personas based on their prior knowledge about the target audience (GOTHELF, 2012; PINHEIRO et al., 2019).

On the one hand, designers and developers recognize the qualitative UX data from lean personas as useful to elicit users' needs and preferences in their interaction (ISHERWOOD, 2020). On the other hand, they struggle to identify and extract meaningful insights about UX from qualitative data in textual format which turns the elicitation of UX-related requirements into a complex task (RHIE; YUN, 2017). The extraction of requirements involving text interpretation requires defining meanings to provide different values to the raw data (RHIE; YUN, 2017). Without defining meanings, the analysis depends on the judgment of each researcher, and the result may vary in relation to knowledge and experience about UX (RHIE; YUN, 2017; DESAI, 2019).

Visual data representations can guide designers and developers during the requirement elicitation (KASHFI; FELDT; NILSSON, 2019) on the exploration of textual data which reduces the effort to obtain information based on patterns and emergent data (DESAI, 2019). Considering the problem of browsing and exploring the textual UX data provided by lean personas, this paper presents three visualizations that deliver UX data that are obtained from lean personas previously elaborated. These visualizations lead designers and developers to explore the lean persona data gradually ranging from the data overview to data in detail taking into account Shneiderman's mantra (SHNEIDERMAN, 1996).

This paper is organized as follows. Section 5.2 presents the fundamental background that underlies the design and development of the visualizations and the related work. We explain all activities made in our systematic process to design the visualizations, including the visualizations' rationale and characteristics, in Section 5.3. In Subsection 5.3.1 we explain the dataset used as a baseline to design the visualizations. We discuss the study validity and present the planning, execution, and analysis of the visualizations evaluation in Section 5.4. The analysis of the evaluation results and our findings about the use of visualizations to get information by UX data are presented in Section 5.5. In Section 5.6 we presented conclusions and suggestions for future work.

5.2 Background

5.2.1 Requirements Elicitation and Personas

Software professionals have a common understanding that a good process of requirements elicitation promotes support for defining requirements (ORMEÑO; PANACH, 2013). Software requirements have two main classifications which are named functional and non-functional. Quality requirements like UX requirements fit into the definition of non-functional and involve understanding system tasks, user expectations, ease of use, and user satisfaction (ORMEÑO; PANACH, 2013).

As mentioned before (see Section 5.1), personas have been used as an artifact to support UX-related requirements elicitation due to the details about users' needs and characteristics these artifacts provide to developers and designers (FAILY; LYLE, 2013; TEIXEIRA.; ZAINA., 2022). However, the construction of traditional personas requires efforts of time and a great amount of data collection (BILLESTRUP et al., 2014). Lean personas is a technique that impresses more agility to the personas creation process by considering the prior knowledge of the software practitioners about the target audience (GOTHELF, 2012).

Pinheiro et al. (2019) extended Gothelf's lean persona and proposed a four-quadrant template that contained guideline questions to guide the lean persona artifact construction. The four quadrants organize the lean persona information as follows: Q1 - Demographic data; Q2 - Objectives and needs; Q3 - Behaviors and preferences; and Q4 - Difficulties. The four quadrants of the template and the guideline questions enhance the description of users' information while keeping the characteristic of being a lean artifact (TEIXEIRA.; ZAINA., 2022). To automatize the creation of lean personas based on the four-quadrant template, Teixeira. and Zaina. (2022) developed the Lean Persona+ tool.

The results about the effectiveness and efficiency of adopting personas artifacts during requirements identification were presented by Salminen et al. (2020). The authors compared the use of personas with requirements elicitation conducted using data from an analytics tool (SALMINEN et al., 2020). The results revealed that the developers spent more time identifying the requirements that were caused by the need to interpret a great amount of data from the analytics tool. On the other hand, the personas artifacts presented end-user data in a straight way which makes the developers work faster (SALMINEN et al., 2020).

5.2.2 UX data visualization

Techniques such as lean persona produce qualitative data about user characteristics, needs, preferences, and experiences (GOTHELF, 2012). This qualitative UX data can aid designers and developers in setting priorities for designing a better user experience during the requirement elicitation process (LUTHER; TIBERIUS; BREM, 2020). Besides, it can bring insights about *why* users need some software features, *what* software features are suitable to fulfill users' needs and preferences, and *how* users can interact with the features (HASSENZAHL, 2018). However, the alignment on the meaning of UX can insert difficulties in using the UX qualitative data to requirement elicitation (CHOMA et al., 2022; RHIE; YUN, 2017).

Different UX definitions are found in the literature. Hassenzahl provides a content-model definition that categorizes UX into three levels based on interaction action (HASSENZAHL, 2018). The *why* level encompasses the motivations and needs that lead an individual to use the product. The *what* level refers to product features that fulfill users' needs. The *how* level is about user activity on the product to interact with the functionalities. When the user reaches their goal (i.e., *why* level) through successful use of the concrete actions (i.e., *how* level) of product functions (i.e., *what* level) a feeling of wellbeing is awakened (HASSENZAHL, 2018). Hassenzahl's content model provides a perspective of the UX information linked to users' interaction with the product (*what* and *how* levels) and with the motivations and needs for having these interactions (*why* levels). Each level of Hassenzahl's content model focuses on information related to UX which facilitates the levels of association with UX qualitative data (ZAINA; SHARP; BARROCA, 2021).

Nonetheless, qualitative data provided from lean personas are based on text data which turns the data exploration for requirement elicitation harder (SLONE, 2009). The visualization of data in plain text format does not assist users to infer patterns and find trends or outliers (WARE, 2012). The adoption of visual representations can assist the navigation of qualitative data guiding professionals to explore and interpret UX data instead of relying on statements of individuals that analyzed the results (SLONE, 2009; WARE, 2012).

The elaboration of visual representations can be supported by Information Visualization (i.e., InfoVis) techniques (MUNZNER, 2014; WARE, 2012). InfoVis offers methods and techniques to enhance human perception in the recognition and interpretation of data (WARE, 2012). Concerning Human-Computer Interaction (HCI), InfoVis aims to build visualizations that are more adherent to the needs of a target audience and prevent users from employing excessive efforts to interpret data (WARE, 2012; MUNZNER, 2014).

5.2.3 Related Work

As far as we know, there is no previous work that proposes visualizations to support the use of data from personas. The work of personas related to requirement elicitation mostly discusses the potential of uncovering user-centered data available from the personas (PINHEIRO et al., 2019). Lean Persona+ tool developed by Teixeira. e Zaina. (2022) allows the visualization of personas data; however, the visualization is available from its template format. Considering that our approach is based on visual representations, we explored studies that provide contributions to the visualizing UX data.

Móro, Daráz e Bieliková (2014) proposed a bar chart view for visualizing gaze-tracking data focused on testing dynamic web applications. An assessment of the usability of the visualization showed that users want to visualize the data in other formats, for instance heat map graphs. To view website navigation data, four visualizations (i.e., Arc Diagram, Word Tree, Sankey Diagram, and Node-Link) were proposed by Buono et al. (2020a) to support novice evaluators in usability testing. An evaluation by applying the System Usability Scale questionnaire (SUS) with fifteen participants revealed that the Sankey Diagram was the visualization most appreciated by the evaluators for showing navigation data quickly and objectively. To support the analysis of the results of usability and UX assessments based on user tasks, Bernhaupt et al. (2020b) has created three visualizations. These visualizations were built by applying the Shneiderman (1996) (i.e., overview first, then zoom and filter, details on-demand) with two levels of granularity: overview and detailed.

Considering visualizations of qualitative data, Bakiu e Guzman (2017) proposed an approach to automatically detect software feature issues, usability, and UX dimensions (e.g., satisfaction, comfort, and motivation), and sentiment polarities (i.e., positive, negative, or neutral) according to user ratings in apps reviews. Similar to Bernhaupt et al. (2020b) work, Shneiderman (1996) was applied in building the views, visualizing the results at two levels of granularity: overview and detailed. The authors do not report an assessment of the visualizations.

5.3 UX Data Visualization Design

To design the visualizations, we followed two approaches: the Design Activity framework and the Nested Model for Visualization Design proposed by McKenna et al. (2014) and Munzner (2014), respectively. Design Activity Framework (DAF) encompasses a process of four activities to support the design of visualizations: *Understand* - encourage designers to adopt a user-centered design thinking about requirements, goals, and opportunities of visualizations; *Ideate* - motivate the brainstorm of ideas and thus the selection of one that best fulfills the needs discovered in the *understand* activity; *Make* - prototype the selected idea in a visualization (or a set of visualizations), and *Deploy* - deploy an effective visualization system which can be explored and evaluated by end-users (MCKENNA et al., 2014).

Nested Model for Visualization Design (i.e., in short, named Nested Model) is a well-known approach to support the evaluation and design of visual representations (MUNZNER, 2014). It contains four levels that cover (see Figure 15): (i) Domain Situation - which guides designers in understanding the proposal of the visualization; (ii) Data/task abstraction – which helps designers find which data and its formats are used to do a task; (iii) Visual encoding/Interaction idiom – encourage designers to consider the visual representations and approaches to show the data for users (i.e., idiom of visualization); and (iv) Algorithm – develop the visualization (MUNZNER, 2014). The idiom of visualization can be perceived by chart models, interaction features (e.g., zoom and pinch), and visualization systems with the information presented in different degrees of granularity (MUNZNER, 2014).

We decided to adopt both approaches because they present a complementary perspective (see Figure 15). Design Activity Framework (MCKENNA et al., 2014) supporting us in applying the good practices presented in the Nested Model for creating the visualizations. We applied this complementary view for each visualization we have proposed. Taking into account the complementary view of the approaches, we proceeded to design the visualizations based on two steps, i.e., preparation of lean personas dataset, and visualizations elaborations. Each step is presented in the next sections.

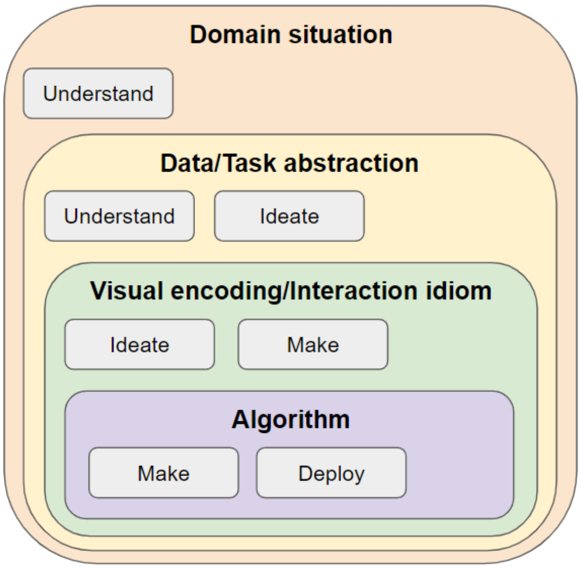


Figure 15 – The complementary view of DAF and Nested Model approaches - DAF in grey boxes and the Nested Mode in colors - adapted from (MCKENNA et al., 2014) and (MUNZNER, 2014), respectively

Considering the complementary view of McKenna et al. (2014) and Munzner (2014) model (see Figure 15), we conducted the *understand* step about the *domain situation*, i.e., how the visualizations could minimize the designers and developers efforts to browse the qualitative data of lean personas. Taking into account the great amount of qualitative data, we decided to adopt a funnel perspective to visualize the UX data as used in (BAKIU; GUZMAN, 2017; BERNHAUPT et al., 2020b). In both works, the authors applied two levels of the Shneiderman's mantra (1996) (i.e., overview and detailed). In our proposal, we planned to design three visualizations to disclose the UX data gradually throughout the three levels (i.e., *overview*, *zoom and filter*, and *details*). After that, we searched for a lean personas dataset that could be used in the visualization construction. Besides, we conducted the *understand* of the data available to then *ideate* how this data could support the conception of the three visualizations (see *data/task abstraction* level in Figure 15) as we described in the next section.

5.3.1 Lean personas dataset model

Our first activity was to obtain the persona dataset. We contacted the Lean Persona+ tool authors because we considered they would have a dataset of virtual data (TEIXEIRA.; ZAINA., 2022). We asked for access to a dataset with the lean persona artifacts and they provided a dataset containing 23 lean personas specifications in the four quadrants template (see the quadrants template explanation in Section 5.2). According to the authors, this dataset was produced from the studies conducted to evaluate the Lean Persona+ tool (TEIXEIRA.; ZAINA., 2022). The dataset encompassed lean personas that represented the target users of tourism mobile apps (e.g., TripAdvisor[1]). Each lean persona is composed by *quadrants*, *questions*, and *answers* which are the base elements of the lean persona template (see an example of a persona's data in the Figure 16-III) (PINHEIRO et al., 2019).

5.3.2 Labeling lean personas dataset

Besides the feature of creating lean persona artifacts, the Lean Persona+ tool provides a labeling tool. The labeling tool allows the activity of marking chunks of text available in the quadrant which are part of lean personas descriptions. The labeling tool provides *labels* based on the levels proposed by Hassenzahl (i.e., *why*, *what*, and *how*, see details in Section 5.2.2). With these labels, we can conduct the *labeling* of chunks of qualitative data to identify UX-relevant excerpts presented in the quadrants of lean personas. Different labels could be assigned to the same chunks of text. In a conversation with the authors tool, we have permission to use the labeling tool to prepare the dataset for being used in our visualization proposals. Three experts with (i) knowledge of Hassenzahl's content model; and (ii) academic and practical experience in Human-Computer Interaction and in Requirement Engineering participated in the labeling activity. The profile of experts was (i) one professor with 15+ years of experience; (ii) one PhD candidate with 5+ years of experience in the software industry; and (iii) one junior researcher with 1+ years who also works in the industry as a developer. After labeling 23 lean personas available in the dataset, we produced 165 labels in total (i.e., 55 *why*, 77 *what*, 33 *how*).

[1] https://www.tripadvisor.com.br/

5.3.3 Design the visualizations

Taking into account the lean personas data labeled with Hassenzahl's content model levels, we moved to *ideate* and *make* the *visual encoding* level with the participation of the three UX experts that labeled the lean persona dataset (see experts profile in Section 5.3.2). These activities were performed for each visualization that was associated with the three mantra levels (i.e., *overview*, *zoom and filter*, and *details*).

In sequence, the *make* and *deploy* activities were carried out for each visualization creation to complete the *algorithm level* (see Figure 15). To provide a functional version of the visualizations, we developed a web application. It is worth noticing that although we used a predefined lean persona dataset, our visualizations proposal are extensible to other datasets that follow the data structure presented in Subsection 5.3.1.

Our three visualizations intend to orchestrate a gradual navigation to the UX data of lean personas. In this sense, the following description illustrates how the visualizations are connected one each other. By interacting with the *overview* visualization designers and developers can first filter the dataset by choosing one of the labels (i.e. *why*, *what*, and *how*) to see the personas that have information about that selected label. With the *zoom and filter* visualization they see the quantitative data related to each persona-quadrant relationship and can select one persona to be examined in detail. *Details* visualization shows the text data and the labels assigned to each lean persona that highlight the role of that chunk of text in providing UX-relevant information. Such information can become a UX-related requirement or trigger insights to search for other requirements.

The following sections discuss each visualization design. For each visualization, we will provide information about our *rationale* for ideation that visualization, the *visual representation format* selected and its respective *characteristics*, and the *chart* and its details that implemented that visualization.

5.3.3.1 Visualization to *Overview* the UX data.

Considering the organization of lean persona artifacts data described in Subsection 5.3.1, we noticed that there is no direct way (i.e., without performing calculations) to visualize the quantitative data about the labels (i.e., UX data). Primary navigation through lean persona artifacts is performed without an overview by selecting a particular persona in a tabular view to see all data.

Rationale: An *overview* visualization avoids the need to open each persona to observe the existence of labels. This type of visualization (i.e. overview) helps users see the scope of data and choose what information they need to see in detail (SHNEIDERMAN, 1996; MUNZNER, 2014). By looking at the labeling distribution, designers and developers can concentrate their attention on the lean personas that may contain UX data that they want to look for guided by the label that abstracts the meaning of the data. Developers and designers can take advantage of information such as: if there are labels; the number of labels, and what kind of existing labels.

Visual representation characteristics: In the Sankey chart, the line can represent a relationship between categories while the thickness of the line represents quantitative data. *The objective of this visualization* is to get an overview of all available data to guide users on the data exploration. Our *overview* visualization (see Figure 16-I) can represent: the number of labels that were found in each quadrant (see Figure 16-I-A)); the number of labels that exist of each label type (see Figure 16-I-B)); and the number of labels of each type is present in each quadrant (see example in Figure 16-I-C)). *The visualization interaction is* Mouse hover in left axle to show the amount of all existing labels in that quadrant; Mouse hover in right axle to show the amount of all occurrences of that label; Mouse click in right axle to open the *zoom and filter* visualization; Mouse hover on lines to show the amount of a given label in a given quadrant.

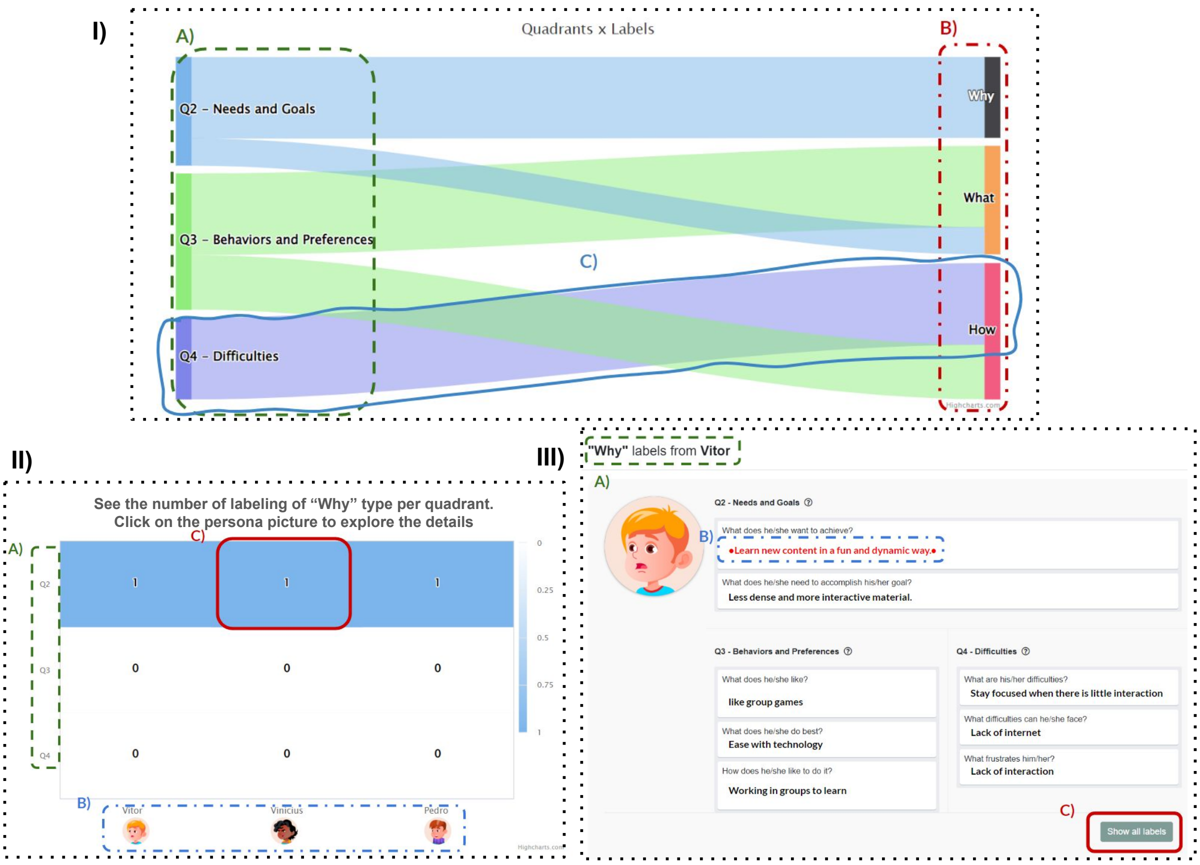


Figure 16 – I) Sankey visualization for overview, II) Heat Map showing the labeling distribution for zoom and filter, and III) Card visualization to see the detail of qualitative data of the persona | High resolution PDF on Google Drive

5.3.3.2 Visualization to *Zoom and Filter* the data.

The *overview* visualization (see Subsection 5.3.3.1) provides the view of the representativeness of each type of label assigned to all the lean personas available in the dataset. Taking into account the label selected (i.e., *why*, *what*, and *how*) previously in the *overview* visualization, designers and developers can explore the quantitative data by choosing which lean personas have the selected label.

Rationale: Considering our objective of eliciting requirements from UX data present in lean persona, *zoom and filter* visualization aids to reduce and focus on a more detailed level of information (i.e. only personas with the select label) achieving the goal of presenting the data in a funnel perspective.

Visual representation characteristics: We chose Heat Map as *zoom and filter* visualization because it can show the quantitative relationship between two categorical variables by highlighting the variation of this relationship through colors (i.e., the color scale is associated with the value scale). *The objective of this visualization* is to show distribution of a pre-selected label in quadrants-persona relationship. In our proposal (see Figure 16-II) the title of the quadrants (see Figure 16-II-A) that organize the lean personas information was inserted on the Y-axis, and the identification of the personas on the X-axis (see Figure 16-II-B). The quantitative information presented in the Heat Map is the number of labels that exist for each set of quadrants and lean personas (see Figure 16-II-C). *The visualization interaction is* Mouse click in x-axis to show all information about the selected persona; Mouse hover in cells to show a textual explanation following the pattern: "[PERSONA_NAME] has [LABEL_AMOUNT] in [QUANDRANT_NAME]".

5.3.3.3 Visualization to see the *Detail.*

After selecting the cell in *zoom and filter*, the visualization of *details* delivers the UX qualitative data in the lean persona essence by showing all the quadrants descriptions and the excerpts highlighted with labels.

Rationale: The visualization of the lean personas detailed provides the UX data more contextualized due to being available all the qualitative data. Card visualization presents small clusters of information, i.e., lean personas quadrants, which allows the delivery of content appropriate to the domain, i.e., description of needs and preferences from lean personas.

Visual representation characteristics: In the *detail* visualization, by default, only labels that match the selected label type on the *overview* visualization will be loaded with highlights, however exists the possibility of seeing all the labels to give software practitioners an *overview* of the labeling data. *The objective of this visualization* is to access all information for a pre-selected persona. In our proposal (see Figure 16-III) the card header (see Figure 16-III-A)) shows the label chosen in the *zoom and filter* visualization.

Labels are identified by colors plus a symbol. By looking at the personas, software practitioners can clearly recognize where the labels are placed and where they are not (see Figure 16-III-B)). Finally, the visualization provides a function that enables the viewing of all the labels that were assigned to that persona (see Figure 16-III-C)). *The visualization interaction is* Mouse click to enable/disable the display of all persona labels (i.e., by default only shows labels of the pre-selected type in the Sankey chart).

5.4 UX Data Visualization Evaluation

We conducted an experimental study with 20 participants following the guidelines proposed by Lazar et al. (2017) and Wohlin et al. (2012). We aimed to evaluate the facility of interpreting the three visualizations. The objective of the visualizations was to assist in extracting requirements based on the data present in the personas, for this reason the interpretation of the visualizations was analyzed. Subjects are sampled by convenience. Our study was approved by the ethical committee of the Federal University of São Carlos (UFSCar) under the process number 68524023.0.0000.5504. The next sections present the study details.

5.4.1 Apparatus

We prepared four artifacts to support our study that is described below.

An *online questionnaire* was elaborated to gather participants' consent to use the data collected (i.e., Informed Consent Form and their profile data (i.e., demographic data, years of experience, or whether they had no experience, position in the company, and knowledge of information visualization and personas techniques).

As personas descriptions are tightly associated with the domain of the product, we prepared a *scenario description* to explain the nature of the data that participants will see. We got the same scenario description that the Lean Persona+ tool authors used to produce the dataset. The scenario presents users' interaction with a tourism app: *"A tourist usually uses a mobile application (mobile, tablet, etc.) to plan and guide their trips. The app displays places in a city or region according to the interest and searches made by users. These places can be hotels, monuments, museums, parks, and restaurants, among others. For each location, the apps show details such as the location name, photos, addresses, and feedback from other individuals that often are ranked by a score from 0 to 5. Moreover, these apps can present estimates of expenses to spend in the places and also comments to assist the user to decide and plan visits to places that have good ratings"*.

We constructed a questionnaire using Google Forms to evaluate the *interpretations* of the visualizations. The questionnaire was composed of nine questions (see Table 5) and guided the participants in tasks that encouraged the exploration of the three visualizations (i.e., *overview*, *zoom and filter*, and *detail*. At the beginning of the questionnaire, an image of the *overview* visualization was presented with numerical labels to support the participants in identifying each part of the visualization the questions were referring to (see Figure 17).

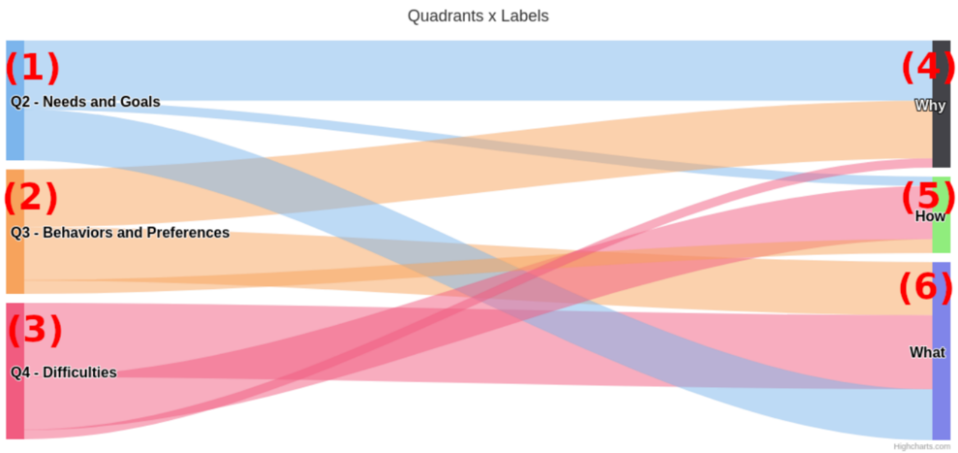


Figure 17 – Overview visualization image with numerical labels

5.4.2 Data Collection

The study was carried out with 20 participants, i.e., 14 undergraduate and 6 graduate students. The subjects were invited in the context of the "Human-Computer Interaction - (HCI)" course that was held in Computer Science at the UFSCar. The participants accepted to take part in the study voluntarily and no type of reward was involved. Although the participants had different levels of experience in relation to UX and data visualization, they mostly were professionals with experience in software development in practice. The study was conducted by two researchers, a senior researcher with 15+ years of experience and a researcher with 5+ years of experience at a laboratory at UFSCar, and lasted 3 hours.

We initially presented the concepts about data visualization (based on (MUNZNER, 2014)), levels of UX (based on (HASSENZAHL, 2018)), and lean persona (based on (PINHEIRO et al., 2019)) to level off the participants' knowledge. After that, we asked the participants to navigate into the visualizations to explore their features and understand the purpose of the personas' quadrants. During this step, the participants could clarify doubts about the concepts and the tool. We then proceeded by presenting the study's objective and after that, we asked the participants to read and answer the Informed Consent Form that state the agreement on the use of the data and images for academic ends. After that, the participants filled out the *profile questionnaire*.

In the next step, all participants received the links to: (i) the description of the *scenario* about the tourism app; (ii) the online questionnaire that guided the participants'

ID	Visualization objective	Question	Answer type	Evaluation purpose
TQ1	*Overview*	Is it possible to clearly visualize the distribution of the quantities of each label (i.e., *what*, *why*, and *how*) in relation to the quadrants?	We used a four-point Likert scale as answer options ranging from strongly disagree to totally agree	Verify whether that visualization informs the participants of the labels with the highest and lowest amounts in relation to the quadrants without the need for absolute numerical verification
TQ2	*Overview*	If you want to know more information about UX data consider: [(a) application functionality and their relationship with the persona, (b) motivations and needs of the persona, (c) concrete details of the user's interaction with the application] you should click on:	The answers present the numbers that are available in the overview image in the questionnaire (see Figure 17). The participant should provide three answers, one for each item (i.e., items a), b), and c))	Aim to verify whether the participants identified the visualization element that provides access to information from each label (i.e., *why*, *what*, and *how*). The labels have semantic meaning that represents the functionalities, user information, and forms of interaction
TQ3	*Overview*	Which quadrant has the fewest number of labels?	Titles that identify the quadrant	Identify particular information about the distribution of labels presented in the visualization
TQ4	*Overview*	Which label appears four times in the "Q4 - Difficulties" quadrant?	Name that identifies the label	Evaluate the use of interactive functions to identify specific information about the number of labels in a specific quadrant
TQ5	*Zoom and filter*	Select the incorrect alternative(s): a) The persona Helena presents information about the needs and motivations for using the tourism application in all quadrants; b) The persona Jorge presents information from the tourists' feedback feature; c) Quadrant "Q2 - Needs and Goals" quadrant is suitable to know the types of concrete actions that can be performed during the use of the tourism application; d) Quadrant "Q3 - Behaviors and Preferences" is suitable to know the types of concrete actions that exist during the use of the tourism application	Multiple-choice alternatives with the statements as options	Verify whether by interacting with the *zoom and filter* and *details* visualizations the participants could evaluate the information in the alternatives; see the participants interpretation on the visual elements.
TQ6	*Zoom and filter* and *details*	Which persona(s) provided more labels that can give insights about features to the tourism app? Justify your answer	Open-question answer from which the participants write up the personas' name and the motivation for choosing a persona	Verify whether by exploring the *zoom and filter* and *details* visualizations the participants were able to identify which personas are suitable for getting insights of the application features
TQ7	*Overview*, *zoom and filter*, and *details*	What is the persona name that displays the information "*When the post text field has limited characters? In which quadrant is this excerpt found? What label was used to highlight this excerpt?*"	Open-question answer to write up the persona name, quadrant where the information is placed, and type of label assigned to the excerpt	Verify whether by navigating in all visualizations (i.e., *overview*, *zoom and filter*, and *details*) the participants could find specific information of a persona

Table 5 – Questionnaire to conduct users during the visualizations evaluation - Question TQ2 is a cluster of three questions

interaction with the three visualizations and evaluates the *interpretation* of the visualizations (see details in Subsection 5.4.1); and (iii) the three visualizations.

5.4.3 Data Analysis and Study Validity

We adopted a descriptive approach to examine the data collected. To analyze the *interpretation* of the visualizations, we first split the participants' responses by the perspectives of *overview*, *zoom and filter*, and *details* considering the arrangement of the question presented in Table 5.

We considered the conclusion, construction, internal and external threats to discuss our study validity according to Wohlin et al. (2012) recommendations. To deal with the conclusion threats, three researchers adopted the same UX levels definition to label the data in the personas dataset which avoided the possible bias that could be introduced by different interpretations of UX data meaning.

To mitigate construct problems, we provided a set of artifacts (i.e., a scenario of the tourism apps and the questionnaire) to support the study conduction. All the participants interacted with the same visualizations developed with the same dataset, which prevents different results from being due to the types of data or visualizations. We also delivered a presentation of the relevant concepts to the study (i.e., visualization, levels of UX, and lean persona). A hands-on exercise ensured the participants had some experience with the visualizations. These two activities allowed us to mitigate the impact of little knowledge of the participants about the study concepts. Moreover, these activities benefited participants as they could clarify any doubts about the concepts.

The internal threat related to participants' fatigue during the study. We proposed a lightweight set of tasks guided by questions that stimulated the participants to explore the visualizations. Besides, the execution of tasks on the visualization lasted 1 hour at most. Participation in the study was voluntary without any financial compensation. The participants were invited to freely and optionally participate in the study, and a few chose not to participate. It is worth mentioning that there was no reward to the participants and the students who did not take part were not affected in any way. The participants' motivation was due to the fact that the topics covered in the study were popular and of interest to them. In the case of an external threat, it commonly refers to the fact that the study includes software developers who are still students. However, Salman et al. provide evidence of few performance differences between students and professionals when performing new activities (SALMAN; MISIRLI; JURISTO, 2015). In the case of this study, the activity was new for all participants as none of them had experience exploring UX data using visualizations.

5.5 Results and Discussion

We present our results in three sections: (i) the participants' profile, (ii) the discussion about the *interpretation* of the three visualizations, and (iii) the comparison of our results with the literature.

Participants' profiles: The profile questionnaire results (see Table 6) reveal that 70% (14) of the participants had a position in technology companies. Considering the knowledge background of the participants, 65% (13) have practical and/or theoretical knowledge about InfoVis. On the other hand, more than half of the participants (i.e., 60% (12)) had their first contact with the concepts of personas and lean personas in the HCI course. We also noticed that the more experienced professionals (i.e., > 1 year) reported having little knowledge of personas technique before attending the HCI course.

Identification			Background		
ID	**Experience**	**Position**	**InfoVis**	**Persona**	**Lean persona**
1*	1 to 3 years	Developer	2	c	b
2*	> 8 years	Developer	3	b	a
3	< 1 year	No job	1	a	a
4	< 1 year	No job	1	a	a
5*	> 8 years	Developer	2	a	a
6	1 to 3 years	Analyst	3	a	a
7	< 1 year	No job	1	b	b
8*	> 8 years	Developer	2	b	b
9	< 1 year	Developer	1	a	a
10	1 to 3 years	Developer	3	a	a
11*	< 1 year	Support	3	b	a
12*	6 to 8 years	Developer	3	a	b
13	< 1 year	Analyst	2	a	a
14	< 1 year	No job	1	a	a
15	< 1 year	No job	1	a	a
16	< 1 year	No job	3	b	a
17	1 to 3 years	Developer	2	a	b
18	< 1 year	Developer	1	c	b
19	1 to 3 years	Researcher	4	b	b
20	1 to 3 years	Analyst	2	a	a

Table 6 – Participants marked with * were graduate students | Participants profile - Background on InfoVis: (1) I have no knowledge about the area; (2) I have theoretical knowledge about good practices; (3) I have practical knowledge with visualizations, but I don't know good practices; and (4) I have practical knowledge and theoretical knowledge about good practices to develop visualizations. Background on persona and lean persona: (a) My first contact with personas techniques was in the HCI course; (b) I already knew the personas techniques but never used them in practice, and (c) I already knew the personas technique and have also used it in practice.

The interpretation of the three visualizations: Table 7 shows the participants' answers to the questions that guided the participants' tasks (see the questions in Table 5). The last line of the table provides the correct answers of the questions. In next sections, we will present the results for each visualization individually and the use of the three visualizations together, and an overall discussion of the results.

In the *Overview* visualization the participants did not have difficulties in identifying and interpreting quantitative data about labels to select their first step for navigation. We saw that only two participants (one participant for each question) did not answer correctly questions TQ3 and TQ4 (see Table 5) about the distribution of UX labels, i.e., *why*, *what* and *how*. We see that almost all participants agreed with the statement on visualizing the distribution of the quantities, i.e., 45% (9) of the participants "partially agree" and

ID	TQ1	TQ2-A	TQ2-B	TQ2-C	TQ3	TQ4	TQ5	TQ6	TQ7
1	Strongly agree	(6)	(4)	(5)	Q2: Needs and Goals	Why	b, d	Claudia, Leandro	(Janaína), (Q4), (How)
2	Strongly agree	(6)	(4)	(5)	Q2: Needs and Goals	Why	d	Leandro	(Janaína), (N/A), (How)
3	Strongly agree	(6)	(4)	(5)	Q2: Needs and Goals	Why	b, c	Leandro, Clara, Cláudia	(Janaína), (Q4), (How)
4	Strongly agree	(6)	(4)	(5)	Q2: Needs and Goals	Why	b, c	Clara, Leandro, Claudia	(Janaína), (Q4), (How)
5	Strongly agree	(6)	(2)	(1)	Q2: Needs and Goals	Why	a, c	Claudia, Lucas, Luiz, Janaína, Lyn, Elena	(Janaína), (Q4), (N/A)
6	Strongly agree	(6)	(4)	(5)	Q2: Needs and Goals	Why	b, c	Leandro, Claudia, Clara	(Janaína), (Q4), (How)
7	Partially agree	(4)	(6)	(5)	Q2: Needs and Goals	Why	b, c, d	Clara, Leandro, Claudia	(Janaína), (N/A), (How)
8	Partially agree	(5)	(4)	(6)	Q2: Needs and Goals	Why	b, c	Fábio	(Leandro), (Q3), (N/A)
9	Partially agree	(5)	(4)	(6)	Q2: Needs and Goals	Why	c	Fábio, Francisco	(Janaína), (Q4), (N/A)
10	Partially agree	(6)	(1)	(2)	Q2: Needs and Goals	Why	a, b, c	Clara, Leandro, Claudia	(Janaína), (Q4), (N/A)
11	Strongly agree	(6)	(4)	(5)	Q2: Needs and Goals	Why	b, c, d	Clara, Leandro, Claudia	(Janaína), (N/A), (How)
12	Strongly agree	(6)	(4)	(5)	Q2: Needs and Goals	How	b, c, d	Carlos, Clara, Leandro, Claudia, Luiz, Francisco, Jorge, Adriana, Helena, Rogério, Patrícia	(Janaína), (Q4), (N/A)
13	Strongly agree	(6)	(4)	(5)	Q2: Needs and Goals	Why	c, d	Clara, Leandro, Claudia	(Janaína), (Q4), (How)
14	Strongly agree	(6)	(4)	(5)	Q4 - Difficulties	Why	d	Jorge, Adriana	(Janaína), (Q4), (How)
15	Partially agree	(6)	(2)	(5)	Q2: Needs and Goals	Why	a, c	Leandro	(N/A), (Q4), (Why)
16	Partially agree	(6)	(4)	(5)	Q2: Needs and Goals	Why	c, d	Carla, Lauren, Francisco, Jorge	(Janaína), (Q4), (How)
17	Strongly agree	(6)	(4)	(5)	Q2: Needs and Goals	Why	a, d	Ana	(Janaína), (Q4), (How)
18	Partially agree	(6)	(1)	(2)	Q2: Needs and Goals	Why	a, b	Patrícia	(Janaína), (Q4), (N/A)
19	Partially agree	(6)	(5)	(4)	Q2: Needs and Goals	Why	a, c	Fábio	(Janaína), (Q4), (N/A)
20	Partially agree	(6)	(1)	(5)	Q2: Needs and Goals	Why	c, d	Leandro	(N/A), (Q4), (How)
Key answers	Agree	(6)	(4)	(5)	Q2: Needs and Goals	Why	c, d	Clara, Leandro, Claudia	(Janaína), (Q4), (How)

Table 7 – Participants' answers to the tasks on the three visualizations - TQ2-A to C: answers were guided by Figure 17; TQ5: see the alternatives in Table 5; TQ7:"N/A" was assigned to participants that did not report an answer

55% (11) "strongly agree" with the statement (see TQ1 in Table 5). We observed more than 70% success rate in answering correctly questions on finding UX labels throughout the entire process of navigating into the visualizations, i.e., *overview*, *zoom and filter*, and *details* (see TQ2 in Table 5 and Table 7): (i) 85% (17) of respondents correctly identified where *what* labels were placed in the visualization; (ii) 65% (13) of respondents correctly pinpointed where *why* labels could be found; and (iii) 70% (14) of respondents correctly identified the *how* labels. These results revealed the effectiveness of the Sankey chart in (i) clearly showing the distribution of labels between the quadrants; and (ii) allowing navigation and data considering a specific purpose.

We motivated the participants to interact with the three different visual representations (i.e., *overview*, *zoom and filter*, and *details*) based on statements about the personas' data to use filtering feature to achieve the target information (see question TQ5 in Table 5). We understood that TQ5 suffers the influence of previous knowledge about the UX labels. We observed that 70% (14) of the participants could select at least one right answer (i.e., options *c* and *d* in Table 7). Our findings suggest that participants' performance on question TQ5 was influenced by their prior familiarity with UX labels, with 70% (14) successfully selecting at least one correct answer from the available options.

In question TQ6, the participants should examine the personas to find the ones that provide insights about the features of the tourism app (see Table 5). To accomplish the *detail* visualization to explore the lean personas information, we motivated the participants to interact with *zoom and filter* and *detail* visualizations more than once and brought an interpretative approach to the task. To correctly answer this question, the participants should pinpoint three personas (key answers: Clara, Leandro, and Claudia). Our results showed that more than half of the participants (i.e., 11 participants - 55%) answered the question by mentioning the three personas or at least one of them (Table 7). Our findings indicate that the funnel-based approach helped participants (i.e., 11 participants - 55%) to successfully address the question by mentioning at least one of the key answers.

To get a perspective of the interpretation with the three visualizations connected one each other, we proposed question TQ7 (see Table 5). In this question, the participants should explore the different paths in the visualizations and browse the *overview* and *zoom and filter* visualizations to achieve the answers available from the *detail* visualization to thus select the lean persona. The results showed (see Table 7) that (i) 85% (17) of the participants selected the right lean persona; (ii) 80% (16) selected the right quadrant; and (iii) 60% (12) selected the right label. From this result, we could conclude that the elements and the interactions provided by the three visualizations supported the search for particular UX data.

The results showed that the participants did not have difficulties in finding specific information about UX data (see questions TQ1, TQ3, TQ4, and TQ7 in Table 5 and their results in Table 7). Our results revealed that all participants got at least 50% of the questions right. We understand that the knowledge in the UX labels could influence the use of the visualizations in particular to determine which label has the desired UX data (e.g., questions TQ2-A, TQ2-B, TQ2-C, TQ5, and TQ6 in Table 5).

We noticed the results of the questions TQ2-A and TQ6 were connected. In the TQ6 question the participants should inform which persona had the information about features. In a lean way, in TQ2-A question they should inform where in the visualization they had to click to access information about features (i.e. which label). Two of the three participants who chose the wrong answer in TQ2-A also selected the wrong option in TQ6.

This result suggests that some participants could struggle to understand the relationship between labels and the embedded UX data of each label. To mitigate this issue, we conducted a warm-up to leverage the participants' knowledge about UX labels; however, we concluded that we needed to find alternatives to present the UX label meanings in the visualizations.

Comparison with the related work: We can see three main differences in relation to the use of visualization to explore personas data, the adoption of complementary approaches for the conceptions of the visualizations, and the use of different chart formats to view the data.

Firstly, we proposed three visualizations that together and connected aid developers and designers in streamlining the search for useful UX data into lean personas textual information. As far as we know there is no other similar proposal that works to make easier the exploration of personal data although personas are recognized as a proper technique for UX-related requirement elicitation as we mentioned in (PINHEIRO et al., 2019).

We have the consciousness that there are many works that adopt the Nested Model of Munzner (2014) to support the elaboration of visualizations. However, we used a different approach by putting harmonically together the good practices proposed by the Nested Model and the systematic process of visualization construction, i.e., Design Activity Framework (DAF) of McKenna et al. (2014) (see Section 5.3). The complementary perspective of the approaches supported us in following a sequence of steps (DAF), observing the details of the visualizations, and clearly specifying the elements of each one (Nested Model).

While the approaches above guided our process of visualization elaboration, Shneiderman's Mantra sustained the conception of the visualizations by introducing ways to gradually conduct developers and designers to the lean personas data. Although Bernhaupt et al. (2020) and Bakiu and Guzman (2017) have applied Shneiderman's Mantra, they do not use all three levels of visualization. The visualizations proposed by Bakiu and Guzman (2017) do not apply zoom and filter levels. Bernhaupt et al. (2020) work presents a close perspective of ours by applying the mantra; however, the authors focus only on the levels of *overview* and *detail*.

Taking into account the chart formats, we see that Buono et al. (2020) also used the Sankey chart and that this chart was highly accepted in the evaluations conducted by the authors. In the work, the authors built the Sankey chart to show the number of visits (i.e., quantitative data) for each page (i.e., categories) that the users navigated on the website. Similar to the authors, we adopted the Sankey chart to show the relationship between quantitative data (i.e., numbers of labels) and categorical data (i.e. *why*, *what*, and *how*). At the design granularity of *zoom and filter*, we adopted the Heat Map to show a three-leg relationship. We associated a particular UX data with the lean personas and the quadrants. In this visualization, developers and designers can identify, through the color pattern, where UX data is concentrated (i.e., *why*, *what*, and *how*). Móro et al. (2014) work mentions the Heat Map as a visualization that is desired by users but they do not implement it.

5.6 Final considerations

In this paper, we presented a new visual approach to UX requirements gathering from lean personas. The three visualizations were developed to assist developers and designers in searching and exploring UX data available in the textual descriptions of lean personas. The visualizations are intended to guide the elicitation of UX-related requirements following Shneiderman's mantra (1996), i.e. provide the *overview* of data first, make available *zoom and filter* of the data, and observe the *details*. In this way, we proposed a funnel-based approach for data exploration. By connecting the three visualizations, we provided ways of exploring the data gradually throughout the visualizations. We labeled the lean personas descriptions by using Hassenzahl's UX levels. It aimed to give meaning to the UX data available since qualitative data can introduce subjectiveness to the requirement elicitation. We evaluated the three visualizations with 20 participants with regard to the interpretation of the visualizations.

Overall, our findings revealed that the three visualizations provide easy ways of identifying and finding UX data available in the lean personas. Nonetheless, the interpretation evaluation revealed the need to add some kind of help that aids the better understanding of the UX level which was used to provide meaning to the data. The participants' feedback showed us that improvements on the *zoom and filter* features are requested to give more flexibility to users' search and find information. We also presented a method contribution by using Nested Model (MUNZNER, 2014) and Design Activity Framework (MCKENNA et al., 2014) in a complementary perspective for building the visualizations. This same combination of approaches can be replicated in other designs of the visualizations.

In future work, we intend to improve our visualizations to add the features suggested by the evaluation participants. We also intend to conduct new studies in real scenarios (i.e., software projects with designers, requirement analysts, and developers) with different samples of lean personas. Regarding the evaluation, we intend to analyze the usability and efficiency of the visualizations to facilitate the organization and speed up the understanding of the UX information collected through lean personas.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 and grant 309497/2022-1 and

147915/2022-8, Consellho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - Brazil).

Chapter 6

Exploratory study: second round

This chapter presents the article "Mind the Gap Between UX Data and Visualization Proposals: An Approach Emerged From the Grey Literature to Support the Analysis of User Dissatisfaction" (MACEDO; ZAINA, 2024), authored by Maylon Macedo and Luciana Zaina. This article was published in 2024 at the Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais (IHC), where it received an award in the innovative ideas track. An extended version of this work was submitted to the Journal of the Brazilian Computer Society (JBCS). This chapter represents a further development within Step B of the overall thesis methodology, presenting a more contextualized study that evaluates visualizations designed to gain insights into problems within a mobile application. The contributions of this work include deepening the understanding of challenges involved in interpreting UX data through visualizations and validating the premises that eventually led to the creation of the taxonomy, especially regarding the difficulty in correlating data without a clear context or predefined structure. This article mentions a "three-leg approach" that is unrelated to the concept of "three-legged approach", well-known in the HCI community. In this article, the "three-leg approach" refers to the relationship between visualization, purposes, and UX data, which is explored by the taxonomy.

6.1 Introduction

Design focused on User eXperience (UX) improvements can create intuitive interfaces to meet user needs and preferences (NORMAN; NIELSEN, 2018). The UX encompasses the quality of human interaction with products and services, going beyond mere usability to include emotional aspects before, during, and after usage (HASSENZAHL, 2018). The evolution of UX has influenced technology development, leading to the emergence of related subjects like UX Research and UX data. UX research plays a crucial role in understanding how users interpret and interact with digital products and services to develop digital products user-friendly and aligned with user expectations (MARTINELLI; LOPES; ZAINA, 2022).

Analyzing UX data generated by the practice with users is crucial to avoid superficial insights in UX research (CONVERTINO; FRISHBERG, 2017). UX data refers to information that allows mapping users' interactions with interactive systems and can be analyzed to enhance user experiences and optimize products (LUTHER; TIBERIUS; BREM, 2020; TONG et al., 2022). The literature points out that UX data is generated during UX research (ZAINA; SHARP; BARROCA, 2021) and is also crucial for fueling analyses that help create successful, user-centered design solutions (KOESTEN; SIMPERL, 2021). Software professionals can apply different techniques to analyze UX data to gain meaningful insights into user behavior and preferences, ultimately leading to enhanced user satisfaction (FRITZ; BERGER, 2015).

Visualizations are a suitable approach to help software professionals break through the barrier of UX data abstraction and obtain relevant insights from the data (BUONO et al., 2020a). The InfoVis area covers aspects related to the design and functionalities of visualizations; this area assists in developing visualizations that are more in line with the users' needs (MUNZNER, 2014). In parallel with the HCI area, the InfoVis area aims to provide the user with visualizations that facilitate understanding of data that can be complex, abstract, and large (WARE, 2012). The use of visualizations has the function of expanding human perception capabilities, facilitating interpretation and decision-making (CARD; MACKINLAY; SHNEIDERMAN, 1999).

Based on the grey literature research (i.e., UX websites and blogs), we selected 144 articles that contained definitions of UX data and visualizations used to explore it; we found three legs that can support UX data analysis: the visualization approach, the purpose, and the data definition. We selected four dissatisfaction reports and constructed five visualizations to explore UX data related to a mobile app for airline tickets. We evaluated the visualizations with 31 software professionals to identify insights about events that triggered the users' dissatisfaction reports.

Our primary concern was to collect the participants' interpretations of the UX data available from the five visualizations to evaluate how well the visualizations helped participants identify aspects of the mobile application that resulted in reports of user dissatisfaction. The participants evaluated the use of visualizations positively, and a good number of participants (i.e., 84%) pointed out that the proposed visualizations could be appropriated to support their daily work.

This study's innovative idea is the methodology to explore UX data using visualizations developed based on a closed-loop study focused on the software professionals. The visualizations and data definitions were developed based on the investigation of grey literature (i.e., practical knowledge published by software professionals, mainly UX blogs); then, they were evaluated by software professionals to understand the suitability of the visualizations to get insights into UX issues and also the applicability of the visualizations in software practitioners work.

The results are considered emerging, as the data used and the research questions discussed are excerpts from a PhD research focused on building a taxonomy on UX data. We expected that with this approach, software professionals, regardless of their affinity with UX, can explore the potential of available data in their daily work, focusing on recognizing available data and applying visualizations to obtain insights that ultimately improve the user experience.

This article is an extended version of an article published in the XXIII Brazilian Symposium on Human Factors in Computing Systems (MACEDO; ZAINA, 2024), which received an honorable mention. Following this recognition, we were invited to submit an extended version, in which we expand our investigation by introducing a new research question, new data from the study, and new discussions and findings. The new research question (i.e., RQ3) focuses on the potential improvements that UX data visualizations can bring to interactive systems. To address RQ3, we conducted a new analysis of the grey literature articles, emphasizing how these visualizations contribute to interactive system improvements. In addition, we included more collected data from the study and performed a further examination incorporating new insights from participant comments. These comments were then compared with the findings from RQ3, allowing us to explore how UX data visualizations can be effectively utilized within software companies. Consequently, the discussion and conclusions were updated to reflect these new insights. Furthermore, adjustments were made throughout the text to align with the expanded analyzes and findings presented in this extended version.

The sections are organized as follows: related work is discussed in Section 6.2; Section 6.3 presents the methodology, execution, and results of the research carried out in grey literature; the visualizations developed for the evaluation are presented in Section 6.4; Section 6.5 presents the organization, apparatus and ethical aspects of the study; Section 6.6 presents the results; the discussion is presented in Section 6.7; and Section 6.8 summarizes the conclusions.

6.2 Related work

Buono et al. (2020a) developed four visualizations to help visualize the paths followed by test participants while navigating websites to perform specific tasks, highlighting areas where users may face difficulties or where usability issues may arise. The goal is to provide a tool for novice evaluators to identify usability problems and understand their causes more effectively during usability evaluations. The results suggest that using visualizations as a tool to explore UX data is relevant. However, the evaluation showed that the visualizations developed focused on only one type of analysis (i.e., navigation data), while common industry practices involve different data sources and analyses.

UXmood (FRANCO et al., 2019a) is a tool for integrating different data types (e.g., audio, video, text, and eye-tracking) into a dashboard with coordinated visualizations synchronized with a temporal slider. The tool allows replay test sessions while offering real-time sentiment analysis feedback, enabling UX specialists to focus freely on specific moments of user tests. The tool aims to help UX specialists solve the problem of synchronizing data from different sources, and it relies on UX professional experience to select what data to see and to do the analyses.

UXSENSE (BATCH et al., 2023) is a visual analytics system that utilizes machine learning (ML) techniques to extract user behavior from audio and video recordings as parallel timelines in a web-based interface. The tool uses ML to extract user sentiment, actions, posture, and spoken words from audio and video recordings; then, it uses the parallel timelines visualization to allow UX researchers to search, filter, and annotate the extracted data. The tool focuses on data preparation and leaves the professional to select the relevant data.

InteracDiff (DITTRICH; HOF; WIETHOFF, 2019a) is a UX visualization tool to interpret and understand UX data after testing and evaluation sessions. The tool uses interactive charts and game-like features to make the data analysis process more enjoyable to users outside the UX research community. The tool proved adequate in tests to help novices digest collected UX data. However, its interactions may present a high cognitive load for professional use.

Each related work discussed above contributes to UX data exploration through visualization techniques and data integration methods. However, they each have distinct limitations. Buono et al. (2020a) tool effectively highlights user navigation issues but is restricted to navigation data alone. UXmood (FRANCO et al., 2019a) offers a comprehensive integration of various data types but focuses primarily on data synchronization rather than UX analysis. UXSENSE (BATCH et al., 2023) focuses on user behavior extraction using machine learning and applying parallel timeline visualizations to explore the data but only caters to professional researchers. InteracDiff (DITTRICH; HOF; WIETHOFF, 2019a) introduces gamified interactions to make UX data analysis more accessible to novices, although it might impose a high cognitive load on experts. These insights underscore the need for a more versatile approach to support software professionals in tasks involving UX data. That approach must integrate diverse data sources, provide both broad and deep analyses, and cater to both novice and expert users.

6.3 UX data in practice according to grey literature

To publish a scientific article, authors must follow a rigorous methodological process, report relevant findings, and undergo peer review. However, software professionals often generate practical insights that remain unpublished in the scientific literature, as the traditional publication process is labor-intensive and unsuitable for documenting smaller empirical contributions (GLASS; DEMARCO, 2006). In contrast, grey literature aggregates valuable yet non-peer-reviewed information from sources such as UX blogs, professional social networks (e.g., LinkedIn), Q&A platforms, and personal pages (ADAMS; SMART; HUFF, 2017). Analyzing grey literature is an effective way to capture the state of practice on a given topic (GAROUSI; FELDERER; MäNTYLä, 2016), revealing how software professionals adapt and apply UX data visualizations in real-world scenarios.

To understand how UX data visualizations support software professionals, we systematically analyzed grey literature, identifying common practices, challenges, and conditions necessary for integrating UX data analysis into the development workflow. We followed the guidelines proposed by Garousi, Felderer e Mäntylä (2019) to ensure methodological rigor in our bibliographic study. Our primary goal was to identify **definitions of UX data, the visualizations used to explore it, and benefits to use these visualizations**. To structure our investigation, we formulated three research questions: *RQ1. What are the visualizations that help developers in the task of exploring UX data?*, *RQ2. What are UX data characteristics, and how is it used in visualization approaches?* and *RQ3 - How can UX data visualizations improve the design of interactive systems?*. The following sections detail the method adopted to conduct a grey literature review and discuss our findings in relation to these research questions.

6.3.1 Conduction

For a grey literature search, the researcher needs to define the search string by an iterative process, first exploring the related works and then defining the relevant terms for the search (GAROUSI; FELDERER; MäNTYLä, 2019). We defined eighteen strings based on the most cited keywords in related works; each search string is formed by two terms (see Table 8). The configuration of two terms avoids using logical operators, as they are not common in internet search engines. Google was chosen as the search base due to its popularity in studies involving grey literature and because it is a general internet search engine (GAROUSI; FELDERER; MäNTYLä, 2019). The defined search strings, the number of results returned, and the number of grey articles selected for the extraction step are presented in Table 8.

The search process returned 2262 grey articles. However, 183 grey articles were unavailable (i.e., broken links or return errors). Each available grey article was downloaded as a PDF and registered in an electronic spreadsheet. For organization, each grey article

ID	Search string	Results	Selected
1	“quantitative UX”	99	14
2	“UX” + “data”	195	35
3	“UX” + “Data Analysis”	130	11
4	“UX data” + “analysis”	195	18
5	“UX data” + “charts”	67	4
6	“UX data” + “dev”	119	4
7	“UX” + “data-driven”	229	27
8	“UX data” + “evaluation”	124	3
9	“UX” + “data extraction”	164	4
10	“UX data” + “information visualization”	28	0
11	“UX data” + “InfoVis”	32	0
12	“UX data” + “patterns”	119	5
13	“UX data” + “platform”	130	6
14	“UX data” + “product development”	135	4
15	“UX” + “data scientist”	183	8
16	“UX data” + “software development”	138	0
17	“UX” + “information visualization”	93	0
18	“UX” + “InfoVis”	82	1
-	Total	2262	144

Table 8 – Search strings and the number of articles returned and articles selected for the extraction stage. The ID also represents the order in which the strings were executed.

received a unique ID respecting the template “*n1-n2*”, where *n1* represents the ID of the search string and *n2* the sequential ID that identifies the result. The selection of articles was carried out by the first author with the support of three exclusion and one inclusion criteria. Table 9 presents the criteria and number of articles excluded after applying each criteria. After applying all inclusion and exclusion criteria, 144 grey articles were approved.

ID	Criteria	Articles removed
CE1	Content is duplicated	406
CE2	It’s just advertising content	241
CE3	Not in textual format (i.e., videos or podcasts)	85
CI1	The grey article needs to describe something related to the use of UX data	1203

Table 9 – Inclusion and exclusion criteria and the quantity removed from the total number of articles by each criteria | CEn means Exclusion, CIn means Inclusion

6.3.2 Extraction

To establish a systematic extraction process, we explored the scenarios and guiding questions proposed by Lam et al. (2012) for evaluating visualization-based approaches to information processing. These questions are categorized into four key scenarios: UWP (Understanding Environments and Work Practices), which focuses on understanding information processing practices; VDAR (Visual Data Analysis and Reasoning), which examines how visualizations support data analysis; CTV (Communication Through Visualization), which assesses the communicative effectiveness of visualizations; and CDA (Collaborative Data Analysis), which evaluates whether visualizations facilitate collaborative work. Based on these scenarios, we selected three questions from Lam et al. (2012) as our "extraction questions" to guide our analysis. The questions chosen are related to the UWP scenario. The extraction questions and their relationship with the research questions are presented in Table 10.

EQ	**Question**	**RQ**
EQ1	What types of visualizations are currently in use?	[RQ1]
EQ2	What data is currently used and what tasks are performed on it?	[RQ2]
EQ3	What is the context of use of visualizations?	[RQ3]

Table 10 – Extraction questions (EQ) adapted from the Lam et al. (2012) proposal. The RQ column represents the intersection between the extraction question and the research questions.

In the extraction step, each approved grey literature article was fully read. We created a structured form to guide the extraction process, which was based on the extraction questions. The form included a dedicated field for each question (i.e., see the questions in Table 10). The extraction process resulted in 165 relevant excerpts used to address the research questions (see Section 6.3). The number of excerpts extracted per question was as follows: EQ1 (53), EQ2 (112), and EQ3 (111). Due to limited space, citations to grey articles will be made using the "ACn" format (i.e., Article Citation), and the bibliographic information of grey articles can be consulted in this electronic spreadsheet.

Figure 18 presents an example of extraction carried out on the grey article AC53. All the extracted excerpts are available in a spreadsheet in Google Drive. Considering the grey literature analysis, we could answer three RQs. We discuss the results per RQs in the following sections.

6.3.3 RQ1. What are the visualizations that help developers in the task of exploring UX data?

The most reported visualization in the grey literature is the table visualization. This visualization format (i.e., table) is easy to read and organizes information into lines and columns [AC1, AC2, AC3]. The table visualization gives rise to other visualizations, such as UX scorecards and Personas [AC70, AC132]. UX scorecards use tables to present user experience metrics over time [AC132], while Personas use this visualization format (i.e., tables) to list user needs, motivations, and behaviors [AC91].

There are visualization models that are present in many tools for analyzing UX data, for example, bubble charts [AC42, AC123], scatter plots [AC34], bar charts [AC123], bubble charts [AC123], and tree diagrams [AC141]. These visualizations are generally

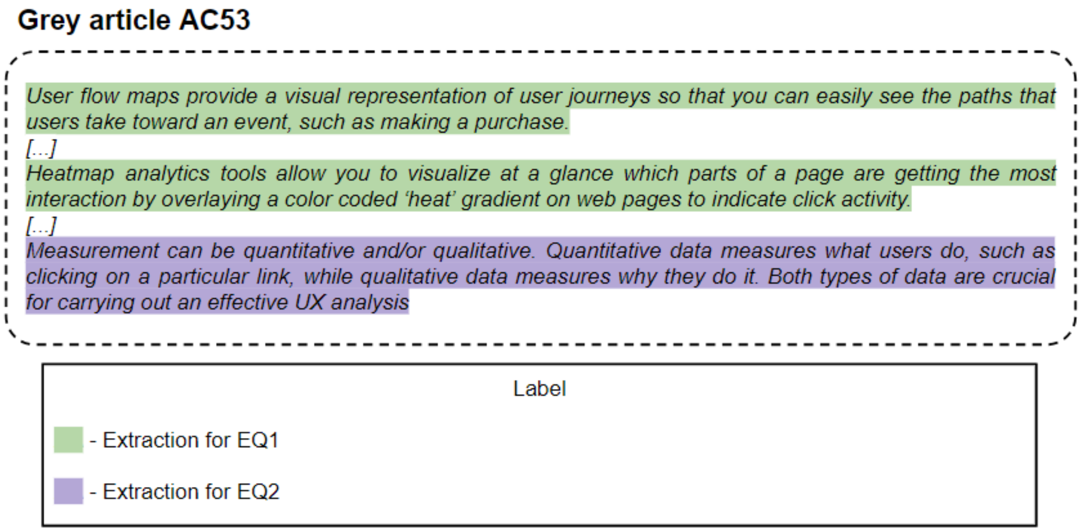


Figure 18 – Example of extractions from grey article AC53, data relating to extraction questions EQ1 and EQ2 (see Table 10).

associated with tasks such as monitoring performance metrics, tracking user navigation in software, identifying behavioral patterns, and analyzing usability test results, among other applications [AC37, AC82, AC101, AC125, AC138].

Considering the variety of visualization formats available, dashboards are the common approach used to aggregate several visualizations into the same system [AC116]. Dashboards are customized according to the professional's needs and generally include visualizations such as heat maps, similarity matrices, and dendrograms [AC82, AC87, AC116, AC120, AC125, AC126, AC138]. The authors of the grey articles point to the heat map as an essential visualization to be added to dashboards, as it makes it possible to obtain an overview of user behavior in the system and identify areas of most significant interest or recurring interaction [AC40, AC42, AC54, AC57, AC65, AC71, AC120].

Some visualizations are customized to serve the purpose of organizing and structuring non-standardized UX data (i.e., like the UXmood tool (FRANCO et al., 2019a)). Non-standardized UX data involves a mix of data that cannot be expressed only with numbers and texts, for example, videos recorded of user interaction with a system, click sequence in a mobile application interface, answers to open questions in questionnaires and user feedback [AC35, AC41, AC63, AC101, AC116, AC125, AC138]. Some examples of popular visualizations to show this data are affinity diagrams, empathy maps, and user flow maps [AC71, AC98, AC131]. A common approach to building a personalized visualization is to associate quantitative and qualitative data to achieve a more comprehensive view of the software's performance in terms of user experience. This approach encourages the engagement of the different professionals involved in the project because personalized visualizations serve multiple purposes (i.e., bring insights into various aspects of the product) [AC19, AC23, AC64, AC136, AC140].

6.3.4 RQ2. What are UX data characteristics, and how is it used in visualization approaches?

UX data is gathered from information collected and analyzed to understand how users interact with software, as in general, approaches that use UX data are focused on user-centered design [AC1, AC40, AC88, AC92, AC102, AC110, AC121, AC127, AC130, AC131, AC139]. To generate valuable UX data, software professionals needs to define a tracking plan that details which events and data should be monitored, how they are aligned with business objectives, and the goals of improving the user experience [AC80, AC99, AC135]. A point highlighted in the grey literature is that UX data must be interpreted in conjunction with other research methods [AC108, AC120, AC129].

The first emerging characteristic of UX data is its general classification, divided into behavioral and attitudinal data [AC25, AC32]. Behavioral data refers to information about users' real behavior and involves interactions with interfaces, task time, abandonment rate, and error rate [AC32, AC82]. Attitudinal data provides information about the user's subjective perception and attitudes, which involves users' opinions, feelings, and perceptions about the software [AC63, AC79].

A relevant approach is the combination of attitudinal data and behavioral data to mitigate limitations, as attitudinal data helps to understand the user's subjective perception that may not be evident with behavioral data alone [AC25, AC66, AC86]. Some of these approaches that stand out in the grey literature are creating personas and empathy maps to gain insight into user motivations [AC98], assembling a user journey map to identify interfaces that represent bottlenecks and have opportunities for improvement [AC98, AC142], and combine performance metrics with user feedback to understand behavior patterns when using the system [AC92, AC131].

Considering the nature of behavioral data and attitudinal data, both can be qualitative or quantitative [AC16, AC40, AC96]. Quantitative data reveals interaction patterns and trends, while qualitative data helps understand user attitudes' reasons [AC85, AC99]. Qualitative data is related to users' qualities and feelings regarding the product and may include data from user feedback and user interviews [AC57, AC89], usability tests [AC48, AC50], and observations that help understand the user's motivations and expectations [AC85]. Quantitative data is related to numerical and measurable values, allowing a more objective view of user behavior, such as conversion rate, session time, and pages visited [AC99]. A common approach to using UX data is to combine qualitative methods (e.g., interviews, usability testing, and diary studies) with quantitative methods (e.g., analytics, A/B testing, and surveys) to analyze the events that triggered the users' dissatisfaction reports [AC13, AC62, AC77, AC138].

6.3.5 RQ3 - How can UX data visualizations improve the design of interactive systems?

The integration of UX data visualizations into the design of interactive systems facilitates a data-driven methodology characterized by continuous iteration and validation. This approach ensures that development strategies remain aligned with end-user needs, fostering refinement throughout the design process [AC2, AC80, AC142]. By systematically collecting UX data over time, professionals can employ metrics to validate development progress [AC5], generate innovative, user-informed ideas [AC51], maintain alignment between user experience and software professionals objectives [AC53], and mitigate discrepancies between envisioned and actual user interactions [AC13].

Informations gathered from UX data visualization enables the creation of products that are not only aesthetically refined but also highly functional [AC17]. Grey literature articles highlight that the primary motivation behind leveraging UX data visualizations in user behavior analysis is to deepen the understanding of how individuals interact with software [AC51, AC59, AC5, AC54, AC53, AC65, AC133]. Additionally, these visualizations fosters a more empathetic, user-centered design approach by bridging the gap between software professionals and their target audiences [AC23, AC81]. The synergy between UX data insights and the expertise of software practitioners leads to more effective design solutions, aligning products more closely with market demands, thereby benefiting users [AC100].

Furthermore, UX data visualizations enables professionals to explore usability challenges, identify improvement opportunities, and understand user preferences trhought qualitative and quantitative analyses [AC65, AC69, AC64]. Qualitative data can be gathered through methods such as interviews and user testing, and uncovers the motivations behind observed behaviors [AC49, AC19, AC78], while quantitative data can be gathered through key performance metrics like abandonment and conversion rates, and provides insights into behavior patterns [AC4, AC92, AC19, AC13, AC62].

By analyzing both quantitative and qualitative data during the UX research, professionals can compare the effectiveness of different interfaces [AC52, AC38, AC60] and assess the impact of specific modifications on user navigation and overall product performance [AC126, AC120]. These visualizations provide detailed insights into user navigation patterns, viewed content, and in-software actions [AC125, AC37, AC126].

These analyses using both quantitative and qualitative data also contribute to practices, such as performance and success measurement [AC40, AC105, AC132]; understanding user behavior [AC26, AC9, AC131, AC18, AC111, AC76]; optimizing products and user experiences [AC1, AC130, AC139, AC132, AC85, AC100]; identifying usability issues and opportunities [AC4, AC92, AC102, AC42, AC100]; supporting decision-making processes [AC85, AC96]; and supports the development of user personas (e.g., with demographic and behavioral insights) [AC34, AC98, AC65].

UX data visualizations contribute significantly to enhancing the understanding of user experience. They facilitate usability improvements, feature prioritization, and stronger connections between software practitioners and end users [AC25, AC32, AC38, AC66, AC86]. Using UX data visualizations during the design processes requires a user-centered approach [AC120, AC122, AC124, AC129], and combining UX data with other methodologies further strengthens outcomes related to user satisfaction, engagement, and overall experience quality [AC120, AC124, AC129].

6.3.6 Reports of dissatisfaction

When analyzing UX data visualizations discussed in grey articles, we found an implicit development pattern (i.e., the three-leg approach) where the UX data definition is combined with a visualization format to accomplish a specific data analysis purpose. A common purpose of the reported UX data visualizations was to identify insights about events that triggered users' dissatisfaction reports [AC120]. We wrote four reports of dissatisfaction (RD) based on the most common issues in the grey literature (i.e., summarized in article AC120) to support our three-leg approach application to develop the study visualizations.

Our reports of dissatisfaction did not directly represent users' dissatisfaction; they were written to represent dissatisfaction on the part of the management team about the health of the mobile application that was receiving users' dissatisfaction reports. The first report (i.e., RD1) was based on three quantitative metrics that represent the percentage of users who complete a task (i.e., Success rate), the percentage of users who encounter a blocker (e.g., confusing navigation) and failed the task (i.e., Error rate), and average time users took to complete their task. The report was written as: *RD1 - Many people access the mobile application, but few complete the purchase.*

The RD2 report aims to find the issues that users are facing by organizing UX data around data that represents system events. The report was written as: *RD2 - The reason for the decrease in sales after February 2022 is unknown.* The RD3 report relies on tracking search usage and user search terms to find particular trending material to engage new users. The report was written as: *RD3 - Users who like discounts cannot find the mobile application in internet searches.* The RD4 report is focused on analysis that highlights patterns to identify recurring issues and users experiencing identical or close interaction problems. The report was written as: *RD4 - People say they do not understand how the cashback policy works.* Figure 19 shows an example of the extractions from grey literature that were used to develop the reports of dissatisfaction.

6.4 Visualizations to explore UX data

To develop the visualizations, we used the three-leg approach by matching the reports of dissatisfaction (RD) with the purposes discussed in grey literature; then, we could

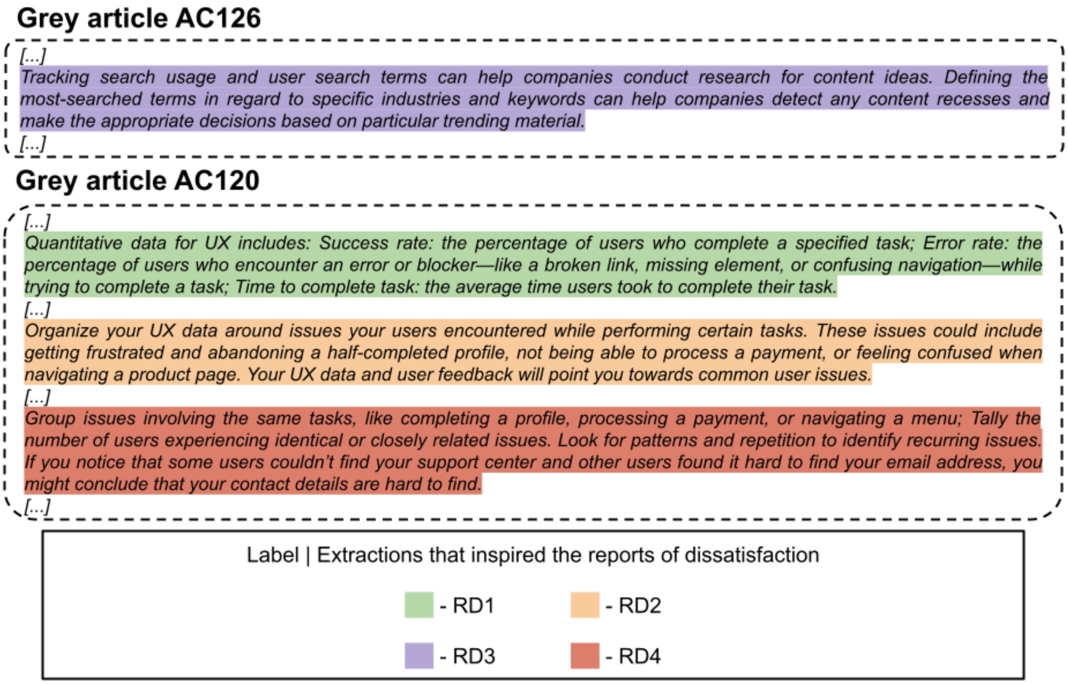


Figure 19 – Example of extractions from grey article AC120 and AC130 relating to reports of dissatisfaction.

select the suitable visualization and the necessary UX data. We generated the UX data (i.e., by using artificial intelligence) considering the common UX data definitions found in grey literature, the reports of dissatisfaction, and the fictional company (i.e., TecFlyFind) scenario developed for the study; for this reason the visualization examples presented in the following sections were built using UX data contextualized for TecFlyFind. All details about TecFlyFind will be presented at Section 6.5.

The visualizations developed for this study were made as medium-fidelity prototypes. A prototype is a candidate solution to a specific problem; the prototype's fidelity level (i.e., low, medium, high) means how close it looks to a working solution from the user's point of view (NORMAN; NIELSEN, 2018). We developed one visualization to assist in analyzing each of the four dissatisfaction reports (i.e., RD1 to RD4) explained in Subsection 6.3.6. Due to the comprehensive scope of the RD1, we decide to develop two visualizations for it. The following sections present the visualizations developed and discuss the rationale for using the visualization format, the objective of the proposed UX data visualization, the purpose of the analysis, and the contribution of each one to reports of dissatisfaction analyses.

6.4.1 Last pages visited

We chose the user flow chart and thumbnails of the application interfaces to develop a personalized visualization that relates the number of users with the last interface viewed before leaving the application. **Rationale:** Visualizing how the users use the product (i.e., user flow visualization) can help identify potential weak areas, providing a starting point for further UX investigation [AC16]. **Objective:** to overview the number of users who abandon the application in each available interface. **Purpose:** to help software professionals identify whether there is an interface that has a higher occurrence of abandonment. **Dissatisfaction report:** to collaborate in analyzing the report *RD1 - Many people access the mobile application, but few complete the purchase*, Through this visualization, we can see that the warning page regarding the cashback policy was the last interface viewed by the majority of users who abandoned the application. Figure 20 presents the visualization made available to participants.

Figure 20 – Visualization representing the total amount of access and what was the last interface seen by users before leaving the APP.

6.4.2 Time in the system

Based on a severity matrix, we built a visualization that groups the number of users who interacted with the product (i.e., the value within the cells) according to the progression achieved in the purchase process (i.e., X-axis) and the time spent on interaction (i.e., Y-axis). **Rationale:** A severity matrix visualization simplifies the analysis of how often each event occurred and enables the researcher to evaluate how bad it was [AC123]. **Objective:** to observe the grouping of users according to the time spent to complete the steps necessary to complete a purchase. **Purpose:** to see the concentration of users who abandon the application without completing the task. **Dissatisfaction report:** to assist in the analysis of the *RD1 - Many people access the mobile application, but few complete the purchase* because through this visualization, we can observe that the interaction time spent in the purchase task is linear in relation to the number of steps to complete the task, suggesting that there are no interaction problems. Figure 21 presents the visualization made available to participants.

Time spent by the user in the system	15 minutes	0	0	0	0	12
	12 minutes	0	0	0	14	3
	10 minutes	0	5	9	19	38
	7 minutes	0	1	23	101	46
	5 minutes	6	2	65	15	23
	2 minutes	2	0	0	0	0
	< 1 minute	2	0	0	0	0
TecFlyFind		20%	40%	60%	80%	100%
		Homepage	List	Cart	Warning	Finishing
		Steps to finalize the purchase				

Figure 21 – Visualization representing the user's progression in the task and the time spent on the APP - Only users in the 100% column completed the task, the others are dropouts.

6.4.3 Interaction with specific user interface

We apply a heatmap over an image representing the homepage interface to determine the areas where users are clicking. **Rationale:** Heatmaps visualization is a popular method that software professionals count on to discern how users spend time; this visualization makes it possible to visualize users' undesirable actions, such as rage clicks, incomplete scrolls, and confusing mouse movements [AC65]. **Objective:** to summarize data on the journey of all users that interacted with the homepage. **Purpose:** presenting the most and least "pressed" areas of the interface through a color scale without presenting the absolute quantity of interactions. **Dissatisfaction report:** to get insights into the report *RD4 - People say they do not understand how the cashback policy works*, because through visualization it is possible to notice that users rarely click on the link with information about the cashback policy. Figure 22 presents the visualization made available to participants.

6.4.4 Calendar of events

We take advantage of a calendar's structure with tag functionality to insert textual information that represents infrastructure events. **Rationale:** Software professionals needs to visualize system-level key performance indicators (e.g., bugs count, usage frequency, up-time) to avoid the narrow vision only in the user data and understand more about the system context [AC59]. **Objective:** overview the relationship between events that affect the mobile application infrastructure and data about the sales department. **Purpose:** to enable software practitioners to directly observe whether variations in access and sales data occur in conjunction with events that affect the product's functioning (e.g., updates, instabilities, reported bugs). **Dissatisfaction report:** to assist in the analysis of the

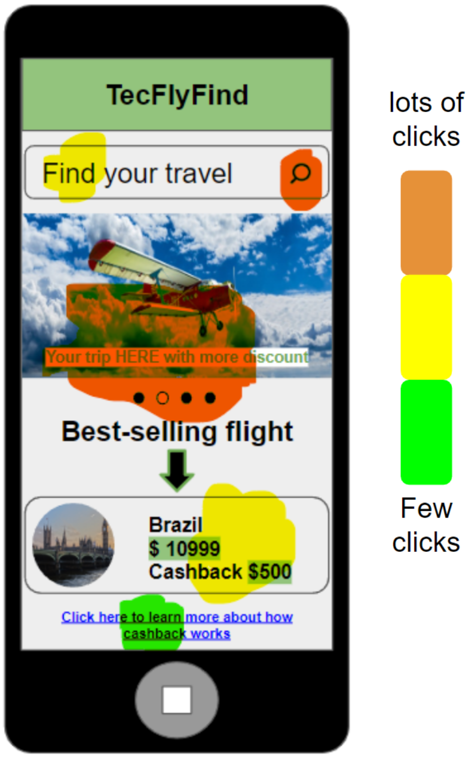


Figure 22 – Visualization showing which areas are most clicked on the TecFlyFind homepage.

report *RD2 - The reason for the decrease in sales after February 2022 is unknown*, as the drop in access and sales numbers occurred after the inclusion of the warning about cashback policy (Figure 25-D). Figure 23 presents the visualization made available to participants.

6.4.5 Most searched terms

We developed a custom visualization composed of a word cloud and a table of baseline terms. **Rationale:** The Search Terms used by users to find your product deserve a more extensive analysis to turn your content into user-centered language [AC30]. **Objective:** to compare the baseline terms used by the team to promote the product with the word cloud highlighting the most used terms for a particular subject in internet search engines (e.g., Google). **Purpose:** to facilitate the analysis of the representativeness of the terms chosen to promote TecFlyFind in relation to the most searched terms on the internet for purchasing airline tickets. **Dissatisfaction report:** to assist in the analysis of the report *RD3 - Users who like discounts cannot find the mobile application in internet searches*, showing that the keywords chosen for promotion are not the ones most used by users. Figure 24 presents the visualization made available to participants.

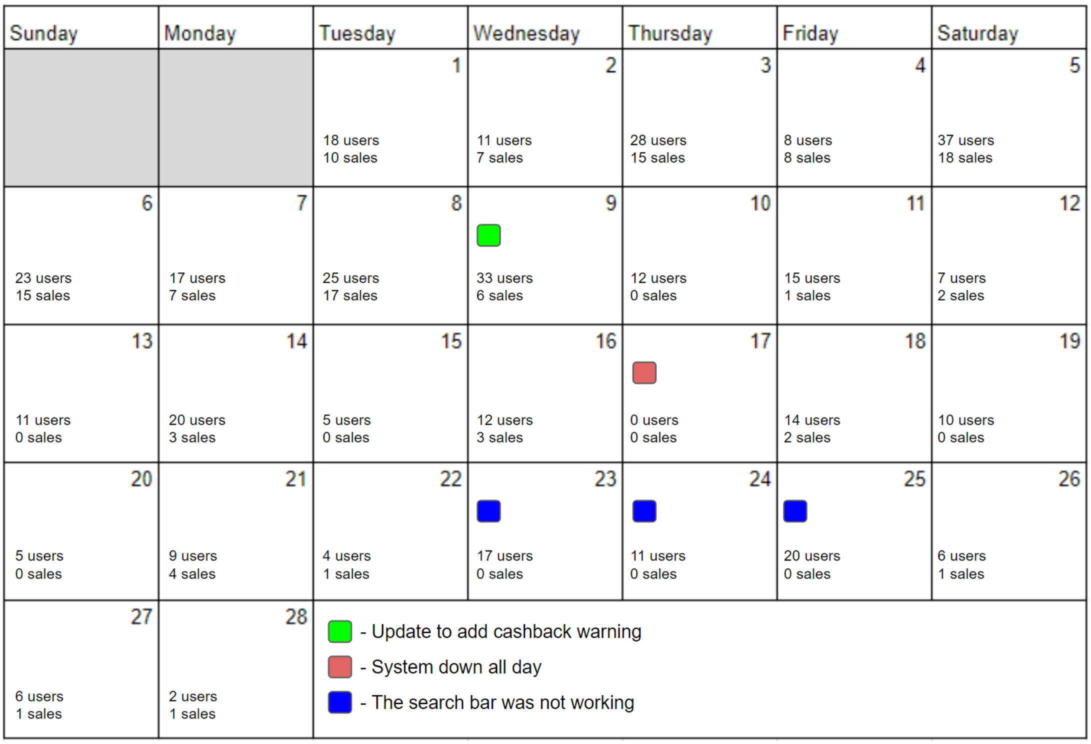


Figure 23 – Calendar-shaped visualization representing the number of accesses, the number of sales and events in relation to the TecFlyFind infrastructure.

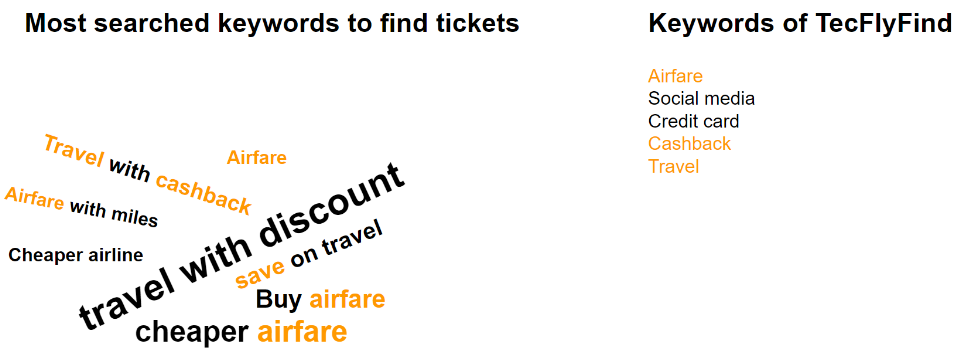


Figure 24 – Visualization representing the relationship between the terms most used on the internet to find airline tickets (word cloud) and the terms used to promote the app (table).

6.5 The study

This study was a proof of concept of the relevance of the three-leg approach (i.e., visualization, purposes, and UX data) to develop visualizations (see Section 6.4) to support the investigation of causes of users' dissatisfaction reports (see Subsection 6.3.6). We aimed to evaluate how well the visualizations helped software professionals identify aspects of the mobile application that resulted in users' dissatisfaction reports. We developed some apparatus to support the study: (i) the airline ticket company scenario; (ii) a fictional mobile app (i.e., TecFlyFind) with app interfaces and usage flow; and (iii) a questionnaire to collect participants' interpretations of the UX data available on the visualizations. The following sections describe the apparatus. We chose the airline tickets domain because it is a common sense domain that avoids the effort of the participant to interpret the application domain.

6.5.1 The company scenario

The fictional company aims to be the largest airline ticket seller on the internet. The company's difference is that it provides cashback on airline tickets through disclosure agreements with the ticket buyer. The target audience is people who are active on social media and want to travel by plane at a cheaper price. The company works as follows: (i) The company offers a mobile application (i.e., named TecFlyFind) to search for airline tickets; (ii) The customer who publishes on their social networks that they purchased an airline ticket using the TecFlyFind app is entitled to cashback; and (iii) After the company analyzes the publication on the social network, the cashback will be credited to the customer's account.

6.5.2 TecFlyFind application

The purpose of the mobile application is to search for airline tickets, inform about the company's cashback policy, and finalize the purchase. The application has six interfaces, with the main flow that directs the user through the process of purchasing tickets consisting of five interfaces; the sixth interface explains the rules of the cashback program, and access is optional. Figure 25 shows the interfaces and usage flow of TecFlyFind.

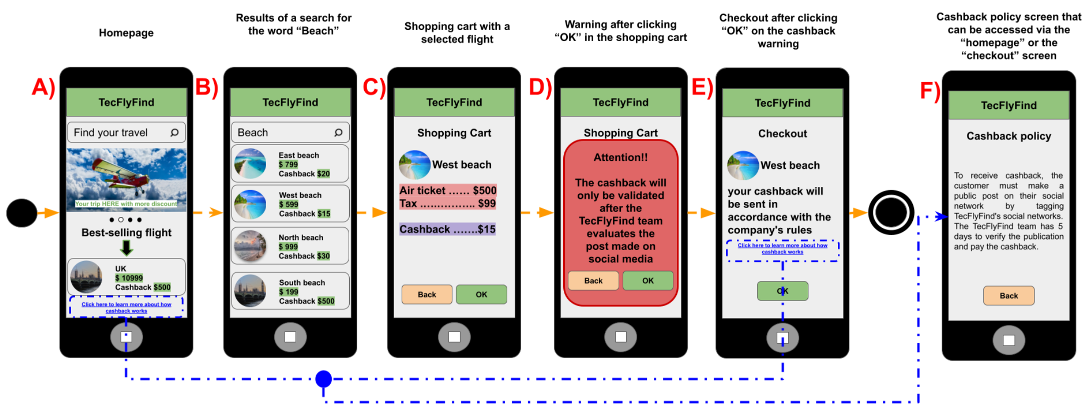


Figure 25 – User flow of TecFlyFind to contextualize participants - Arrows with orange dashes represent the main flow - Arrows with blue dashes and dots represent the flow to read the cashback policy page.

The homepage (Figure 25-A) features an advertising banner, the best-selling flight, and the option to search for a destination. After searching for a destination, the application lists prices and the amount of cashback offered (Figure 25-B). After selecting a flight, the application adds it to the shopping cart (Figure 25-C). After proceeding through the cart, a warning is displayed that the cashback will only be credited after analyzing the publication on the social network (Figure 25-D). After closing the warning, the purchase confirmation interface (Figure 25-E) is displayed. On the homepage (Figure 25-A) and in the confirmation interface (Figure 25-E), there is a link that redirects to the explanation of the cashback policy (Figure 25-F).

6.5.3 Questionnaire

We developed a questionnaire to collect data and guide participants through the study's steps. To reach the largest and most diverse pool of participants, we chose an online questionnaire format. The questionnaire was divided into five sections: (i) Informed Consent Form (ICF) for academic purposes, (ii) profile questions, (iii) presentation of TecFlyFind, (iv) gathering perceptions about visualizations, and (v) collecting participants' feedbacks on the relevance of the visualizations.

The *profile section* included questions about the participant (e.g., age, years of experience, position), the company (e.g., type, segment, year of foundation, whether it allows access to customer feedback), interests related to the study (e.g., InfoVis, UX, Analytics, visualizations), and personal knowledge about UX data and software used to visualize UX data. In the *TecFlyFind section*, we presented information about the fictional company (detailed in Subsection 6.5.1) and the mobile application (described in Subsection 6.5.2).

For gathering data on participants' interpretations of the UX data available in the visualizations, we presented each of the five visualizations (described in Section 6.4) and asked two questions per visualization (ten questions in total). The questions were identical for each visualization: participants were asked to identify, for each visualization, (1) which reports of dissatisfaction could be analyzed using that visualization, and (2) what insights they identified from interpreting the visualization. For the first question, participants had the option to select from the four dissatisfaction reports described in Subsection 6.3.6, with the addition of an option for "none" and a free-text field. In the second question, participants could choose from ten predefined options describing UX-related insights (see Table 11 for the list of options) to capture their interpretation of the data displayed in the visualization.

At the end of the questionnaire, two additional questions were included to assess the relevance of the visualizations for participants' daily work. The first question (discussed in Subsection 6.6.3) was multiple choice, asking participants to identify the most appropriate target audience for each visualization: (a) Not useful; (b) Appropriate for designers; (c) Appropriate for developers; or (d) Appropriate for stakeholders. The final question (discussed in Subsection 6.6.4) was open-ended, allowing participants to describe the

ID	Information
IE1	I see reasons why the user cannot complete the purchase
IE2	I see parts of the system that the user is having difficulty with
IE3	I see opportunities to improve access to the system
IE4	I see functions that can be added to the system
IE5	I see opportunities to improve the usability of the system
IE6	I see who my main users are
IE7	I see reasons for users not to use the system
IE8	I see errors/crashes in the system
IE9	I see opportunities to improve access to system information
IE10	I don't see any useful information

Table 11 – Pre-defined options to collect the participant's perception of what information they were able to extract from the study visualizations.

potential benefits and usefulness of applying the UX data visualizations in their daily work.

6.5.4 Study conduction

The invitation to participate in the study was made through social networks and email lists of the authors' research group network. In total, 31 participants answered the questionnaire. While filling out the questionnaire, the researcher was available via an instant messaging application to answer any participant's questions if necessary. In addition to the questionnaire, participants also had access to the PDF with the TecFlyFind interfaces.

6.5.5 Ethical aspects

Our study considered the regulation document 510/2016 of the Health National Board in Brazil and was approved by the ethical committee of the Federal University of São Carlos (UFSCar) under the process number 68524023.0.0000.5504. About the collection, processing, and availability of data: We prepared an invitation message to the participants informing them who the researchers in charge of the study, the study's aim, and how data collection would happen, and we explained that the data would be strictly used for scientific purposes; we developed and delivered to participants the Term of Informed Consent took into account the recommendations of regulation document 510/2016; we also make clear that our evaluation focused on the visualizations and not the participants' expertise; we provided the authors' contact details to participants and informed them that they could request the data at any time.

Regarding care for possible impacts arising from research and its conduct, we invited participants to the study and made it clear that participation was free and they could withdraw at any time; to mitigate participants' fatigue and stress, the ideal time to complete the questionnaire was thirty minutes. We reduced the time needed to complete the questionnaire by presenting the five visualizations and the ten questions (see Subsection 6.5.3) on the same page of the questionnaire. Additionally, we decided to use only the homepage in the *Interaction with specific user interface* visualization to reduce the study time. The participant was free to choose when to respond to the questionnaire, and we did not make any demands on where the participant should be. We allowed participants to pause their responses and come back to respond later if they felt tired.

Concerns about transparency, auditability of results, dissemination, and use of research products: The Term of Informed Consent clarifies that all the data used in publications would be anonymized and made available to allow the reader to replicate the analysis presented; the raw data would only be accessed by researchers; after the end of data collection all identity information would be erased from the raw data; and all participants can have access to data and research results.

6.6 Results

The majority of responses originated from Brazil, with 30 participants based there and 1 Brazilian participant living and working in Malta, Portugal. The Brazilian participants were from São Paulo (27), Paraná (1), Alagoas (1), and Santa Catarina (1). Our participants were members of software teams and had roles as designers, software developers, software engineers, product managers, and UX researchers. Participants had relevant professional experience, with more than 80% (N = 25) declaring working in a big company and 21 participants with more than four years of experience. Regarding the participants' interest in topics related to the research, it was noted that the average indicates that everyone has a reasonable interest in the topics covered. Table 12 presents the data collected about the participants' profiles. In the following sections, we present the results organized based on the focus of the questions presented in the questionnaire:

6.6.1 Participants knowledge about UX data

As previously discussed (see Section 6.1), the definition of UX data is not yet well defined among software developers. In our evaluation, in the profile questionnaire, participants were asked about their personal knowledge about UX data considering their daily work context (see Subsection 6.5.3). For this question, there were five pre-definitions answers for UX data (i.e., D1 to D5) and an open text field where participants could enter their own definition (i.e., D6). The pre-defined answers were related to the use of the product (i.e., D2), the target audience (i.e., D5), the application's context (i.e., D4), and collection from the user (i.e., D1 and D3). Participants were free to select as many definitions as they wanted. Table 13 presents the answers that participants could choose and the number of votes each received.

We noticed that participants were confident in choosing only the definitions directly related to the product and its use (i.e., D1, D2, and D3). In contrast, less than 15%

Profile			Company		Interest			
ID	Age	Experience	Type	Explore UX data?	InfoVis	UX	Analytics	Charts
P1	27	+ 4 years	Company	Sometimes	3	2	2	2
P2	28	+ 4 years	Company	Always	3	3	3	3
P3	42	> 2 years	Startup	Sometimes	3	4	4	3
P4	27	+ 4 years	Startup	Always	4	4	4	4
P5	28	+ 4 years	Company	Always	4	3	3	3
P6	27	+ 4 years	Company	Sometimes	3	2	1	2
P7	27	+ 4 years	Company	Sometimes	4	4	4	4
P8	29	+ 4 years	Startup	Always	4	3	3	3
P9	34	+ 4 years	Startup	Never	3	3	2	3
P10	26	> 2 years	Company	Always	4	3	3	3
P11	26	+ 4 years	Company	Sometimes	4	4	3	3
P12	22	> 1 year	Startup	Always	3	3	2	4
P13	25	> 2 years	Company	Sometimes	4	3	3	4
P14	30	+ 4 years	Company	Always	4	4	4	4
P15	27	+ 4 years	Company	Never	4	4	3	4
P16	27	+ 4 years	Company	Sometimes	4	3	3	4
P17	27	+ 4 years	Company	Always	3	2	2	2
P18	33	+ 4 years	Company	Always	2	4	2	2
P19	31	< 1 year	Company	Always	4	2	4	3
P20	23	> 1 year	Company	Sometimes	3	4	3	3
P21	32	+ 4 years	Company	Rarely	3	2	1	1
P22	27	+ 4 years	Company	Sometimes	3	3	3	3
P23	27	> 2 years	Company	Always	4	4	4	4
P24	24	+ 4 years	Company	Always	4	4	4	4
P25	33	< 1 year	Company	Always	4	4	3	3
P26	26	+ 4 years	Company	Sometimes	3	4	3	3
P27	27	> 2 years	Startup	Sometimes	4	3	2	3
P28	32	+ 4 years	Company	Sometimes	3	4	4	3
P29	31	+ 4 years	Company	Always	4	4	4	4
P30	30	+ 4 years	Company	Always	3	4	3	4
P31	45	> 2 years	Company	Always	4	4	4	4

Table 12 – Profile data of participants | Label: 1 - No interest, 2 - Little interest, 3 - Reasonable interest, 4 - A lot of interest.

ID	Definition	Votes
D1	Data produced by a UX professional	9
D2	Data collected regarding product use (e.g. logs, screen recording)	26
D3	Data collected through an instrument (e.g. questionnaire)	14
D4	System performance data (e.g. processor and memory usage)	3
D5	Data about the user (e.g. name and age)	4
D6	My own definition	4

Table 13 – Pre-defined options that participants could choose to define UX data.

(N = 4) of participants chose options related to the user and the context of use (i.e., D4 and D5). Only four participants (i.e., P12, P14, P24, and P29) used the field to inform their own definitions of UX data (i.e., D6). Participants P12 and P29 pointed out that UX data can have origins external to the user's interaction with the product, such as the company's business rules [*"P12 - both business and functional data that impact the customer experience user"*], and the user's interest and satisfaction with the use of the product [*"P29 - Set of metrics and indicators, I can mention here CSAT [Customer Satisfaction Score], CES [Customer Effort Score], summed with usage and engagement metrics"*].

According to P14, UX data can be *"user behavior data such as usage tracking"* and can be *"produced with qualitative and quantitative collection methods"*, while the responsibility for generating this data can be attributed to *"designer/researcher/cx/or another area of* *the company"*. P14's answer shed light on the practical definition of UX data and also brought definitions about its origin and those responsible for the data. Finally, participant P24 selected only definition D1 and used the free text field to inform that UX data is *"data collected to improve UX in products"*.

6.6.2 Using the visualizations to get insights

Based on the data from the question about which reports of dissatisfaction could be analyzed using the observed visualization, we observe that most votes align with the key answers (i.e., answers considered correct by the authors). These results suggest that participants were confident in extracting insights into the reports of dissatisfaction that the visualizations were designed to meet. However, we noticed that participants found new meanings for three of the five visualizations presented: the case of *Last pages visited* and the *Time in the system* visualizations, where participants pointed out that they can also help analyze the RD4 report that talks about *people say they do not understand how the cashback policy works*; and the visualization of *Calendar of events* which was highlighted as relevant to analyze the RD1 report that talks about *Many people access TecFlyFind, but few complete the purchase.* Table 14 summarizes the results and presents the key answer for each visualization.

Chart	**RD1**	**RD2**	**RD3**	**RD4**	**None**	**Key answer**
Last pages visited	25	1	2	21	0	RD1
Time in the system	20	0	4	13	4	RD1
Interaction with specific user interface	5	1	3	27	0	RD4
Calendar of events	11	20	3	9	2	RD2
Most searched terms	1	1	29	2	0	RD3

Table 14 – Quantitative of participants' responses regarding which report of dissatisfaction (explained in Subsection 6.3.6) benefited most from the observed visualization.

Considering the question of what insights the participants identified when interpreting the observed visualizations, we noticed in the *last pages visited* visualization that the participants pointed out that it was possible to extract insights about reasons for the user not to complete the purchase (i.e., IE1) and that it was also possible to find opportunities to improve the usability of the system (i.e., IE5) and access to information (i.e., IE9); the *Time in the system* and *Interaction with specific user interface* visualizations contributed to finding opportunities to improve the usability of the system (i.e., IE5), improve access to system information (i.e., IE9), and also helped participants see parts of the system where the user was having difficulties (i.e., IE2).

The visualization of *Calendar of events* was the most highlighted by participants to help in the search for failures and errors in the system (i.e., IE8) and also to clarify reasons for users not completing the purchase and not using the system (i.e., IE1, IE7). Finally, the visualization of *Most searched terms* was highlighted as relevant to discover who the main users are (i.e., IE6) and to find ways to improve access to the system and system information (i.e., IE3, IE9).

6.6.3 Applying visualizations in participants daily work

Considering the context of the company where they worked, participants should list which professionals could obtain meaningful insights using the study's visualizations or whether the visualizations would not be helpful. All the visualizations received at least one vote as not applicable. In this question, participants should choose only one option. The data shows that in three of the visualizations, the participants were divided into two roles as being those most interested in using the visualization as a helpful tool; this division only did not occur in the *Interaction with specific user interface* visualization (i.e., which had the majority of votes directed to designer) and in the *Most searched terms* visualization (i.e., which was understood as useful to stakeholders). Designers and stakeholders were chosen as preferred for the *Last pages visited* and *Time in the system* visualization, while developers and stakeholders were appointed for the *Most searched terms* visualization.

6.6.4 What improvements could UX data visualizations bring to company's work?

As discussed in Subsection 6.5.3, we proposed an open-ended question for participants regarding the improvements that the use of general UX data visualizations (i.e., not just the visualizations presented in this study) could bring to their work in software development. Twenty-one out of the thirty participants provided contributions. We then compared their responses with the findings of *RQ3 – How can UX data visualizations improve the design of interactive systems?* (see Section 6.3). Through a qualitative analysis of participants' responses, we categorized their contributions into four themes (see Table 15 for the themes and their definitions), which highlight improvements in UX data utilization related to the development of interactive systems.

Theme	**Definition**
Understand users	leveraging UX data visualizations to support the creation of more empathetic, user-centered designs, ensuring that products align with actual user needs rather than assumptions.
Self-knowledge about the system	UX data visualizations to tracking feature usage, aligning product vision with user expectations, and assessing system-wide interactions to enhance functionality and business alignment.
Handling development resources	UX data to identifying performance gaps, tracking user engagement, and justifying development investments.
Improve system interaction	focuses on analyzing qualitative and quantitative data to uncover critical touchpoints, assess feature effectiveness, and enhance the overall user experience through data-driven design decisions.

Table 15 – Themes and definitions derived from participant comments about using UX data visualizations in their company.

The first theme was about Understanding users (see participants comments on this theme in Table 16), and the participants' comments aligned closely with the research findings. Several participants emphasized the role of UX data visualizations in understanding users' pain points and behavioral patterns, which resonates with the notion that these visualizations deepen insights into user interactions with software [AC5, AC51, AC53, AC54, AC59, AC65, AC133]. For instance, P2 highlighted that this process helps companies empathize with users, reinforcing the idea that UX data fosters a more user-centered design approach [AC23, AC81]. Additionally, participants P9 and P31 pointed out that UX data helps in better understanding user profiles, aligning with findings that such data supports persona development through demographic and behavioral insights [AC34, AC65, AC98]. Furthermore, P10's comment on prospecting new users suggests that UX data visualizations not only aid in refining existing experiences but also in identifying opportunities for expanding user engagement, which is consistent with the research's emphasis on using these tools to enhance user satisfaction and interaction quality [AC120, AC124, AC129].

ID	**Participant comment**
[P2]	This process helps to *understand the user's pain*, allowing the company to put itself in the user's shoes
[P9]	Better *understand the user's profile*
[P10]	Help to *prospect new users*
[P31]	help us to better *understand the profile* of those who access the system and we can (or not) associate them with a persona

Table 16 – Participants comments aligned with theme Understand users. Words in italic represents the rationale to be select on the actual theme.

Participants' comments also reinforced the research findings about Self-knowledge about the system (see participants comments on this theme in Table 17). P2 highlighted that UX data visualizations facilitate the understanding of the purpose of the application, which aligns with the idea that systematically collecting UX data helps mitigate discrepancies between envisioned and actual user interactions [AC13]. Similarly, P6 mentioned that such tools provide a better view of the product we sell, supporting the notion that UX data insights enable professionals to refine both functionality and market alignment [AC17, AC100]. P20's remark on the importance of aligning the key words of the business resonates with the research's emphasis on leveraging UX data to assess the impact of modifications on user navigation and overall product performance [AC120, AC126]. Finally, P27's comment about improving product visibility is consistent with findings that UX data visualizations help optimize products, prioritize features, and strengthen connections between software practitioners and end users [AC25, AC32, AC38, AC66, AC86].

We noted that Participants' comments also highlighted the ability of UX Data visualization in helping with Handling development resources (see participants comments on this theme in Table 18), reinforcing key findings from the research. P9 emphasized that these visualizations improve efficiency in managing product processing resources,

ID	Participant comment
[P2]	facilitating the *understanding of the purpose of the application*
[P6]	I believe it would help our company to have a *better view of the product we sell*
[P20]	it would be *good to align the key words of the business*
[P27]	Improve *product visibility*

Table 17 – Participants comments aligned with theme Self-knowledge about the system. Words in italic represents the rationale to be select on the actual theme.

aligning with the idea that UX data facilitates performance measurement and supports decision-making processes [AC40, AC85, AC96, AC105, AC132]. P20 noted that such tools could help developers take action more quickly in relation to a drop in users, which resonates with the role of UX data in validating development progress and ensuring strategies remain aligned with user needs [AC2, AC5, AC80, AC142]. P24 pointed out that using these resources in production systems enhances problem tracking and conversion into actionable tasks, supporting the research finding that UX data analysis enables professionals to compare interface effectiveness and optimize development workflows [AC38, AC52, AC60]. Additionally, P26 highlighted the challenge of tracking everything that happened in large companies, reinforcing the benefit of UX data in facilitating feature prioritization [AC25, AC32, AC38, AC66, AC86]. Finally, P30 emphasized that UX data helps justify the company's investment in user-centered development, which aligns with the research's emphasis on leveraging UX data insights to refine design decisions and improve financial outcomes for stakeholders [AC85, AC96].

ID	Participant comment
[P9]	*Improve efficiency* in the use of product processing resources
[P20]	could help developers *take action more quickly* in relation to a drop in users
[P24]	Using these resources in production systems (or even approval), the problem *tracking process becomes more agile*, especially when converting these problems into actions to be performed by the development teams
[P26]	I thought the idea of the calendar showing everything, even product updates, was a great idea. It would help in many moments here at the company, we have a lot of difficulty *tracking everything that happened* in the products given that it is a very large company
[P30]	They can *justify the company's investment* in user-centered development, with more participatory dynamics that could provide a better financial return for stakeholders and the entire ecosystem

Table 18 – Participants comments aligned with theme Handling development resources. Words in italic represents the rationale to be select on the actual theme.

Finally, the majority of the comments was about Improve system interaction (see participants comments on this theme in Table 19), which also was highlighted in grey literature. Several participants emphasized the importance of UX data visualizations in understanding and optimizing user interaction. P2 noted that these visualizations help improve navigation processes, aligning with research findings that UX data facilitates usability improvements and enhances user experience [AC25, AC32, AC38, AC66, AC86]. P6, P8, and P9 mentioned that UX data enables the identification of points of improvement, failures in the customer journey, and usability problems, supporting the idea that these tools help detect usability challenges and refine system functionalities [AC4, AC42, AC92, AC100, AC102]. Additionally, P10, P11, and P20 highlighted that UX data allows teams to map user interactions, identify critical paths in the system flow, and analyze the relevance of each screen, reinforcing research findings on how UX data visualizations provide insights into navigation patterns and software usage [AC37, AC126, AC125]. P17 and P18 pointed out that these insights reveal how users actually interact with the system, rather than relying solely on self-reported data, which aligns with the importance of combining qualitative and quantitative analyses to assess product performance [AC64, AC65, AC69, AC120, AC126]. Furthermore, P13, P15, and P16 mentioned the role of UX data in understanding past interactions, structuring optimal event sequences, and analyzing screen layout impacts on user decisions, which corresponds to research findings that UX visualizations support iterative product refinement and foster a more user-centered approach [AC120, AC122, AC124, AC129]. Finally, P30 emphasized that by providing clarification on how users interact, UX teams can strengthen or redesign user flows, reinforcing the grey literature's assertion that UX data helps bridge the gap between software professionals and end users to create more intuitive and effective interfaces [AC23, AC81].

ID	**Participant comment**
[P2]	Improve *navigation processes* [interaction with the system]
[P6]	It would make it easier to *identify possible points that could be improved*
[P8]	Identify *points of failure in the customer journey*
[P9]	Detect *usability problems*
[P10]	*Identify critical paths in the system flow*, identify strengths and weaknesses of use on specific screens
[P11]	*Mapping user interaction* with the product
[P12]	We could use it to collect information about the *acceptance and usefulness* of new features in our application
[P13]	I believe that the calendar of events and accesses is interesting for an organization and strategic review of *what happened previously*
[P13]	Interaction and permanence in the system also becomes useful to *discover possible screens or flows* that the user does not access or accesses and ends up leaving because they do not understand
[P15]	The flow would help designers and researchers *think about the best sequence of events*
[P16]	UX team understands *how screen layouts and screen order are influencing* purchases
[P17]	It is good for us to know *how the user actually uses the system*, and not what they answer in questionaires
[P18]	It provides useful *information about user behavior* that can help design screens that are more targeted to the information the user is looking for, and helps with the investigation of the root causes of a problem
[P20]	helps in a very visual way to *understand the complete user flow* to make a purchase and understand exactly what result each flow can lead to
[P20]	It helps to *know which are the most critical parts of a screen*
[P22]	to understand *how relevant each screen is to the user*
[P27]	In *identifying problems* in the functionalities
[P27]	In *analyzing the user's interaction* with the product and reaction to updates
[P28]	It could help in *identifying points of improvement in the system*, from the communication, usability and design aspects, to tracking failures that harmed sales conversion
[P30]	By providing *clarification on how the user interacts*, the design team can strengthen or redesign the user's intention of interaction

Table 19 – Participants comments aligned with theme Improve system interaction. Words in italic represents the rationale to be select on the actual theme.

6.7 Discussion

The first lesson learned that stands out is the participants' need for more confidence in defining UX data. According to reports in the grey literature, all definitions offered to participants are considered UX data; however, participants clearly only felt confident in stating that UX data is only data directly related to the use of the product, indicating a funneled view of the concept of "User eXperience" focusing only on the moment of interaction. Another point to highlight about the definition of UX data is that half of the participants supported the idea that UX data can only be generated and explored through the intervention of a UX professional. At this point, the three-leg approach (i.e., visualization, purposes, and UX data) proves to be a relevant solution because by establishing a purpose, the professional will be able to know the possible visualizations and the necessary UX data.

Participants felt motivated to use visualizations to seek insights into all reports of dissatisfaction provided. This motivation was perceived by the fact that participants consistently pointed out that the *Last pages visited* visualization and *Time in the system* visualization were also suitable for providing insights into the report *RD4 - People say they do not understand how the cashback policy works* (i.e., these visualizations was intended for RD1 report). The motivation of the participants in pointing out that these visualizations were suitable to get insights about the RD4 report is corroborated by the fact that in the question about insights identified in the visualization (see Subsection 6.5.3) participants pointed out that visualization brought opportunities to improve the usability of the system and also improve access to information (i.e., IE5, IE9). Unlike what is pointed out in the grey literature, which states that a specific visualization serves one purpose, this observation shows that a visualization can be multipurpose even if it is not an overview visualization.

Similarly, participants pointed out the usefulness of the *Calendar of events* visualization for the report *RD1 - Many people access the mobile application, but few complete the purchase* (i.e., this visualization was intended for RD2 report). The participants' choice to point out contributions to RD1 is justifiable because when answering the question about insights present in the visualization, participants pointed out that through it, they were able to identify reasons for the user not to complete the purchase and not use the system (i.e., IE1 and IE7). We developed these visualizations to yield insights just for a report of dissatisfaction. However, this observation by participants to notice other contributions embedded in the visualizations shows that visualizations depend not only on the experience of the professional who develops them but also on the experience of the user who will use them.

Based on participants' comments about the exploration of UX data visualization to search opportunities to improve the product, we get valuable insights into the role of UX data visualizations in software development. First, UX data visualizations play a crucial role in understanding users, enabling professionals to identify pain points, refine user personas, and develop more empathetic, user-centered designs. Second, they are instrumental in improving system interaction, as they help detect usability issues, optimize navigation flows, and assess feature effectiveness. Third, UX data visualizations enhance self-knowledge about the system, allowing software practitioners to track feature usage, align product vision with user expectations, and improve internal communication about system performance. Finally, they support handling development resources by facilitating feature prioritization, validating design choices, and justifying investments in UX-driven development.

When exploring the study visualizations, participants felt confident in extracting insights to investigate the events that triggered the reports of dissatisfaction, including making recommendations that the visualizations would help analyze other reports of dissatisfaction and that they would be suitable for the work of other professionals such as designers, developers, and even stakeholders. This scenario suggests that software professionals (i) are aware of the existence of UX data but do not have the skills to characterize them; (ii) recognize that through visualizations of UX data, it is possible to obtain relevant information about product improvements; however, they trust only UX professionals to manage the data and build the visualizations; (iii) understand that UX data can yield varied insights according to the responsibilities of each position within the software development flow; however, they do not feel like protagonists in the process of collecting, managing, and analyzing UX data; and (iv) feel motivated to explore UX data in search of opportunities to improve the product; however, they are not always part of a company with a culture of access to UX data.

6.8 Conclusions

This study explored the relevance and applicability of the three-leg approach (i.e., visualization, purpose, and UX data) by analyzing UX data visualizations with 31 software professionals. The evaluation was grounded in the analysis of 144 grey literature articles discussing the practical use of UX data visualizations in software development practices. The research revealed that software professionals often have a narrow view of UX data, typically restricting it to product usage metrics generated by UX professionals. However, participants recognized the broader potential of the proposed visualizations, highlighting their applicability across various UX-related analyses. They acknowledged the adaptability of these visualizations, noting multiple utilities beyond their initial design purposes; Reinforcing the importance of the software professionals experience and interpretation in extracting meaningful insights from visualizations.

Our findings shed light on four key areas that professionals could benefit from using UX data visualization (i.e., Understanding users, Improving system interaction, Enhancing self-knowledge about the system, and Handling development resources) and the importance of fostering confidence among all software practitioners to effectively collect, manage, and analyze UX data. Participants acknowledged that UX data visualizations can bridge the gap between software professionals and end users, fostering a more user-centered design approach. In future work, we intend to further refine the three-leg approach by expanding the scope of visualizations and exploring their applicability in diverse domains. By bridging the gap between theoretical visualization methods and real-world applications, we aim to provide software professionals with practical tools (e.g., Taxonomys) that enhance user satisfaction and drive more effective UX-driven development processes.

6.9 Final Considerations

This chapter concluded its investigation into the relevance and applicability of a three-leg approach (i.e., visualization, purpose, and UX data) through an evaluation with software professionals. The study found that participants often hold a confined view of UX data, primarily associating it with direct product usage and data generated by UX specialists. Despite this, participants recognized the broader utility of the proposed visualizations, frequently identifying insights beyond their initial design objectives and highlighting the importance of the professional's experience in interpreting data. The findings emphasize that UX data visualizations play a crucial role in enhancing user understanding, improving system interaction, augmenting self-knowledge about the system, and optimizing development resource allocation. The research also revealed a gap in confidence among software professionals regarding collecting, managing, and analyzing UX data, despite their motivation to explore it for product improvement. This work reinforces the need for structured approaches to bridge the conceptual and practical divide in UX data handling.

Acknowledgements

We thank the support of grant 309497/2022-1, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - Brazil), and grant by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Chapter 7

Technology acceptance of UX data tools

This chapter presents an article titled "Handling UX Data through visualization tools: an investigation of technology acceptance with tech-users", authored by Erik da Silva, Maylon Macedo, and Luciana Zaina. This article was prepared for submission to Interacting with Computers (IwC) and is an English synthesis of a work developed in Portuguese in collaboration with a scientific initiation student. This work constitutes Step C of the overall thesis methodology. Its objective is to delve into the challenges and factors influencing technology acceptance within the UX data context, specifically exploring the effectiveness and acceptance of data visualization tools from the perspective of software development professionals. The contributions of this work include providing a broad context for the acceptance challenges that tools in the UX data area might face, with its conclusions offering insights into the practical barriers to adopting artifacts designed to assist with UX data.

7.1 Study design

This study, conducted using a controlled experiment setting, aimed to evaluate the acceptance of UX data visualization tools from the perspective of software developers. The research focused on understanding how these tools impact developers, taking into account both task performance and their perception of tool use. The study was structured into four main steps: planning, execution, analysis and conclusion.

During the **planning** step, following the guidelines of controlled experiment methodology, was defined the study objective, apparatus (i.e., tasks, tools, datasets), and formulating hypotheses. The metrics to be used in data analysis were also established to answer the previously raised research question. The **execution** step involves the practical application of the previously defined procedures.

The **analysis** step involved evaluating the data collected during the study, focusing on acceptance of the UX data visualization tools. Acceptance was measured using the Technology Acceptance Model Version 3 (TAM3) questionnaire, which addressed issues such as usability, satisfaction, and overall experience. Furthermore, statistical tests were applied to assess the significance of differences between the groups and tools, with the aim of confirming or refuting the hypotheses.

Finally, in the **conclusion** step, the study results were properly interpreted, allowing a understanding of the implications of UX data visualization tools in terms of acceptance. The limitations of the study were also acknowledged, focusing on aspects that could affect the study's replicability, such as uncontrolled variables and possible flaws in participant sampling.

Planning the experiment required prior definition of several crucial aspects. Initially, the specific research context was defined, establishing the scenario for the tool's application. Hypotheses were then formulated, along with the definition of metrics for their evaluation.

Hypotheses	**Metrics**
H_{01} - There is no statistically significant difference between the acceptance perceived by developers using the Looker tool compared to Amplitude.	Regression analysis by construct
H_{A1} - There is a statistically significant difference between the acceptance perceived by developers using the Looker tool in relation to Amplitude.	

Table 20 – Hypotheses and metrics defined for the study.

7.2 Exploring Visualization Tools

Choosing the tools to use is an important step in the study, as these tools enable the creation of various visual representations for UX data. There are many data visualization tools available on the market, but to facilitate subsequent analysis, we chose to study a limited set.

The methodology of this study involved surveying UX data visualization tools based on three main criteria: popularity, usefulness, and suitability for the study objectives. The objectives were to identify the main differences between the selected tools and to analyze their acceptance from the perspective of technical users who used them to create UX data visualizations.

The tools were classified into two categories: **generic tools**, which stand out for their versatility and adaptability to different contexts of use, and **specific tools**, which have more limited functionality and are geared toward specific objectives (e.g., as user retention and conversion rate analysis). This classification enabled a comparative analysis between the two categories of tools, highlighting the strengths and weaknesses of each.

After an initial survey of nine tools (see Table 21), a final selection was made based on criteria that would ensure the study efficiency and quality. The tools cost was considered, with preference given to free options to facilitate the study execution; the variety of chart types and features available for comprehensive visual analysis; and the ability to integrate with various external data sources, since UX data would not be collected directly by the tool.

Name	**Category**	**Main features**	**Link**
Amplitude	Specific	Visualization templates, product analytics, campaign reporting, and support for various data types and integration.	<https://amplitude.com/>
Google Analytics	Specific	Variety of charts for visualization, event management, and conversion analysis.	<https://analytics.google.com/>
Heap Analytics	Specific	Analysis charts, integration with other services, data history and SSO.	<https://www.heap.io/>
Hotjar	Specific	Focused on heatmap analysis, automatic data capture and user recording.	<https://www.hotjar.com/>
Kissmetrics	Specific	Used for analysis and tracking of human behavior.	<https://www.kissmetrics.io/>
Looker Studio	Generic	Multiple data sources and charts, integration with Google tools.	<https://lookerstudio.google.com/>
Mixpanel	Specific	Flow analysis and user segmentation	<https://mixpanel.com/>
Plausible	Specific	Data analysis and retention, data protection following European regulations.	<https://plausible.io/>
Tableau	Generic	User visualization and analytics dashboards.	<https://www.tableau.com/>

Table 21 – Main tools raised in the study.

Generic tools are versatile and allow you to create charts for various types of data, applicable to different contexts. Examples include Looker Studio and Tableau, which offer flexibility for a variety of analyses, such as market monitoring. Their main advantage is adaptability to various scenarios, but they may not offer specialized features for more complex data.

Specific tools, on the other hand, are designed to meet specific analytical needs, such as analyzing user retention or conversion rates. Examples include Google Analytics and Mixpanel, which are highly effective for specific metrics but have limitations when applied to other contexts.

After exploring and testing each tool's offerings, only two of the nine initial tools were selected for use in the study. **Looker** was chosen from the generic tools primarily because of its flexibility in generating various types of graphs and its ability to connect to multiple data sources. **Amplitude** was selected as the specific tool because of its availability of a free plan with adequate features and its ease of integrating pre-existing data, eliminating the need for additional data processing.

Among the generic tools, Tableau was not selected because during the testing period it proved difficult to integrate with external data from a separate database. Google Analytics and Hotjar were not chosen because they are more useful in contexts that did not fit the study's purpose, either because they do not allow analysis of external data or because they are focused solely on heatmap analysis. Tools such as Kissmetrics and Plausible, despite their promise, were not chosen because they only offer paid plans. Finally, Mixpanel was one of the tools initially considered for the study, but we decided not to use it due to significant difficulties in integrating external databases.

7.3 Databases and Tasks

Since the study did not include UX data collection, a dataset was chosen from the public databases available in large repositories, such as the UC Irvine Machine Learning Repository[1] and Kaggle[2]. Initially, five databases containing UX data were selected (see Table 22). Of these, only two were chosen as final: Clickstream Data for Online Shopping (i.e., used as main database) and Customer Satisfaction in Airline (i.e., only used as a training for participants).

Dataset	**UX data**	**Link**
Clickstream Data for Online Shopping	Contextual, temporal and interaction data	<https://archive.ics.uci.edu/dataset/553/clickstream+data+for+online+shopping>
Student Performance	Demographic data	<https://archive.ics.uci.edu/dataset/320/student+performance>
E-commerce Customer Behavior	Demographic data and user preferences	<https://www.kaggle.com/datasets/uom190346a/e-commerce-customer-behavior-dataset>
Women's E-Commerce Clothing Reviews	Demographic data	<https://www.kaggle.com/datasets/nicapotato/womens-ecommerce-clothing-reviews>
Customer Satisfaction in Airline	Demographic data e emotional data	<https://www.kaggle.com/datasets/yakhyojon/customer-satisfaction-in-airline>

Table 22 – Main databases analyzed.

The selection of these databases was guided by the criterion that the data should relate not only to UX, but, above all, to the interaction between the user and the system. This would ensure that relevant insights about the system would be obtained through the visualizations to be generated. To identify the most suitable database, the fields in each were thoroughly analyzed, aiming to understand all the details surrounding the data, interpreting how they are used within the context of the database and how they can be useful in building the visualizations.

Furthermore, tests were performed on the previously selected tools (i.e., Looker and Amplitude), using these same databases, to evaluate their compatibility and performance. Therefore, databases such as Student Performance, E-commerce Customer Behavior, and Women's E-Commerce Clothing Reviews were discarded because they did not provide relevant information about user interaction or useful data for UX visualization. This criterion was essential to facilitate the analysis and interpretation of the results.

[1] <https://archive.ics.uci.edu/>
[2] <https://www.kaggle.com/datasets>

Scenarios were created to allow study participants to explore the tools' functionalities while building visualizations. These scenarios, connected to the data types, contextualize what should be done to achieve a defined objective. These stories were transformed into two specific tasks (see Table 23) to be performed by study participants during the visualization construction stage.

Task	Task 1	Task 2
Description	Consider you're a developer working for an online clothing store, using the provided "Clickstream data for Online Shopping" database. Build a visualization that allows you to see a customer's navigation within the store (i.e., the sequence of pages they accessed within a session). The visualization should be able to show the percentage of customers who accessed a given page, displaying the flow of each page.	You've just joined a development team for a clothing e-commerce site. You've been asked to use the provided "Clickstream data for Online Shopping" database to build a visualization that identifies the position on the screen where the photo of the item of clothing the customer is clicking on is located. The visualization should be able to show the order of clicks in each position and which positions were most frequently clicked.
Data that may be useful	Event	Event
	Session ID	Session ID
	Order	Order
	Page	Location
	Duration (sec)	Model photograpy
Requirements	Show the percentage of customers who accessed a specific page, displaying the flow of each page.	Show the order of clicks in each position, and show which positions were clicked the most.

Table 23 – General information for Tasks 1 and 2.

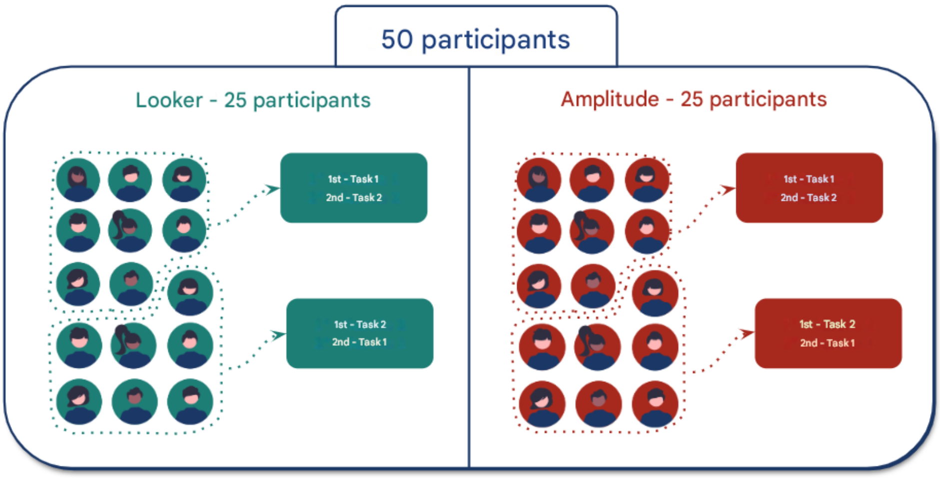
7.4 Execution

A profile questionnaire was administered to collect participant data, aiming to characterize the sample group and better understand their experiences and skills. The questionnaire included information about their workplace, areas of interest, roles held, length of experience at the company, and their level of knowledge of UX data and data visualization tools. Furthermore, the Informed Consent Form was presented along with the questionnaire, ensuring that participants were aware of the research objectives, their voluntary participation, and their rights.

Fifty professionals with experience in software development participated in the study. All participants were students enrolled in courses at the Federal University of São Carlos (UFSCar), and the invitation to participate was disseminated internally within the UFSCar academic community. The study was conducted in person at UFSCar, in two separate rooms, one for each tool. The initial division of participants (see Figure 26) was carried out into two distinct groups, based on the alphabetical order of their names, to ensure heterogeneity between the groups.

Group A, composed of 25 participants, used the Looker tool, while Group B, also with 25 participants, used Amplitude. Afterwards, participants were directed to separate rooms, where they participated in a warm-up exercise. The warm-up was different from the study tasks (i.e., Task 1 and Task 2) and used a separate dataset from that used in the main tasks, to ensure that all participants had a consistent level of initial knowledge of the assigned tool.

Figure 26 – Separation of participants into groups and subgroups.



After completing the warm-up task, participants completed the profile questionnaire and began the visualization construction stage, following the tasks established in the study plan. To minimize potential bias in the results, the groups in each room were subdivided into two subgroups. The first subgroup began constructing the visualizations with Task 1, while the second subgroup began with Task 2 (see Table 23).

At the end of the study, all participants completed the TAM questionnaire, which assessed their level of acceptance of the tool used on a scale ranging from 1 to 7. Participants recorded their final visualizations of each task on a specific form. In total, the study lasted two hours, resulting in the production of 100 visualizations, 50 for each tool (i.e., Looker and Amplitude), with 25 visualizations per task.

7.5 Data Analysis

To answer the question *(RQ1) Which tool has better acceptance?*, a regression analysis was performed. Regression analysis is a statistical method used to investigate the relationship between variables, allowing us to understand whether this relationship is positive or negative (CRUZ et al., 2022). This analysis helped identify which questions contributed most significantly to explaining the variability in the constructs. The coefficient of determination (R^2) was used as a measure of model fit, indicating the proportion of the constructs' variance that was explained by the questions (CRUZ et al., 2022). Therefore, the variance was classified positively or negatively based on the trend of the results obtained on the question scales. An electronic spreadsheet was used to store and organize all the collected data.

7.6 Study limitations

The study's validity was assessed and addressed according to the guidelines established by Wohlin et al. (2012). Regarding internal threats, the issue of participant fatigue stands out, which was mitigated by limiting the duration of the study session to a maximum of two hours. Furthermore, participants' lack of experience with the tools used was mitigated through a warm-up session, allowing them to familiarize themselves with the environment before beginning the main tasks.

Regarding the threat to conclusion validity, this was addressed through the use of appropriate metrics, ensuring that the analysis of the results was conducted systematically. The different metrics were analyzed to provide a more complete understanding of the results without negatively affecting the validity of the conclusions.

Finally, to address threats to external validity, all participants selected for the study had prior software development experience, minimizing the risk of results being limited to a specific group of individuals without the necessary expertise. This approach aims to increase the generalizability of the results.

7.7 Results

The 50 participants were mostly between 21 and 23 years old. Most participants (54%) work at established companies (i.e., more than 5 years in the market), where 64% have up to 2 years of experience and 68% of these companies work specifically with software development. The participants' understanding of the definition of UX data predominantly (78%) indicates that it refers to data collected regarding product use. Finally, it is worth highlighting the fact that only 8% (n = 4) of them were familiar with Amplitude and 16% (n = 8) were familiar with Looker.

When analyzing the responses obtained in both Looker's TAM and Amplitude's TAM (see Figure 27 and Figure 28), it is noted, mainly in the first and last questions, that in the case of Looker the values are better distributed along the response ranges. Amplitude presenting more values within ranges that show disagreements, which could indicate a greater tendency of disagreement in relation to the acceptance of Amplitude.

Therefore, an analysis was performed using linear regression, evaluating the impact that each question (see Table 24) had on their respective constructs through the coefficient of determination R^2, angular coefficient β and standard error ϵ (see Table 25, Table 26, Table 27, Table 28, Table 29).

Considering the PU construct (see Table 25), when examining the coefficient of determination R^2 in Looker, the question Q1 presents the highest value (0.817), indicating that approximately 81.73% of the variation in the response can be explained by the model of this specific question. In contrast, question Q3 presented the lowest R^2 value (0.434),

Figure 27 – TAM responses when using Looker.

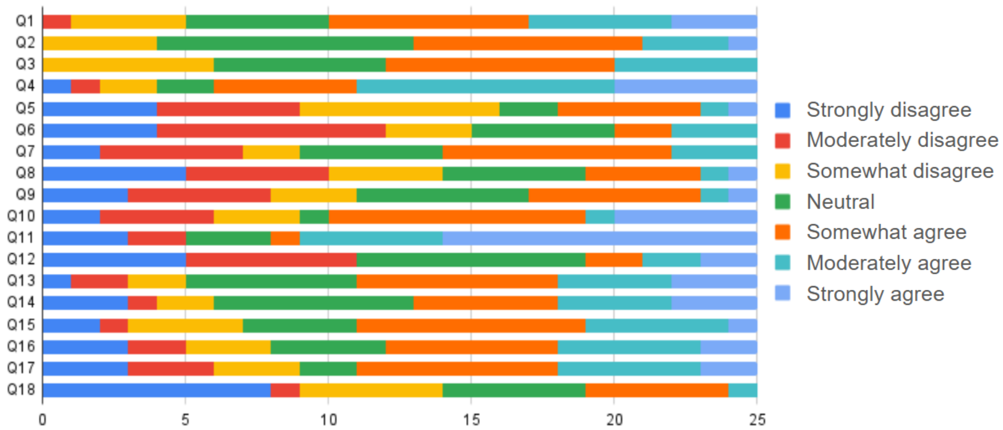
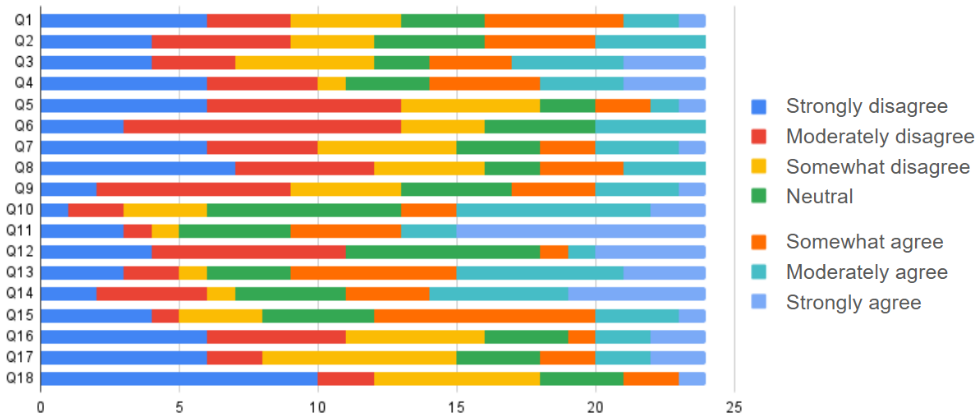


Figure 28 – TAM responses when using Amplitude.



therefore the least explanatory result addressing only 43.4% of the variation. The result suggests that the answers to question Q1, which deals with performance improvement, may be more aligned with what concerns **perceived usefulness** compared to Q3.

For amplitude, conversely, question Q3 was the question that explained the most response variation (0.892), while Q4 had the lowest value (0.746). Regarding the other parameters, the angular coefficient β serves to indicate the strong relationships between the questions and the constructs, while the standard error ϵ suggests the level of precision of the predictions made from these relationships.

While a question may indicate a strong explanatory relationship for the variation in responses to a construct, it is crucial to assess whether this variation manifests itself positively or negatively. In the context of this study, the question with the greatest impact on Looker was Q1, while on Amplitude it was Q3. To investigate the direction of this variation, the measures of central tendency of the most frequently asked scales in the responses, were used to identify the direction (i.e., positive or negative) in which the variation tended, as shown in Table 30.

An analysis of the responses to question Q1, which had the greatest impact on the PU variation for Looker, reveals a tendency for users of this tool to assign a higher score

#	Question	Construct
Q1	Using the system improves my performance in my job	**PU**
Q2	Using the system in my job increases my productivity	
Q3	Using the system enhances my effectiveness in my job	
Q4	I find the system to be useful in my job	
Q5	My interaction with the system is clear and understandable	**PEOU**
Q6	Interacting with the system does not require a lot of my mental effort	
Q7	I find the system to be easy to use	
Q8	I find it easy to get the system to do what I want it to do	
Q9	I have control over using the system	**PEC**
Q10	I have the resources necessary to use the system	
Q11	Given the resources, opportunities and knowledge it takes to use the system, it would be easy for me to use the system	
Q12	The system is not compatible with other systems I use	
Q13	The quality of the output I get from the system is high	**OUT**
Q14	I have no problem with the quality of the system's output	
Q15	I rate the results from the system to be excellent	
Q16	Assuming I had access to the system, I intend to use it	**BI**
Q17	Given that I had access to the system, I predict that I would use it	
Q18	I plan to use the system in the next <n> months	

Table 24 – Questions used from TAM and its associated constructs.

	Perceived Usefulness (PU)					
	Looker			Amplitude		
Questions	R^2	β	ε	R^2	β	ε
Q1	0,817	2,722	0,268	0,804	3,422	0,361
Q2	0,652	3,220	0,490	0,857	3,757	0,327
Q3	0,434	2,533	0,603	0,892	3,275	0,242
Q4	0,707	2,170	0,292	0,746	2,817	0,351

Table 25 – Regression Analysis Parameters for PU.

	Perceived Ease of Use (PEOU)					
	Looker			Amplitude		
Questions	R^2	β	ε	R^2	β	ε
Q5	0,621	2,552	0,416	0,655	2,853	0,441
Q6	0,636	2,636	0,416	0,638	2,871	0,461
Q7	0,677	2,821	0,406	0,859	2,910	0,251
Q8	0,756	2,746	0,325	0,694	2,745	0,388

Table 26 – Regression Analysis Parameters for PEOU.

Figure 29 – Number of responses by question of the Perceived Usefulness construct.

(mode = 5). On the other hand, question Q3, which most influences the variation in PU responses for Amplitude, has a mode of 3. This difference in the most frequent values suggests that the variation for Looker follows a positive direction, while for Amplitude it follows a negative direction (considering that the scale ranges from 1 to 7). This can

	Perceptions of External Control (PEC)					
	Looker			Amplitude		
Questions	R^2	β	ε	R^2	β	ε
Q9	0,477	2,262	0,494	0,615	1,912	0,322
Q10	0,536	2,005	0,389	0,634	2,032	0,329
Q11	0,624	1,919	0,311	0,409	1,268	0,325
Q12	0,303	1,539	0,487	0,008	0,175	0,427

Table 27 – Regression Analysis Parameters for PEC.

	Output Quality (OUT)					
	Looker			Amplitude		
Questions	R^2	β	ε	R^2	β	ε
Q13	0,789	2,437	0,263	0,817	2,430	0,246
Q14	0,703	2,021	0,274	0,825	2,313	0,227
Q15	0,813	2,479	0,248	0,815	2,653	0,270

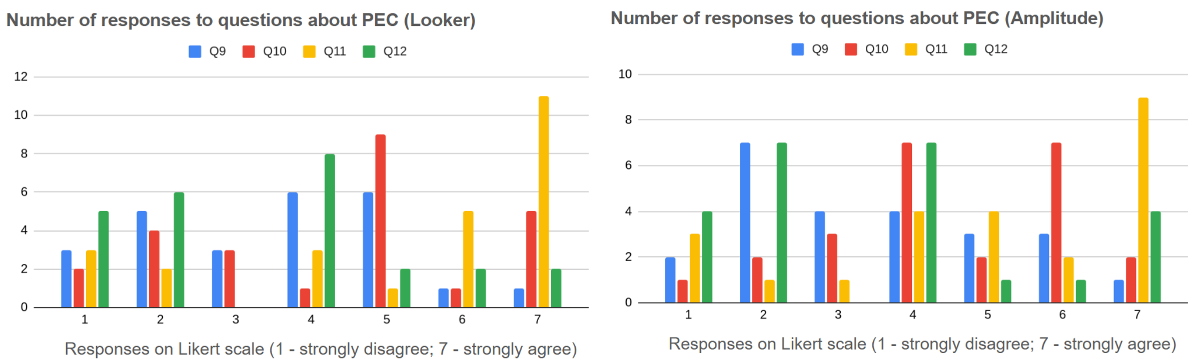
Table 28 – Regression Analysis Parameters for OUT.

	Behavioral Intention (BI)					
	Looker			Amplitude		
Questions	R^2	β	ε	R^2	β	ε
Q16	0,933	2,645	0,148	0,956	2,670	0,122
Q17	0,936	2,565	0,139	0,953	2,679	0,127
Q18	0,719	2,547	0,332	0,819	2,863	0,287

Table 29 – Regression Analysis Parameters for BI.

Figure 30 – Number of responses by question of the Perceived Ease of Use construct.

Figure 31 – Number of responses by question of the Perceptions of External Control construct.



also be evidenced by the average of the most frequent scales for each question in this construct, where it is 5 for Looker and 1.75 for Amplitude.

Construct	Tool	Question	R^2	Mode of the scales	Mean of the scales
PU	Looker	Q1	81,73%	5	5
PEOU	Looker	Q8	75,63%	1, 2 e 4	3
PEC	Looker	Q11	62,35%	7	5,25
OUT	Looker	Q15	81,32%	5	4,67
BI	Looker	Q17	93,65%	5	3,67
PU	Amplitude	Q3	89,24%	3	1,75
PEOU	Amplitude	Q7	85,94%	1	1,5
PEC	Amplitude	Q10	63,38%	4 e 6	3,75
OUT	Amplitude	Q14	82,53%	6 e 7	5,67
BI	Amplitude	Q16	95,63%	1	1,67

Table 30 – Questions of greatest impact by construct.

Figure 32 – Number of responses by question of the Output Quality construct.

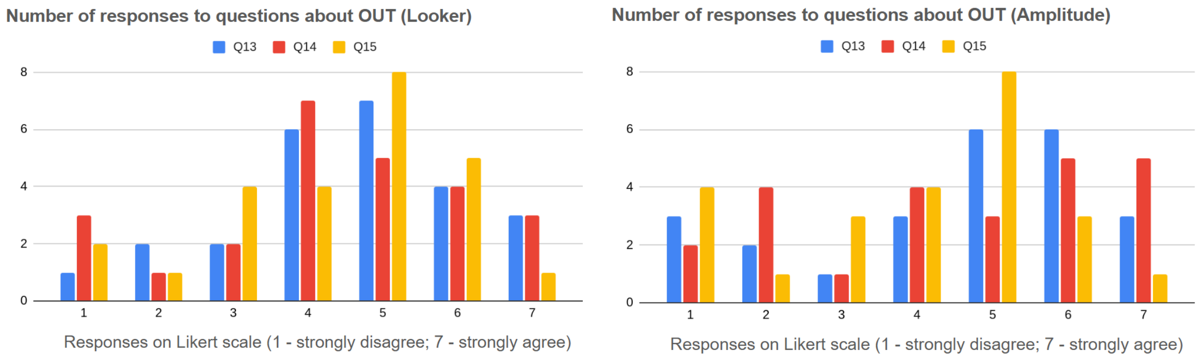
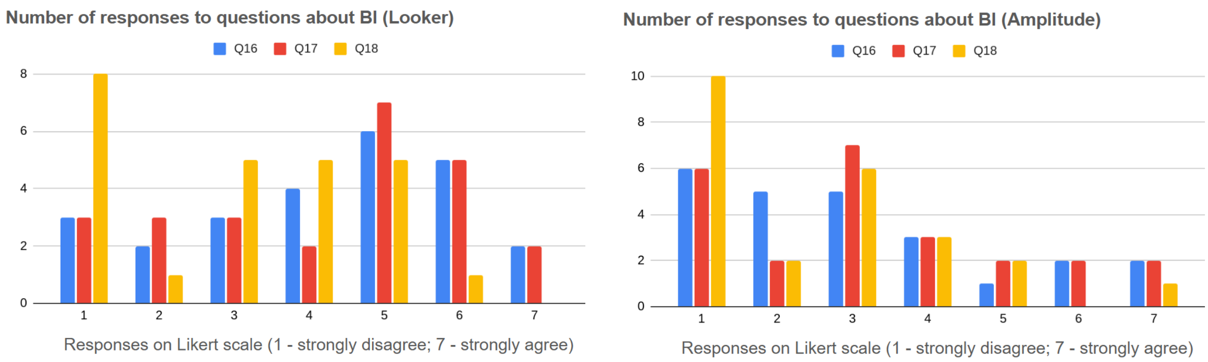


Figure 33 – Number of responses by question of the Behavioral Intention construct.



Although the coefficient of determination is higher for Amplitude (89.24%), when analyzing the magnitude and direction of this impact, it is observed that, in reality, Looker presented positive results in relation to Perceived Usefulness, while the results for Amplitude were more negative. This trend is consistent for all constructs evaluated, with the exception of the output quality (OUT) construct, in which Amplitude performed better, presenting a more positive variation compared to Looker.

Considering the analyzed aspects, it can be stated that Looker received better acceptance from developers, being evaluated as more useful (PU), easier to use (PEOU), controllable (PEC), and with greater intended use (BI) compared to Amplitude. However, as already mentioned, in the Output Quality (OUT) category, Amplitude slightly outperformed Looker. This can be attributed to the fact that Amplitude is a special-purpose tool, which tends to generate more structured results, with more concrete visualizations that do not rely entirely on the user's knowledge to produce good results. In contrast, Looker is a more generic tool, without predefined templates for outputs, meaning that the quality of the visualizations can vary depending on the user's choices.

7.8 Discussion

This study investigated the acceptance of UX data visualization tools among software developers, specifically focusing on perceived usefulness and other related constructs as measured by the TAM3. The findings offer insights into how developers evaluate these tools and highlight the critical role of data understanding in their adoption.

A characteristic of the participant pool was their limited prior exposure to both Amplitude (8% familiar) and Looker (16% familiar), suggesting that their evaluations largely stemmed from their initial interactions during the experiment. Despite this lack of familiarity, a high percentage of participants (78%) demonstrated a clear understanding of UX data as information related to product usage. This foundational knowledge is crucial, as the utility of data visualization tools is inherently tied to the user's ability to interpret the underlying data.

The analysis of Perceived Usefulness (PU) revealed distinct patterns between the two tools. For Looker, the question "Using the system improves my performance in my job" (Q1) exhibited the highest explanatory power (i.e., $R^2 = 0.817$), with responses consistently indicating a positive perception of usefulness (mode and mean = 5). This suggests that developers readily perceived Looker as a tool that could enhance their work performance. Conversely, for Amplitude, while the question "Using the system enhances my effectiveness in my job" (Q3) had an even higher ($R^2 = 0.892$), the modes (3) and means (1.75) of the scale responses pointed towards a negative overall perception of usefulness. This indicates that despite the model explaining a significant portion of the variance in responses for Amplitude, the actual sentiment regarding its usefulness was often unfavorable.

This divergence in perceived usefulness, despite similar explanatory power in some cases, underscores the importance of not solely relying on statistical fit (i.e., R^2) but also examining the direction and magnitude of user perceptions. The overall trend across most TAM constructs (i.e., Perceived Usefulness, Perceived Ease of Use, Perceptions of External Control, and Behavioral Intention) consistently favored Looker, indicating its higher general acceptance among the participating developers.

The exception to this pattern was Output Quality (OUT), where Amplitude demonstrated a slightly more positive variation compared to Looker. This can be attributed to Amplitude's design as a specialized tool, which often provides more structured and predefined visualizations. Such tools may produce more concrete visualizations that do not rely entirely on the user's knowledge to produce good results, as indicated by the study. In contrast, Looker, being a more generic platform, requires greater user input and understanding to generate high-quality outputs. This distinction highlights a key challenge in the adoption of data visualization tools: while specialized tools can offer immediate, high-quality outputs, generic tools demand a deeper understanding of both the data and visualization principles from the user.

The findings suggest that for UX data visualization tools to be accepted by developers, there needs to be a standardized understanding of the data being visualized. While participants generally understood the definition of UX data, the differing acceptance levels between a generic tool (i.e., Looker) and a specialized tool (i.e., Amplitude) imply that merely defining the data is insufficient. For tools like Looker, where users construct their visualizations, a lack of standardized knowledge regarding data structure, metrics, and analytical goals can hinder their ability to generate meaningful insights, thereby impacting perceived usefulness. Even with specialized tools like Amplitude, while the output might be clearer, a deeper understanding of the underlying data still enables more effective interpretation and application of the visualizations.

This study provides an initial indication that fostering a more uniform and comprehensive understanding of UX data principles among developers, beyond just a conceptual definition, could be instrumental in improving the acceptance and effective utilization of both generic and specialized UX data visualization tools. Such standardization could empower developers to better leverage these tools, regardless of their specific design paradigms, ultimately leading to more insightful data analysis and improved product development.

7.9 Conclusions

This study undertook a comparative analysis of Looker and Amplitude, two data visualization tools, assessing their acceptance among 50 software development professionals tasked with constructing UX data visualizations. The methodology employed a controlled experiment design to systematically evaluate participant perceptions and interactions with each tool.

The findings indicate differences in user acceptance, with Looker receiving more positive evaluations across key acceptance constructs of the TAM3, including Perceived Usefulness, Perceived Ease of Use, Perceptions of External Control, and Behavioral Intention. It is important to acknowledge that even in the sole construct where Amplitude exhibited superior results, namely Output Quality, the observed differences between the tools were minimal. This outcome highlights that while Amplitude's specialized nature may offer benefits in specific output clarity, it did not translate into a widespread preference across other dimensions of acceptance.

The preference for Looker suggests that its intuitive interface and perceived ease of use affects in its higher acceptance. As a general-purpose visualization tool, Looker appears to enable professionals, even those with limited prior exposure to the specific tool, to construct effective visualizations with relative efficiency. These aspects, corroborated by the study's results, reinforce the notion that Looker's simplicity and versatility can make it a highly suitable option for many software development teams addressing UX data visualization needs.

Furthermore, the study implicitly highlights a critical underlying factor influencing tool acceptance: the user's foundational understanding of the data itself. Despite a high percentage of participants (78%) understanding the general definition of UX data, the varying levels of perceived usefulness and output quality between a generic tool like Looker and a specialized tool like Amplitude suggest that a deeper, more standardized knowledge of UX data metrics, structures, and analytical objectives is paramount.

For generic tools, this knowledge empowers users to create meaningful visualizations from diverse datasets. For specialized tools, while they may offer pre-configured solutions, a robust understanding of the data still maximizes their utility and the interpretation of their outputs. Therefore, beyond the features and usability of the tools themselves, fostering a more uniform and comprehensive understanding of UX data principles among developers is likely to be a significant determinant in the successful adoption and effective leverage of any UX data visualization solution. This suggests that future efforts aimed at improving the use of such tools should also consider initiatives to standardize data literacy within development teams.

7.10 Final Considerations

This chapter concluded its comparative analysis of Looker and Amplitude, investigating their acceptance among software development professionals for UX data visualization tasks. The study, conducted via a controlled experiment settings, revealed a general preference for Looker across various technology acceptance model constructs, including perceived usefulness and ease of use, despite participants' limited prior exposure to both tools. While Amplitude, a specialized tool, showed marginally better output quality, this did not translate into overall higher acceptance compared to Looker's versatility and perceived simplicity. A key insight from this research is that, beyond the tools' features, the user's foundational understanding of UX data (i.e., its metrics, structures, and analytical objectives) is critical for effective utilization. The findings suggest that fostering a more comprehensive and standardized understanding of UX data principles among developers could significantly enhance the adoption and impactful use of both generic and specialized UX data visualization tools.

Chapter 8

Grey Literature Review and Taxonomy

This chapter presents the article "Beyond Raw Data: A Taxonomy for UX Data Exploration to Support the Design of Interactive Systems" (MACEDO; ZAINA, 2025), authored by Maylon Macedo and Luciana Zaina. This article is forthcoming in 2025 in the International Journal of Human–Computer Interaction. This work represents Step D of the overall thesis methodology. Its objective is to describe how the taxonomy emerged as a central artifact of the research, specifically detailing the grey literature investigation into UX data usage in practice and proposing the taxonomy based on its three core components: objectives, purposes, and UX data characteristics. The contributions of this work include practically confirming the emerging conclusions from previous chapters regarding the unclear definitions and empirical reliance of UX data. It also demonstrates that a visualization framework alone would be less effective without first resolving the fundamental problem of conceptual clarity. The taxonomy, therefore, offers a structured approach to bridge this gap, organizing and navigating these definitions and practices to meet the real needs of professionals.

8.1 Introduction

User Experience (UX) represents the entirety of a user's engagement with a product, system, or service. UX encompassing not only its functional performance and ease of use, but also the crucial emotional and perceptual responses that occur before, during, and after interaction (HASSENZAHL, 2018). This view is formally recognized by the ISO 9241-210 standard, which defines UX as "a person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service" (STANDARD; ISO, 2019). ISO 9241-210 standard further emphasizes that UX is influenced by the system's characteristics, the user's state (i.e., including prior experiences, attitudes, skills, and personality), and the context of use, thereby highlighting its subjective and comprehensive nature.

To sustain the development of user experience in the products, UX design centers on creating intuitive interfaces that cater to user needs and preferences. It encompasses the entire product acquisition and integration process, from branding to functionality, all informed by a rigorous research process (i.e., UX Research) (NORMAN; NIELSEN, 2018; HASSENZAHL, 2018). The research about UX, i.e., UX research, is crucial for promoting an iterative design approach. It often employs diverse methods like interviews and usability testing to refine products based on user feedback and translate theoretical insights into practical solutions (MILLET, 2018; CHRISTENSEN et al., 2020). UX research can take advantage of using UX data, i.e., data gathered from users' interactions and interpretations of digital products, to create more usable designs that meet expectations and avoid superficial insights (CONVERTINO; FRISHBERG, 2017; OLDENBURG; HILLENBRAND, 2024; MARTINELLI; LOPES; ZAINA, 2022).

UX data is naturally generated throughout the UX research process from the application of quantitative and qualitative techniques. It offers significant information for companies to increase profitability and gain a competitive edge when it can be effectively analyzed by UX professionals (ZAINA; SHARP; BARROCA, 2021; FRITZ; BERGER, 2015). Although the concept of UX is established, the literature indicates a lack of a practical definition of UX data that can effectively guide software professionals[1] in leveraging its power to improve UX throughout the development cycle (KASHFI; FELDT; NILSSON, 2019; FABIJAN; OLSSON; BOSCH, 2016). This gap is especially critical for software teams without dedicated UX professionals, as they often lack the practical knowledge needed to leverage user interaction data effectively, underscoring their need for tools and methods to gain insights into user behavior and preferences for creating user-centric solutions (KOESTEN; SIMPERL, 2021).

Recent studies highlight a growing interest in understanding how UX data is handled in design processes and communicated among software team members (KASHFI; FELDT; NILSSON, 2019; KOESTEN; SIMPERL, 2021; ZAINA; SHARP; BARROCA, 2021). However, much of the practical knowledge acquired daily by software and UX practitioners from UX data exploration is not typically published in formal scientific literature, often due to the laborious nature of academic publishing, which may not suit the reporting of smaller, empirical insights (GLASS; DEMARCO, 2006; RAINER; WILLIAMS, 2019). However, much of the practical knowledge acquired daily by software and

[1] In this work, software professionals are the practitioners that played roles directly related to the design and development of software, e.g., UX designers, UX researchers, developers.

UX practitioners from UX data exploration is not typically published in formal scientific literature because it is derived from empirical insights. Considering this context of daily practices, the grey literature arises as an important repository that makes practitioners' knowledge available. The analysis of the grey literature can help to point out the “state of practice” on a given topic (GAROUSI; FELDERER; MäNTYLä, 2016). A grey literature publication aggregates information that is not necessarily peer-reviewed, and is available from different sources such as: personal pages, social networks, UX blogs, forums and Q&A sites (ADAMS; SMART; HUFF, 2017).

In this article, we present a comprehensive Grey Literature Review (GLR) exploring the integration and utilization of UX data throughout the software development lifecycle. Driven by the practical relevance and accessibility of grey literature (e.g., UX websites and blogs) and its alignment with current field challenges (GAROUSI; FELDERER; MäNTYLä, 2019), our study systematically selected 144 articles. From these articles, we gathered visualization examples and data definitions grounded in the practical knowledge of software professionals.

The results provide a detailed account of our findings, structured around four key research questions (RQ): we first identify and discuss the practices and tools employed to integrate UX data into UX activities (RQ1a), clarifying their connection to overarching themes and the role of supporting technologies. Following this, we delve into the motivations and benefits of incorporating UX data into software development (RQ1b), presenting these findings through a categorized structure that also elucidates the UX data flow. We then characterize the nature of UX data and its leverage in visualization techniques (RQ2a), organizing this discussion by the sequential stages involved in effectively utilizing UX data. Finally, we explore the types of data visualizations commonly used to represent and analyze UX data (RQ2b), detailing their evolution. Considering the findings, we proposed a taxonomy to support the software professionals in understanding how the UX data could be used in the construction of visualizations. As theoretical contributions, the paper provides a structured knowledge for advancing UX research and promotes a deeper understanding of the role of UX data in achieving user-centric objectives. From a practical perspective, we offer a taxonomy for exploring UX data, structured around a closed-loop cycle tailored to software professionals. This taxonomy aims to empower practitioners to harness available data and apply appropriate visualizations to derive actionable insights for an enhanced overall user experience.

This article is organized as follows: related work and background are discussed in Section 8.2; Section 8.3 presents the method followed to conduct the GLR; Section 8.4 discusses the results and answers the research questions of the GLR; Section 8.5 presents the design, apparatus and results of the taxonomy; the discussion is presented in Section 8.6; and Section 8.7 summarizes the conclusions.

8.2 Background

This section covers our work fundamentals and related work. In Subsection 8.2.1, we present the fundamentals about UX, and UX data and its linking with visualizations. In addition, we describe (LAM et al., 2012)'s framework which is adopted in our grey literature approach to support the analysis of visualizations. In the next section, we discuss the related work about the classification, structuring, and practical application of UX data (see Subsection 8.2.2).

8.2.1 Fundamentals

User Experience is a broad concept that refers to a user's entire interaction with a product or service (STANDARD; ISO, 2019). To provide a better UX, software and UX practitioners concentrate their efforts on the process of UX design (NORMAN; NIELSEN, 2018). Amongst different factors, UX design is directly impacted by the results obtained through the user research process, i.e., UX research. UX research utilizes diverse methods, e.g., interviews, surveys, and usability testing, to investigate users (MILLET, 2018). UX research outcomes sustain the design decisions, providing data about user needs, behaviors, and interactions with interactive systems (e.g., click sequences on a system interface), i.e., UX data (NARANG; TRIVEDI; DUBEY, 2017; CHRISTENSEN et al., 2020).

UX data can be categorized as quantitative, encompassing measurable metrics like success rates, number of pages visited, and return rates (ALBERT; TULLIS, 2013), or qualitative, which includes user-generated information such as opinions, preferences, and experiences with the system (RHIE; YUN, 2017). Although UX data can aid UX and software development practitioners to enhance user experiences and optimize product performance (LUTHER; TIBERIUS; BREM, 2020; TONG et al., 2022), their analysis is not trivial. Visualizations play a pivotal role in facilitating this data interpretation (CARD; MACKINLAY; SHNEIDERMAN, 1999). By constructing visualizations, UX and software professionals can overcome the abstraction barrier inherent in raw UX data, simplifying the understanding of complex user interactions, feedback, and large-scale behavioral datasets to extract meaningful insights (WARE, 2012; BUONO et al., 2020a).

Taking into account the importance of visualization for UX data interpretation, the understanding of which specific visualization techniques have been employed in practice by professionals becomes essential. However, to appropriately examine the diversity of visualization formats found in grey literature sources (i.e., sites, blogs) as well as scenarios of their applications, a systematic approach is requested. For this purpose, we adopted Lam et al. (2012) framework to carry out a systematic exploration of the information visualization techniques found in the grey literature and answer our research questions related to visualization more precisely. Lam et al. (2012) emphasize evidence-based approaches and outlines seven distinct scenarios where visualizations add value, such as enhancing data comprehension, supporting decision-making, and facilitating collaboration. Additionally, the framework includes a set of guiding questions designed to evaluate visualization techniques and their application contexts.

In our work, we selected and adapted a set of framework questions focusing on those that are in line with our research goal, i.e., the identification and categorization of visualizations used for UX data (see RQs in Section 8.1). To structure our data extraction from grey literature, we utilized defined scenarios in (LAM et al., 2012)'s framework which are: *Understanding environments and Work Practices (UWP)* - helps in understanding how visualizations aid in comprehending work practices related to information processing; *Visual Data Analysis and Reasoning (VDAR)* - allows to assess how visualizations support data analysis and reasoning; *Communication Through Visualization (CTV)* - supports evaluation of the communicative efficacy of a visualization; and *Collaborative Data Analysis (CDA)* - aid to examine a visualization's support for collaborative decision-making. The application of these scenarios and the adapted questions (also detailed in Subsection 8.3.6) provided a systematic lens to identify how different visualizations are used to explore and display UX data in practice.

8.2.2 Related work

This section reviews prior research pertinent to the classification, structuring, and practical application of UX data. To understand the landscape into which our work fits, we examine efforts to categorize qualitative and textual UX information, the development and limitations of existing UX data taxonomies, the ongoing drive to integrate diverse data types (including quantitative and emotional aspects), and the persistent need for frameworks that bridge theoretical constructs with practical, actionable insights. Reviewing these areas is crucial for identifying the specific gaps and opportunities that our research aims to address through the development of a taxonomy focused on UX data and its visualization.

(MENEWEGER et al., 2014) emphasizes the importance of textual data in UX research. However, they argue that this data requires a systematic effort for data classification to enhance the understanding of it in the user interaction context. Drawing on sociological approaches, they propose a theoretical foundation pivotal for categorizing qualitative data from user experience descriptions. This demand for user data in a more structured format is echoed by (LEMON et al., 2020). In a literature systematic review, the authors highlight that variabilities in UX data interpretation can be found and underscore that a standard proposal could improve the consistency of UX deliverables throughout diverse contexts. These two works discussions raise the necessity for a well-defined structure to provide ways to deal with diverse UX data. From this perspective, it is essential to characterize UX data and understand how its characteristics are leveraged in practice. Moreover, a taxonomy can support the management of the diverse types of UX data.

Several efforts have aimed to structure UX data into insights. The Engineering Usability Research Empirical Knowledge and Artifacts Taxonomy (EUREKATAX) (GERMANAKOS; FICHTE, 2020), for instance, organizes qualitative usability feedback into a hierarchical framework. The framework comprises four main categories (i.e., discover, learn, act, and monitor) and eight subcategories with 52 specific items. While its evaluation (with 80 UX experts and business professionals) demonstrated utility for usability test data analysis, its specific focus also highlights an ongoing need for taxonomies applicable to a broader spectrum of UX data sources and types. This study aims to investigate the visualization techniques currently employed to represent and analyze UX data, taking into account a broader range of data inputs.

The call for more comprehensive and nuanced categorization of UX data has been discussed in the context of evaluation. Kalantari e Lethbridge (2022), in their literature review on UX evaluation in software modeling tools, identified methodological gaps and advocated for improved categorization of UX evaluation methods. Their findings suggest that having a clear description of the structure of UX data and a deeper understanding of its nature and relationships could enhance the scalability and applicability of UX evaluations. Similarly, Lin et al. (2023) propose UX metrics grounded in psychological needs to enable comparative studies across varied product types. The emphasis in these studies on the need for comprehensive categorization directly influences the effective use of UX data throughout the development lifecycle, raising questions about the practices and tools for integrating UX data, as well as the motivations and benefits of incorporating such practices into the software development process.

Furthermore, the integration of data about emotions into UX frameworks has gained traction. Condori-Fernández (2017) argues for the inclusion of emotional data as a fundamental element in any comprehensive effort to describe the structure and meaning of general UX data. This perspective is supported by Ebel, Lingenfelder e Vogelsang (2021a), who propose a multi-level user behavior visualization framework categorizing interactions by context and task complexity. These works highlight the diverse and multifaceted nature of UX data, including emotional and contextual factors, which directly reinforces the need to understand its characteristics and how they are explored to gain deeper insights into user experience.

In summary, the literature reviews reveal a consistent drive towards more structured and comprehensive approaches to understanding and categorizing UX data. Key themes emerging include the value of systematically classifying textual data (MENEWEGER et al., 2014), the call for standardization in UX artifacts and deliverables (LEMON et al., 2020), the development of practical focused taxonomies (GERMANAKOS; FICHTE, 2020), the imperative to integrate qualitative and quantitative data (KALANTARI; LETHBRIDGE, 2022; LIN et al., 2023), and the importance of identifying UX data based on contextual, task-specific, and emotional factors (CONDORI-FERNáNDEZ,

2017; EBEL; LINGENFELDER; VOGELSANG, 2021a). These themes emerged from the literature directly informed the scope and focus of our research questions, particularly our investigation into the practices and tools (see Subsection 8.4.1), motivations and benefits (see Subsection 8.4.2), characteristics (see Subsection 8.4.3), and visualization techiniques (see Subsection 8.4.4) of UX data within the software development lifecycle.

Considering the related work discussed above, we recognize the need to examine the use of UX data in practice. While these contributions represent significant progress, the literature still reveals persistent gaps that our research questions aim to address, with the goal of generating practical insights that inform the development of our novel UX data taxonomy. Although existing frameworks are valuable, they are often limited to specific data types (e.g., EUREKATAX for usability test data), may not fully bridge the gap between theoretical rigor and practical application, or lack mechanisms for effectively integrating the diverse, often unstructured, qualitative and quantitative data (e.g., including emotional and contextual dimensions) that characterize real-world user experiences.

8.3 Research Method

We established a systematic process, based on guidelines proposed by (GAROUSI; FELDERER; MäNTYLä, 2019), to conduct a Grey Literature Review (GLR). Our research was conducted to investigate the integration of UX data into the software development cycle to enhance user experience. Specifically, we aimed to identify the attributes and characteristics of UX data and analyze how these characteristics are utilized in data visualization techniques to support UX activities throughout the development process. The GLR was conducted by two researchers (i.e., the two authors). The protocol was conducted by the first researcher (R1), with the collaboration of the second researcher (R2) in steps that needed validation.

The first researcher (R1) is a Ph.D. candidate in Computer Science at UFSCar, Brazil, and an active member of the UXLeris research team. He has over 8 years of experience as a software developer, and his doctoral research focuses primarily on UX data and visualizations aimed at supporting UX practices in the software industry. The second researcher (R2) is an Associate Professor in the Computing Department at UFSCar, Brazil, and a senior UX design researcher with over 15 years of experience. She leads the UXLeris team, coordinating projects that develop solutions to help software professionals integrate UX design activities into agile practices.

8.3.1 Defining the need for a GLR

Garousi, Felderer e Mäntylä (2019) developed a checklist consisting of seven yes-or-no questions (QGL) to assess the necessity of conducting a Grey Literature Review. The inclusion of grey literature in the research is considered justifiable if at least one QGL is answered positively. We highlighted three of the five QGL answered positively for this research as the motivation that guided our decision to do the GLR (see QGL3, QGL6, and QGL7 in Table 31). See the QGL developed by Garousi, Felderer e Mäntylä (2019) and the answers for this research in Table 31.

ID	Question	Answer
QGL1	Is the subject "complex" and not solvable by considering only the formal literature?	No
QGL2	Is there a lack of volume or quality of evidence, or a lack of consensus of outcome measurement in the formal literature?	Yes
QGL3	Is the contextual information important to the subject under study?	Yes
QGL4	Is it the goal to validate or corroborate scientific outcomes with practical experiences?	Yes
QGL5	Is it the goal to challenge assumptions or falsify results from practice using academic research or vice versa?	No
QGL6	Would a synthesis of insights and evidence from the industrial and academic community be useful to one or even both communities?	Yes
QGL7	Is there a large volume of practitioner sources indicating high practitioner interest in a topic?	Yes

Table 31 – Questions used to determine whether the inclusion of grey literature is appropriate for the research context (GAROUSI; FELDERER; MäNTYLä, 2019).

These are our rationale for answering "yes" to the questions: (QGL3) based on our related work search (see Subsection 8.2.2) that revealed a need for more attention to UX data definitions and achievable goals that rise from practical environments (e.g, software development industries); our data source is mainly based on grey literature found in UX sites and blogs that represent industrial setups. (QGL6) practitioners can use our findings to apply integration of UX data and data visualization in the software development lifecycle, and our lessons learned would be valuable for the academic communities to advance studies about integration of UX data and visualizations. (QGL7) due to the great number of blogs and websites (i.e., practitioner sources) that have posts about the attributes of UX data and how these characteristics are leveraged in data visualization techniques.

Based on the justifications for conducting the GLR, our decision was driven by the recognition of critical gaps in the formal literature and the relevance of insights from industry practices. The contextual importance of UX data (QGL3) was evident in our preliminary research, which highlighted a lack of structured definitions and practical applications of UX data within software development environments. Additionally, the synthesis of knowledge from both industrial and academic communities (QGL6) was essential to bridge the gap between theoretical research and real-world UX practices, ensuring that our findings could benefit both practitioners and researchers. Lastly, the high volume of practitioner sources (QGL7) discussing UX data attributes and visualization techniques demonstrated a strong industry interest in the topic, reinforcing the necessity of incorporating grey literature into our study. These motivations provided a solid foundation for conducting the GLR, ensuring that our research captured both theoretical and practical perspectives to advance the understanding and integration of UX data in software development.

8.3.2 Research questions

Considering the opportunities discussed by the related works (see Subsection 8.2.2), we defined two research questions (RQ) to focus on bridging the gap between theoretical and practical aspects of UX data and the use of visualizations to better address the complexity of UX analysis. Each RQ gave rise to a sub-question (i.e., four RQs in total).

- ❏ *RQ1a - What are the practices and tools employed to integrate UX data into the UX activities throughout the software development lifecycle?*
- ❏ *RQ1b - What are the motivations and benefits of incorporating UX data into software development practices?*

For RQ1a and RQ1b our rationale is to understand the current UX data integration practices to identify common patterns or inconsistencies in the process and gain insights about gaps and standardization opportunities (i.e., RQ1a), and determine the potential impact of UX data on software development outcomes and the value proposition of UX data integration (i.e., RQ1b).

- ❏ *RQ2a - What are the characteristics of UX data and how are these characteristics leveraged in data visualization techniques to support the discovered practices?*
- ❏ *RQ2b - What types of data visualizations are commonly used to represent and analyze UX data?*

The rationale here is to understanding what attributes are essential to classify and select appropriate data as UX data to define a suitable visualization technique that communicates insights in a clear, concise, and meaningful way (i.e., RQ2a); and focused only on data visualizations we aim to identify the commonly used data visualization techniques for UX data in the current state of practice and collecting recommendations for improving the use of data visualization in UX research (i.e., RQ2b).

8.3.3 Grey article retrieval

The definition of search string for grey literature is an interactive process where the researcher performs an initial exploration for seminal works and then defines the relevant terms for the search (GAROUSI; FELDERER; MäNTYLä, 2019). To develop the search string, we followed a well-established method used in systematic mapping studies, the PICO acronym (PETERSEN; VAKKALANKA; KUZNIARZ, 2015). It stands for Population, Intervention, Comparison, and Outcomes; this research focuses only on Population and Intervention based on Melegati, Guerra e Wang (2021) approach. **Population**: we focused on searching considering “UX data” as our major term, even recognizing that other terms with similar meaning could be considered. We bypassed broadening the variation of a term with other synonyms (i.e., user data), which could lead to an unmanageably large number of results, many of which would likely be false positives. Following the same procedures adopted by Melegati, Guerra e Wang (2021), we decided to split the "UX data" term into different strings to avoid issues with logical operators and to prevent the search algorithm from assuming that a string with more terms is more important than one with fewer. **Intervention**: we select terms that emerged from the related work describing practices made while using UX data through software development.

Google was chosen as the search engine due to its popularity in studies involving grey literature and by the characteristic of being a general internet search engine (GAROUSI; FELDERER; MäNTYLä, 2019). All search strings (see Table 32) were run in *Google Chrome* in incognito mode to avoid the influence of the saved history and cookies in the search results. The defined search strings, the number of results each returned, and the number of selected grey articles are presented in the Table 32. The researcher R1 made the strings with the revision of the researcher R2.

ID	String	Results	Selected
1	"quantitative UX"	99	14
2	"UX" + "data"	195	35
3	"UX" + "Data Analysis"	130	11
4	"UX data" + "analysis"	195	18
5	"UX data" + "charts"	67	4
6	"UX data" + "dev"	119	4
7	"UX" + "data-driven"	229	27
8	"UX data" + "evaluation"	124	3
9	"UX" + "data extraction"	164	4
10	"UX data" + "information visualization"	28	0
11	"UX data" + "InfoVis"	32	0
12	"UX data" + "patterns"	119	5
13	"UX data" + "platform"	130	6
14	"UX data" + "product development"	135	4
15	"UX" + "data scientist"	183	8
16	"UX data" + "software development"	138	0
17	"UX" + "information visualization"	93	0
18	"UX" + "InfoVis"	82	1
-	Total	2262	144

Table 32 – Search strings and the number of results returned and selected for the extraction step.

The search string results were retrieved using the SEOQuake plugin[2]; this plugin gets all the data displayed in the search result (i.e., website title, description, and URL) and outputs it in a comma-separated values (CSV) file. The CSV has been loaded in a spreadsheet, and each result received an ID respecting the template "*n1-n2*", where *n1* represents the search string ID and *n2* the ranking of the result in relation to the Google search page. The search process generated a set of 2262 grey articles, which are available on Google Drive.

[2] <https://www.seoquake.com/>

8.3.4 Selection criteria

We defined three exclusion criteria, and a single inclusion criterion for the selection process. Researcher R1 conducted the selection process, while researcher R2 served as an evaluator, performing a random sampling assessment of 30% of the grey articles. Before starting the selection process, we conducted the cleaning of data by analyzing the content availability of the URL. We conducted the checking because it was noticed that some results that had their content removed from the internet continued to be indexed by search engines.

We developed an algorithm using *Google Apps Script* to check the HTTP status of the grey article URL (i.e., collected by SEOQuake). By subjecting the URLs of the grey articles to our script we excluded 183 results that returned statuses in the 400 families (i.e., error codes). In addition to the URLs removed due to error status, 20 URLs were also removed for redirecting to sites with no relation to their titles and descriptions (i.e., clickbait). This cleaning process reduced the set to 2059 grey articles.

We started the selection process by applying exclusion criteria to reduce the number of articles that would be read in full. An article would be excluded when *(CE1) the content is duplicate*, *(CE2) it is just advertising content*, or *(CE3) it is not in textual format (i.e., videos or podcasts)*. We analyzed the URL and textual content to determine that a grey article was a duplicate (i.e., CE1); 406 grey articles were removed as duplicates. We excluded 241 grey articles that only presented advertising, such as author biographies, sales of online courses, and promotion of YouTube channels (i.e., CE2). We found 47 URLs that point to videos or podcasts that did not have an official transcript were excluded (i.e., CE3). This exclusion process reduced the set to 1365 grey articles.

For inclusion we defined that *(CI1) The grey article must describe something related to using UX data*. The CI1 criteria was applied in the 1365 grey articles that were not excluded by the exclusion criteria and selected 144 as approved grey articles. All approved grey articles were accessed and their contents were downloaded as PDFs; all PDFs were stored in a folder on Google Drive and named with the same ID as the spreadsheet. Due to limited space, citations to grey articles will be made using the "ACn" format (i.e., Article Citation), and the bibliographic information of grey articles can be consulted in Section 12.5. The Figure 34 shows the number of grey articles removed at each step of the selection process.

8.3.5 Quality assessment

Before proceeding with the analysis of the 144 approved, we carried out a quality assessment of our sample. This step aided us in understanding the nature, origin, and scholarly attributes of the selected literature, thereby informing the subsequent analysis. The complete and detailed quality assessment outcomes for all 144 articles are compiled

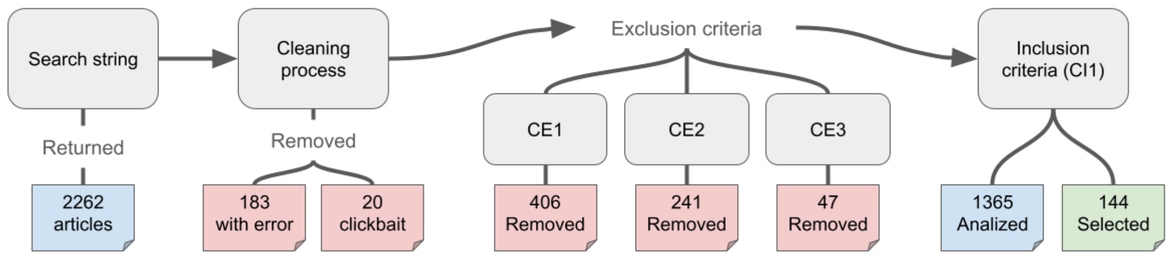


Figure 34 – Quantitative information about the number of grey articles removed in each step of GLR.

and available in an electronic spreadsheet.

A quality assessment form was designed (see Section 12.5) to capture the following information for each article: (a) *authorship profile* - if the work was by an individual author, multiple authors, or an institutional publication; (b) *author's publication record* - check whether the author(s) had previously publications of other works; (c) *context of application* - categorize work as referring to a real-world situation (e.g., an industrial application or case study) or an academic research context; (d) *scholarly attributes* - check the bibliography information and distinct conclusions were presented; (e) *potential vested interest* - evidence of vested interest, such as overt advertisement or the promotion of paid tests or services; (f) *primary contribution(s)* - categorize the main contribution(s) based on a predefined list of options. The list options included: "how to explore/analyze UX Data", "how to visualize UX Data", "how to collect/store UX Data", "how to communicate UX Data findings", "when we should explore/handle UX Data", "what we can extract from UX Data", "what we consider UX Data (definitions)", "what UX Data can tell us", "who manages UX Data", and "information about UX Data tools/approaches".

This quality assessment provided a comprehensive overview of the grey articles selected and aided us in the extraction step and the results presentation.

8.3.6 Data extraction and analysis

To analyze the data, we follow a protocol based on (MELEGATI; GUERRA; WANG, 2021)'s work that uses the thematic synthesis (CRUZES et al., 2015; CRUZES; DYBA, 2011) for identifying, analyzing, and reporting themes. We chose this approach because our research goal is aligned with the goal of determining the current practices performed in the industry and led with a large and diverse body of research (i.e., grey literature).

For the first step of thematic synthesis, the researcher immersed themselves in the data to understand the details of the publications; this step was carried out during the document selection process. The second step (i.e., the extraction step in this paper) involves coding data when interesting concepts are identified throughout the entire dataset.

To establish a systematic process for this extraction step, researcher R1 selected a set of six questions proposed by Lam et al. (2012) (see explanation in Section 8.2). These six questions, called herein "extraction questions" were correlated to this study's research questions (see Subsection 8.3.2). These extraction questions served as themes, guiding researcher R1 in selecting excerpts of text from the grey literature articles to address the research questions.

The extraction questions, their relationship with the research questions, and the Lam et al. (2012) scenarios related to are presented in Table 33. To answer each extraction question, we considered the Lam et al. (2012) scenarios, which are previously explained in the background section (Section 8.2).

ID	(1) Extraction questions	(2) Scenario	RQ	(3) N. Ext.
Q1	In which daily activities should the visualization tool be integrated?	UWP	[RQ1a]	63
Q2	What is the context of use of visualizations?	UWP	[RQ1a]	111
Q3	What types of analyses should the visualization tool support?	UWP	[RQ1b]	128
Q4	How does it support processes aimed at seeking information, searching, filtering, and reading and extracting information?	VDAR	[RQ1a]	12
Q5	What kinds of visualizations are currently in use?	UWP	[RQ2b]	53
Q6	What data are currently used and what tasks are performed on it?	UWP	[RQ2a]	112
			total of extractions	479

Table 33 – (1) Extraction questions adapted from (LAM et al., 2012). (2) Scenario: *UWP - Understanding environments and Work Practices*; *VDAR - Visual Data Analysis and Reasoning*; *CTV - Communication Through Visualization*; *CDA - Collaborative Data Analysis*. (3) N. Ext.: number of extractions from the grey literature.

We developed an electronic form to perform the extraction step. The extraction form consisted of the grey article ID, the six extraction questions (see the questions on Table 33). Each approved grey article was read in full, and its contributions to each question were extracted separately, i.e., an entry in the form for each contribution. Due to the context of the research questions, i.e., UX data and visualization approaches, not all the grey articles presented contributions to all questions in the extraction form.

Figure 35 presents an example of extraction performed in the grey article AC130 where the excerpts found answered two questions linked to **UWP** scenario: *Q3 - What types of analyses should the visualization tool support?* and *Q6 - What data are currently used and what tasks are performed on it?*. Examining the Q3 extraction, we can discuss the benefits of analyzing exit and bounce rates. Based on the Q6 extraction, we can define the attributes of the mentioned UX data.

We performed 479 extractions of relevant excerpts during the extraction process to discuss the research questions. The number of excerpts extracted per extraction question (i.e., see the questions on Table 33) was: Q1 (63), Q2 (111), Q3 (128), Q4 (12), Q5 (53), and Q6 (112). The extracted excerpts are available in a electronic spreadsheet.

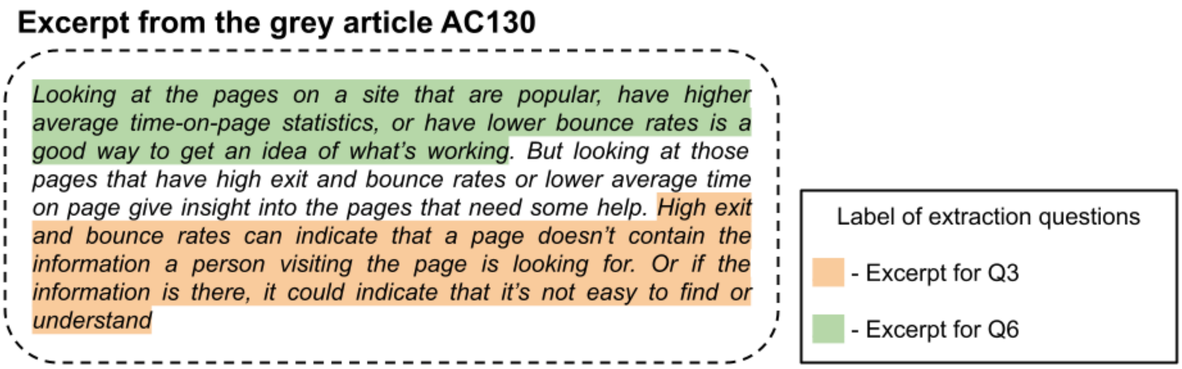


Figure 35 – Example of application of extraction questions in grey article AC130.

8.4 Results

Considering the context of the work, a significant 81% (i.e., 117 articles) indicated that the knowledge and practices described were obtained in industrial work scenarios, while 18% (i.e., 21 articles) originated from academic research scenarios. Only one grey article's context could not be determined [AC1]. In terms of authorship, 75% (i.e., 108 articles) were single-authored, approximately 8% (i.e., 11 articles) had two or more authors, and 17% (25 articles) were published on behalf of an institution. An examination of the authors' publication histories showed that 74% (i.e., 107 of the primary or corresponding authors) have at least one other publication, and notably, a majority of these authors (i.e., 42 authors) have other publications specifically related to UX data.

The following sections will now delve into answering the research questions presented in Subsection 8.3.2. Bibliographic information for all approved grey articles, including their "ACn" citation format, can be consulted in Section 12.5.

8.4.1 RQ1a - What are the practices and tools employed to integrate UX data into the UX activities throughout the software development lifecycle?

Our grey literature review identified a diverse set of practices and tools utilized by software professionals to integrate UX data throughout the software development lifecycle, ultimately enhancing the design of interacting systems and supporting data-driven decisions for user-centered products. These findings are summarized by four key themes that emerged from our analysis (see Table 34). Each theme represents a group of practices based on their common overarching goals or characteristics, and lists the corresponding grey literature references that support each theme. These themes were systematically identified through a content analysis focusing on recurring objectives and actions of the practices described in the grey articles. The next paragraphs discuss them.

The *understanding user needs* theme (see #1 in Table 34) encompasses practices fundamentally aimed at gaining deeper insights about user behavior and preferences throughout the software development process [AC2, AC65, AC131, AC142]. This involves activities such as conducting user research (e.g., surveys, interviews, observations) to gather direct feedback, and analyzing existing user data to identify behaviors [AC57, AC89, AC48]. These practices serve as the foundational step, ensuring that subsequent development efforts are truly aligned with the target audience's requirements and expectations, fostering user-centered design [AC80].

Central to integrating UX data is covered by the theme *data-driven design* (see #2 in Table 34), which emphasizes utilizing empirical evidence to inform and validate design decisions [AC2, AC23, AC41, AC59, AC92, AC98, AC105, AC113]. This theme encompasses key practices such as defining clear UX goals, formulating hypotheses, conducting tests and experiments (e.g., A/B testing, user behavior tracking), and analyzing outcomes to confirm hypotheses [AC5, AC23, AC32, AC41, AC59, AC81, AC82, AC105, AC130, AC134, AC140]. By doing so, software practitioners gain deeper insights into user behavior, enabling more informed decision-making and the creation of experiences precisely tailored to user needs throughout the entire development process [AC2, AC54, AC71, AC118, AC142].

The *improve user experience* theme (see #3 in Table 34) highlights direct actions taken to enhance the overall quality of user interaction with software systems [AC26, AC59, AC68, AC84, AC95]. This includes leveraging UX data to identify patterns and trends that inform design decisions for performance and usability improvements, as well as diagnosing existing issues to propose targeted enhancements [AC30, AC42, AC84, AC105, AC130, AC134]. These practices underscore the continuous and iterative approach inherent in data-driven methodologies for refining work strategies and delivering better user outcomes [AC71].

Finally, the *evidence-based hypothesis validation* theme (see #4 in Table 34) represents a crucial practice for ensuring that design decisions are sustained by empirical evidence [AC5, AC48, AC92, AC98, AC130, AC134, AC138]. This theme encompasses methods focused on systematically testing assumptions about user behavior and product performance. It aligns with activities such as monitoring interface performance through testing hypotheses (e.g., A/B testing, user behavior tracking) [AC32, AC82, AC105] and enhancing communication regarding UX objectives and data-derived insights among the development team (e.g., through personas, data-driven UX discussions) [AC6, AC34].

Beyond the identified practices, a common thread found in the analysis of the grey literature was the crucial role of specialized tools in the facilitation of UX activities [AC17, AC27, AC49, AC79, AC83, AC101, AC112, AC143]. These tools often fall into categories based on their primary functions which are: for visualizing UX data (e.g., [AC63, AC138]), for automated UX data collection (e.g., analytics tools [AC55, AC33, AC40]), and for user

ID	Themes	Grey literature reference
1	Understanding user needs	[AC2, AC51, AC131, AC142, AC65]
2	Data-driven design	[AC2, AC92, AC98, AC105, AC113, AC41, AC59, AC50, AC5, AC69, AC23, AC94]
3	Improve user experience	[AC68, AC26, AC59, AC84, AC95]
4	Evidence-based hypothesis validation	[AC130, AC92, AC98, AC42, AC134, AC48, AC5, AC138]

Table 34 – Emerging practices that use UX data described in the grey articles.

data collection (e.g., user testing and research platforms [AC122, AC116]). The tools identified in the grey literature that support the use of UX data in work activities are listed in Table 35.

The prevalence of tools like Google Analytics, Hotjar, and Crazy Egg, as evidenced by their frequent citation in Table 35, indicates a strong emphasis in grey literature on automating behavioral data collection and visualization (e.g., heat maps, scroll tracking). These tools primarily contribute to monitoring interface performance with a focus on testing hypotheses (e.g., A/B testing, user behavior tracking) [AC32, AC82, AC105]. They assist software professionals in identifying critical areas, such as user drop-off points (i.e., bounce rate), pages with the highest engagement, and actions that lead to increased conversion rates [AC71, AC118, AC58]. In addition to managing performance aspects, these tools also serve to enhance communication regarding UX objectives (e.g., through personas [AC6]) and insights derived from data among the development team (e.g., fostering data-driven UX discussions [AC34]).

Tool	Grey article
Google Analytics	[AC5, AC30, AC41, AC53, AC55, AC71, AC84, AC117]
Hotjar	[AC41, AC53, AC65, AC71, AC87]
Crazy Egg	[AC40, AC71, AC117, AC125]
Mixpanel	[AC55, AC58, AC71, AC84]
Inspectlet	[AC65, AC84]
Mouseflow	[AC65]
FullStory	[AC71, AC84]
Heap	[AC71, AC84]
UXCam	[AC58, AC65]
Flink	[AC71]
Lookback	[AC65]
Amplitude	[AC84]
Contentsquare	[AC65]
AppSee	[AC65, AC71, AC84]

Tool	Grey article
W3Counter	[AC55]
UserFeel	[AC116]
TryMyUI	[AC116]
Qualaroo	[AC116]
Survey Monkey	[AC116]
Data Studio	[AC116]
Dovetail	[AC24]
Aurelius	[AC124]
Ahrefs	[AC124]
Lucky Orange	[AC125]
Optimizely	[AC87]
VWO	[AC87]
lookback	[AC87]
Userlytics	[AC87]
Matomo	[AC55]

Table 35 – UX Data Tools Identified from Grey Literature.

The integration of UX data into the development process is inherently collaborative. It affects the involvement of diverse software professionals, varying significantly throughout the software development lifecycle and reflecting the distinct areas of professionals' expertise and contributions. Figure 36 shows the strength of the connections between the actions support from UX data (on the left side) and the roles of professionals that performed these actions (on the right side); where the thin connections denote less involvement and the thicker connections more involvement. From this analysis, different patterns of engagement are discussed in the following.

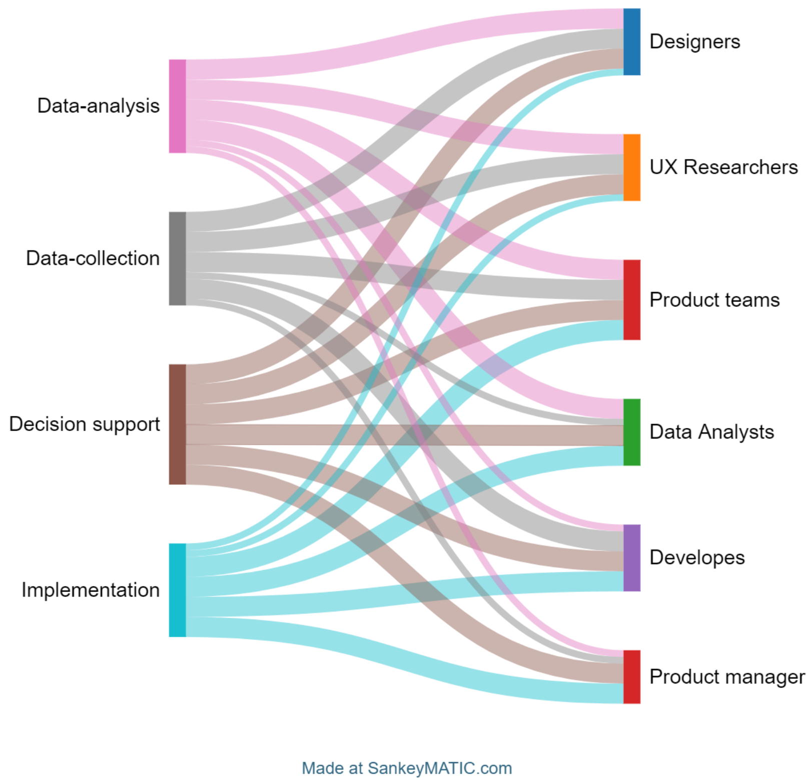


Figure 36 – Connections between the roles and the actions performed with the support of UX data.

Designers are heavily involved in both data collection, gathering user feedback and qualitative insights, and data analysis to interpret user interactions and consequently identify points of improvement. While they moderately contribute to decision-making by informing design choices, their involvement in implementation is relatively limited, focusing primarily on interface design [AC1, AC6, AC29, AC59, AC68, AC109, AC123, AC130, AC131, AC134]. *UX Researchers* primarily lead data collection, conducting tests, interviews, and surveys. They are also deeply engaged in data analysis to interpret findings and provide moderate support in decision-making. However, their role in direct implementation of solutions is minimal [AC30, AC57, AC61, AC86, AC114, AC115].

Data Analysts predominantly focus on analyzing data, processing metrics, and delivering insights that heavily influence decision-making. Although they play a minor role in data collection and implementation, their work is crucial for informed decisions throughout the development process [AC27, AC75, AC91, AC92, AC121, AC138]. *Product Teams* contribute moderately throughout all phases, from gathering metrics to contextualizing data within business goals, to aiding decision-making, and refining products during implementation [AC3, AC35, AC63, AC111, AC124, AC137].

Finally, *Developers* are primarily involved in implementation, leading the development of UX-driven changes. Their involvement in data collection and analysis is modest, typically centered on technical performance metrics, yet they provide essential feasibility insights that support decision-making [AC4, AC48, AC76, AC81, AC87, AC89]. Product Managers contribute most significantly in decision-making, prioritizing goals and guiding team efforts using data-driven strategies. They also play a substantial role in implementation by ensuring decisions align with strategic objectives. However, their direct involvement in data collection and analysis is limited, as they rely on insights provided by other roles [AC2, AC22, AC63, AC91, AC98, AC103, AC117].

8.4.2 RQ1b - What are the motivations and benefits of incorporating UX data into software development practices?

Our grey literature review identified several primary motivations and significant benefits of incorporating UX data into software development practices, which drive a more user-centered and data-driven approach. These findings were systematically grouped into five categories, which are: understanding user behavior; support for decision-making; guidance on optimizing products and experiences; identification of problems and opportunities; and performance assessment. The following discussion elaborates on each of these categories.

A primary motivation for using UX data is to gain a deeper and more objective understanding of how users interact with software, i.e., the *understanding user behavior* [AC51, AC59, AC5, A54, AC53, AC65, AC133]. This involves analyzing user navigation patterns, content consumption, and actions performed in interfaces [AC125, AC37, AC126]. Qualitative analysis, often obtained through methods such as interviews and user testing [AC49, AC19, AC78], provides the “why” behind behaviors observed in quantitative data, offering insights into user motivations and expectations [AC65, AC69, AC64]. This comprehensive understanding fosters a more empathetic and user-centered approach, connecting software professionals directly with their target audience [AC23, AC81]. The continuous collection of UX data over time is also highlighted as crucial to fully understanding actual user experiences and facilitating ongoing product optimization [AC41, AC84].

In the *support for decision-making* category, UX data is seen as fundamental to data-driven development approaches [AC96], significantly enhancing informed decision-making throughout the entire development process [AC2, AC142]. This includes employing metrics to validate development progress [AC5], generating innovative ideas informed by user data [AC51], and ensuring alignment with the needs of software teams [AC53]. Integrating UX data with the expertise of software professionals enables the creation of more effective and successful design solutions [AC133, AC24, AC109] by providing concrete evidence, rather than relying solely on assumptions about users’ interactions [AC37, AC92]. UX data also facilitates the communication of user profiles and realities by creating personas based on demographic and behavioral attributes [AC34, AC98, AC65], thereby supporting evidence-based prioritization of goals and team efforts [AC85, AC96].

The category of *guidance on optimizing products and experiences* underscores how UX data guides the continuous refinement of products based on user feedback and insights. Practices such as utilizing UX data to compare the effectiveness of existing interfaces to select the one that maximizes user engagement represent a key aspect of this guidance [AC52, AC38, AC60]. The goal is to iteratively refine products and experiences, ensuring alignment with evolving user needs and improving overall user satisfaction and engagement [AC25, AC32, AC38, AC66, AC86, AC120, AC122, AC124, AC129]. This also covers usability improvements and feature prioritization.

In *identification of problems and opportunities*, UX data serves as a fundamental tool for identifying discrepancies between the envisioned and real user experiences [AC13], and for uncovering usability problems. Both behavioral (quantitative) and UX research (qualitative) data are utilized to analyze the challenges users face with interfaces [AC126, AC120, AC129, AC122]. This analysis helps to assess how specific changes influence user navigation and overall product performance, thereby leading to the identification of problems and new opportunities for improvement [AC4, AC42, AC92, AC100, AC102].

Finally, the *performance assessment* category manages the incorporation of UX data, which is deeply linked to evaluating overall product performance and contributing to strategic business objectives. This translates into products that are better aligned with market demands [AC100] and optimized user experiences, fostering professional motivation [AC133, AC24, AC109]. Data-driven development approaches leverage metrics to validate progress [AC5], contributing to the success of business objectives and organizational UX maturity [AC40, AC105, AC132].

Taking into account the motivations and benefits categorized above, a structured visualization of the continuous flow of UX data throughout the software development lifecycle can be elaborated as shown in Figure 37. The figure illustrates how UX data progresses through various stages, with each stage representing a key action or outcome driven by the motivations and benefits outlined in the categories described above. It begins with **data collection** from diverse user interactions, surveys, and usability tests (reflecting aspects of *understanding user behavior* and *support for decision-making*). This collected data is then analyzed (i.e., **data analysis**) to uncover patterns and insights (aligning with *understanding user behavior* and *identification of problems and opportunities*) that directly inform design decisions (i.e., *support for decision-making*). Key practices such as **hypothesis testing and validation** (contributing to *guidance on optimizing products and experiences* and *support for decision-making*) are integral to refining these insights, ensuring that product improvements are empirically based on real user behavior. The figure emphasizes the continuous feedback loop within this workflow, illustrating how data continuously drives **product optimization** by validating hypotheses and refining interfaces based on ongoing user feedback (a core benefit of *guidance on optimizing products and experiences* and *performance assessment*). Ultimately, Figure 37 underscores the importance of an iterative, data-driven approach to product development to enable teams to create more user-centered and effective digital experiences, demonstrating the interconnected nature of all identified benefits and motivations.

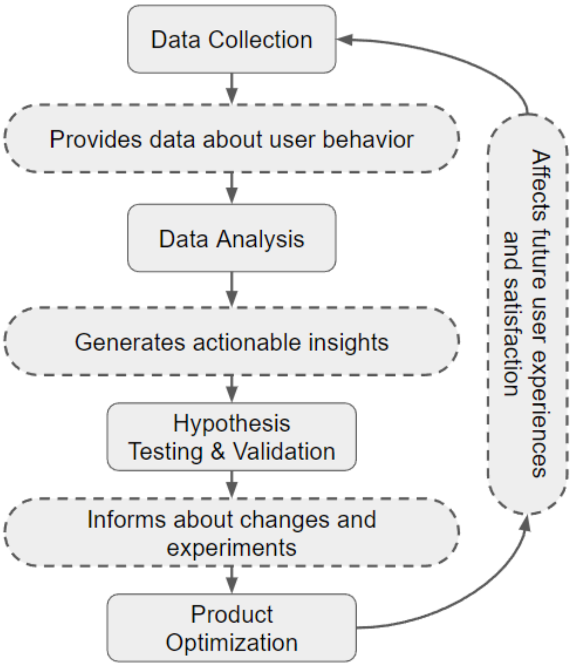


Figure 37 – UX Data Practices in the Development Workflow.

8.4.3 RQ2a - What are the characteristics of UX data and how are these characteristics leveraged in data visualization techniques to support the discovered practices?

The results reaffirmed that UX data can broadly be categorized into behavioral and attitudinal types. These data types inherently contain both qualitative and quantitative information [AC40, AC96, AC16]. Typically, behavioral data are predominantly presented in quantitative forms, often analyzed statistically to calculate metrics [AC32]. In contrast, attitudinal data usually exhibits a qualitative predominance, analyzed to understand users' subjective perceptions [AC25]. Quantitative data provides an objective view of actual user behavior, such as interactions and actions within a system [AC99], while qualitative data captures user qualities, feelings, and attitudes, expressed in text [AC79, AC63].

Beyond this behavioral and attitudinal classification, UX data can also be sorted according to the objectives of UX research and classified in four types: descriptive (i.e., insights into user interaction frequency), diagnostic (i.e., tracking changes over time for trends), prescriptive (i.e., assessing trends for optimal patterns), and predictive (i.e., simulating scenarios for recommendations) [AC55]. Leveraging these diverse characteristics of UX data to enhance user experience involves a sequence of interconnected stages as follows.

The grey literature emphasizes the critical importance of the *planning* stage to establish a tracking plan prior to collecting and visualizing UX data. A robust tracking plan specifies precisely which data should be monitored, how its characteristics align with available tools (e.g., data visualizations) [AC135], and how it contributes to achieving overall user experience goals [AC80, AC99]. Software practitioners should strategically determine, via the tracking plan, whether their existing data can readily address specific UX questions, if additional analysis is required, or if there is insufficient coverage due to missing data [AC13]. Critically, a well-defined tracking plan ensures the relevance of the collected data, thereby preventing the generation of unmanageably large data volumes that often lead to unhelpful visualizations [AC29, AC33, AC122, AC124].

Data *collection* methods vary depending on the type of UX data. Behavioral data, typically quantitative, can be gathered through analytics platforms, usability tests, and observations. Examples of data include conversion rates, session times, pages visited, task times, abandonment rates, and error rates [AC32, AC82, AC99]. Attitudinal data, primarily qualitative, is usually collected via methods such as feedback forms and user interviews [AC57, AC89], usability tests [AC48, AC50], persona analysis [AC79, AC63], and observations that aid software professionals in understanding user motivations and expectations [AC85].

The *integration* of diverse UX data into insights involves methodologies that connect different data natures. Quantitative data can often be directly inserted into visualization tools [AC120, AC122]. In contrast, qualitative data require a more in-depth processing and analysis procedure to derive meaningful insights [AC120, AC124]. Tools and methodologies such as the HEART framework [AC136], thematic analysis [AC78], and CAQDAS tools [AC19] specifically help integrate these varied data types. This integration is crucial because qualitative data provides deep insights into the “why”, while quantitative data reveals the “what” in terms of patterns and trends [AC85, AC99].

The *analysis* stage focuses on synthesizing collected and integrated data to uncover meaningful patterns and draw conclusions. Combining behavioral (i.e., primarily quantitative) and attitudinal (i.e., primarily qualitative) data offers a comprehensive approach, where diverse methods provide complementary insights and mitigate the limitations inherent in each individual approach [AC86]. For instance, integrating qualitative methods (e.g., interviews and usability tests) with quantitative methods (e.g., analytics, A/B testing, and surveys) enables a more comprehensive examination of the user experience. This enables practitioners to effectively hypothesize about “why” behind user behavior [AC62] while simultaneously observing “how” their attitudes manifest during interactions [AC13, AC77, AC138]. Analyzing the challenges users face with system interfaces through both behavioral and UX research data allows the assessment of how specific changes influence user navigation and overall product performance [AC120, AC126, AC129, AC122].

The ultimate *application* of UX data insights is often facilitated through effective data visualization, which enhances the user experience by creating ways to make complex data understandable. By combining behavioral and attitudinal data, various visualization approaches can be employed. These include developing personas and empathy maps to gain a comprehensive view of user perception [AC98], identifying bottlenecks and improvement opportunities through user journey mapping [AC98, AC142], and analyzing user behavior patterns by merging performance metrics with collected feedback [AC92, AC131]. Although personas and empathy maps are derived primarily from qualitative attitudinal data, understanding users' actions within the system can also involve quantitative behavioral data collected from real user interactions [AC108, AC109]. UX data also facilitates the communication of user profiles and realities through personas based on demographic and behavioral attributes [AC34, AC98, AC65].

Figure 38 illustrates the sequential stages involved in leveraging UX data to improve user experience. Each step of the process ensure the effective transformation of raw data into actionable insights. The pipeline emphasizes the interplay between behavioral and attitudinal data while addressing both qualitative and quantitative dimensions.

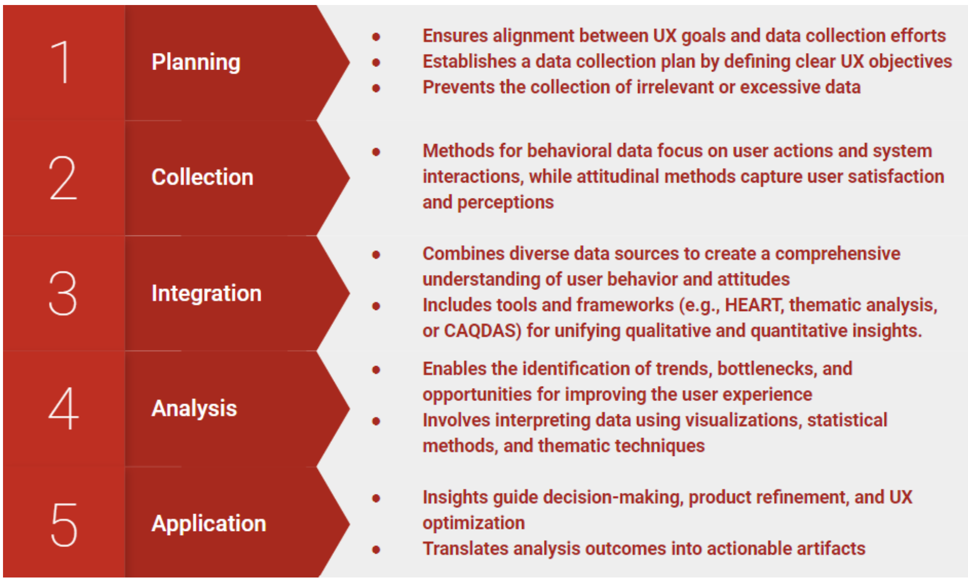


Figure 38 – UX Data Collection and Analysis Pipeline.

8.4.4 RQ2b - What types of data visualizations are commonly used to represent and analyze UX data?

Our grey literature review identified a diverse range of visualization types commonly used to represent and support UX data analysis. These visualizations are often available in tools designed to display data related to performance, click analysis, and user journeys [AC87]. Common UX data visualization types can be broadly categorized into their primary purposes.

Standard charts include fundamental formats like bubble charts [AC42, AC123], scatter plots [AC34], bar charts [AC123], and tree diagrams [AC141], widely used for analyzing performance metrics and click analysis. *User-centric diagrams* are more specialized formats directly focus on user experience, such as affinity diagrams [AC98], personas [AC88], empathy maps [AC16], user journey maps [AC42], and user flow maps [AC53], applied to simplify data interpretation and classification [AC71, AC131].

Dashboards for integrated views act as powerful tools displaying multiple visualizations simultaneously [AC116], often incorporating user flow charts [AC82, AC125], dendrograms [AC138], similarity matrices [AC138], and heat maps [AC87, AC120]. In particular, dashboards with data storytelling features are accessible approaches for stakeholders to extract insights and make informed decisions [AC21, AC22].

Heat maps, in particular, offer diverse applications, such as showing user clicks [AC126], scroll patterns [AC42], and mouse movements [AC40], enabling a deeper understanding of user behavior [AC65, AC71] and identifying areas of interest [AC57, AC120]. *Tabular visualizations* are a widely accepted format due to their easy readability and interpretation [AC1, AC2, AC3]. Tabular visualization is exemplified by UX scorecards, which track user experience metrics over time to monitor performance and maturity [AC132]. Personas also leverage tabular and textual visualization to convey user needs, motivations, and behaviors [AC5, AC70, AC91, AC103, AC143].

While these diverse visualization types are used, combining quantitative and qualitative data in visualizations provides a more comprehensive view of software performance [AC23, AC64, AC136] and fosters participation between project stakeholders [AC19, AC140]. However, some data types lack a standard structure and require custom visualizations, often seen with user flow [AC38], session recordings [AC63], usability tests [AC125, AC126], surveys [AC101], conversion funnels [AC41], and user feedback [AC122].

Our results revealed a progressive evolution in how UX data is utilized and visualized. This progression, from basic metric representations to highly integrated and predictive analytics, underpins the evolutionary stages depicted in Figure 39. This figure aggregates the various visualization types and their applications identified from the grey literature review, showing how the field has moved towards more comprehensive insights.

Initially, visualizations primarily focused on **basic performance metrics** such as page views, session duration, and conversion rates. These were commonly represented through bar charts and scatter plots (i.e., *standard charts* and *tabular visualizations*). Then it evolved to **integration of behavioral data**. As user interactions became more intricate, visualizations evolved to incorporate more varied and complex data. This stage saw the increased use of behavioral insights such as click analysis and heat maps (i.e., *dashboards* and *heat maps*), as well as user journey and flow charts (i.e., *user-centric diagrams*).

The *focus on data storytelling* showed the emergence of more interactive visualizations, allowing stakeholders to view multiple data sources simultaneously and promoting data storytelling approaches (i.e., *dashboards*) to support informed decision-making. Finally, in the current era, UX visualizations have further advanced by integrating **real-time analytics and predictive modeling**. These visualizations leverage all of the common UX data visualization types identified in the study, providing a deeper understanding of user behavior and enabling proactive decision-making. They offer a comprehensive and forward-thinking perspective on the user experience.

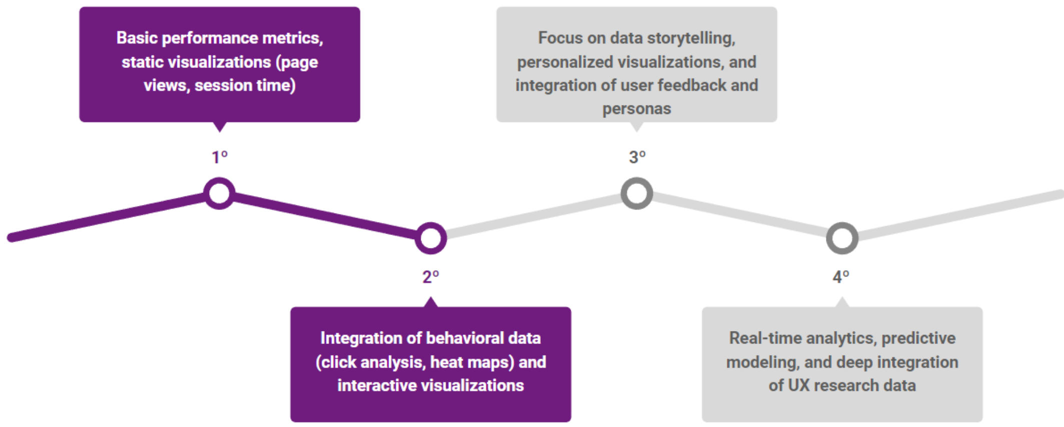


Figure 39 – UX data evolved from simple data representations to highly interactive, predictive UX research visualizations.

8.5 The taxonomy about UX data

The GLR data analysis reveals that the individual RQ responses were linked to one another, and the classification of results could offer perspectives useful for building a taxonomy. Taking into account the RQs answers, we see the characteristics of UX data (RQ2a), the tools and practices for integrating these data into the software development lifecycle (RQ1a), the motivations and benefits of these practices (RQ1b), and thedata visualizations commonly used to explore UX data (RQ2b) (see Section 8.4).

Similar to (GERMANAKOS; FICHTE, 2020), we decided to construct a taxonomy that provides a consistent and progressive flow of thinking and knowledge generation during the UX data analysis process.We considered our extractions that describe the exploration of meaningful *UX data* through actions (e.g., activities, tasks, approaches) undertaken by UX and software professionals with specific *purposes* (e.g., motivations and benefits discussed in Section 8.4) to achieve *objectives* related to UX improvements.

We understand that the taxonomy aids both academic and practitioners of UX and software development in adopting a UX data-driven approaches. From a practical perspective, the GRL pointed out that software professionals can enhance product innovation and engage user experiences that drive positive outcomes (e.g., by improving conversion rates) [AC85, AC127]. Based on UX data, the decision-making approach moves beyond intuition and experience to create a more objective and effective process [AC129]. For Human Computer Interaction community, the taxonomy represents a practical approach designed to understand, integrate, and visualize user interaction data.

8.5.1 Design of the taxonomy

Taxonomies provide a structured representation of concepts, enhancing the understanding and communication of emerging knowledge derived from data (GLASS; VESSEY, 1995; UNTERKALMSTEINER; FELDT; GORSCHEK, 2014). There are two approaches to developing taxonomies, *top-down* and *bottom-up* (UNTERKALMSTEINER; FELDT; GORSCHEK, 2014). In the top-down approach, the classification scheme is created from a composition of previously established themes, allowing the reuse of definitions and categorizations (BROUGHTON, 2019). The bottom-up approach is driven by the extraction of patterns on a subject of interest that are refined into a classification scheme (BROUGHTON, 2019). Taxonomy construction is inherently iterative and often requires revisions and progressive refinements to ensure its accuracy and relevance (GERMANAKOS; FICHTE, 2020).

To construct our taxonomy, we adopted a bottom-up approach by applying an open and closed coding process (SALDAÑA, 2021). First, we conducted open coding procedures, assigning codes to segments of data to provide them with meaning. These codes were not predefined and instead emerged organically from the data (SALDAÑA, 2021). From open coding procedures, three key labels emerged from the RQ answers that summarized the focus of our taxonomy: definitions what **data** is meaningful for UX-related approaches (extracted from the RQ2a answer); **objectives** that can be achieved through the exploration of UX data (derived from the RQ1b); and **purposes** for improving UX, materialized through actions performed by software professionals, e.g., activities, tasks, approaches, (emerged from RQ1a and RQ2b). We defined these three labels, i.e., *data*, *objective*, and *purpose*, as primary categories of our taxonomy.

After that, we applied the closed coding process to all 479 relevant excerpts identified during the GLR extraction step (see Subsection 8.3.6). In closed coding, predefined codes are assigned to chunks of text (SALDAÑA, 2021). To this step, we manually reviewed all 479 and labeled those that could provide meaningful guidance for each main category. Similar to (GERMANAKOS; FICHTE, 2020), our taxonomy data is based on empirical data that may contain statements, suggestions, challenges, lessons learned, and daily findings from real-life scenarios and situation-specific events.

After the manual labeling process, we asked ChatGPT-4o[3] to cluster the excerpts of each category and create a small sentence (i.e., sub-themes) that represents the content of each cluster. Figure 40 shows the flow and the basic prompt used to cluster the excerpts and generate the sub-themes. Due to the high volume of text in the excerpts (see Subsection 8.3.6), we split the excerpts based on the limit of characters of the ChatGPT and use the flow shown in Figure 40 for each part. Sometimes, ChatGPT may vary the answer for the same question. In such cases, we had to use the prompt *"When grouping notes by similarity, what themes would you create?"* to force it to create the small sentence that represents the cluster. In total, 64 sub-themes were categorized: 18 related to data (see Table 36); 20 related to objectives (see Table 37); and 26 related to purposes (see Table 38).

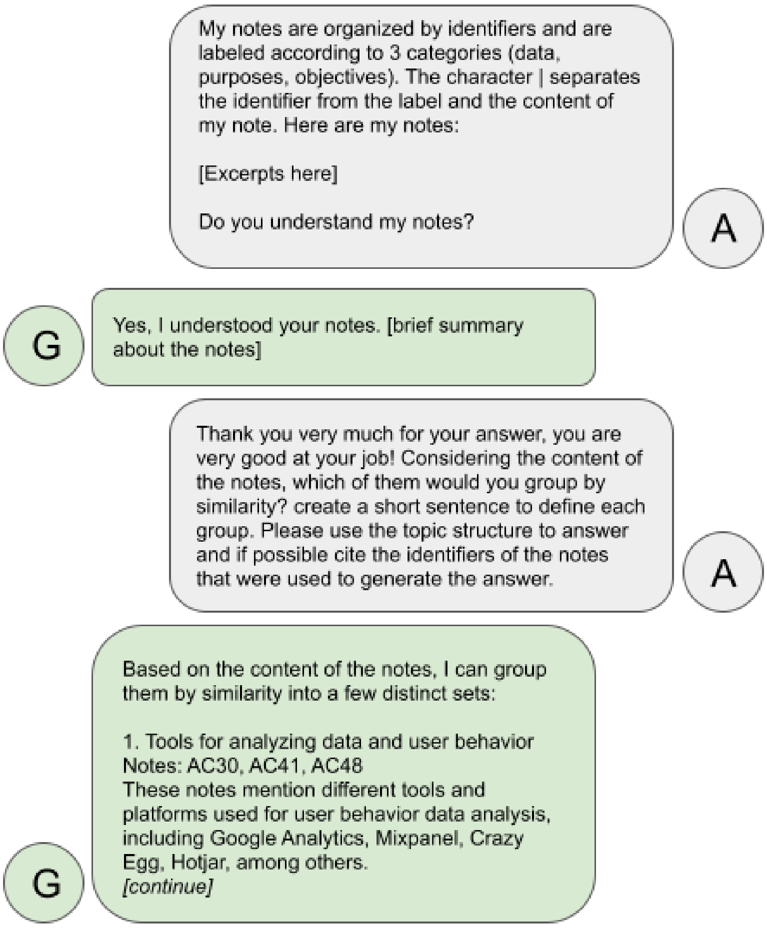


Figure 40 – Flow used with ChatGPT to cluster the excerpts. Letter A represents the authors, and G represents ChatGPT.

After the labeling process, we initiated the task of establishing links (i.e., relationships) between the sub-themes. To create the relationships, we developed two *validate statements*

[3] <https://chatgpt.com/>

ID	Sub-themes	References
1	Personas representing users	[AC63, AC79]
2	User opinions, feelings and perceptions about software usage	[AC57, AC63, AC79, AC89]
3	Documentation of interviews/questionnaires conducted with the user	[AC57, AC63, AC79, AC89]
4	Usability test results	[AC24, AC25, AC30, AC32, AC48, AC50, AC82, AC105, AC131]
5	User demographic information	[AC34, AC65, AC98]
6	User preferences	[AC64, AC65, AC69]
7	User expectations and motivations	[AC64, AC65, AC69]
8	User time using the system	[AC99]
9	User time on each system interface	[AC4, AC92]
10	Time it takes the user to complete the task	[AC24, AC25, AC30, AC32, AC82, AC105, AC131]
11	User visited interfaces	[AC99]
12	Behavior patterns found among users	[AC13, AC19, AC62]
13	User actions and interactions with the system	[AC24, AC25, AC30, AC32, AC82, AC105, AC131]
14	Bounce rate	[AC4, AC92]
15	Error rate	[AC24, AC25, AC30, AC32, AC82, AC105, AC131]
16	Click-through rate	[AC24, AC25, AC30, AC32, AC82, AC105, AC131]
17	Conversion rate	[AC13, AC19, AC24, AC25, AC30, AC32, AC62, AC82, AC99, AC105, AC131]
18	Abandonment rate	[AC4, AC24, AC25, AC30, AC32, AC82, AC92, AC105, AC131]

Table 36 – Sub-themes related to *data* information and its references in grey articles

(see the first and second column in Table 39). Researcher R1 used the two validated statements to check when the sub-themes have a relationship. A relationship is considered valid when the sub-themes match with the two statements to establish an approach for exploring UX data. We do not make statements to establish connections between sub-themes about *data* with the others because these sub-themes were naturally connected with *purposes*, and these connections were stored during the labelling step. Table 39 shows the two validation statements and examples of how they were used in the process (see third column).

These relationships represent a structured perspective of the taxonomy to display information in a logical sequence, tailored to any context related to the taxonomy's subjects (e.g., to know the achievable objectives based on existing UX data). In our taxonomy, **purposes** sets the central link that binds the desired **objectives** with the necessary and relevant **UX data**. This structure not only ensures clarity, but also promotes an iterative and cyclical process, allowing individuals to continuously navigate and refine their understanding of the taxonomy's information as they progress through their work. This design empowers individuals to effectively leverage the taxonomy, making it a tool for addressing UX data-related challenges.

8.5.2 The taxonomy

We structured our taxonomy into three primary categories that represent the essential resources for effectively utilizing UX data to enhance system interactions: objectives,

ID	Sub-themes	References
1	Validate or refute hypotheses	[AC5, AC23, AC41, AC42, AC48, AC50, AC59, AC81, AC92, AC98, AC105, AC130, AC134, AC138]
2	Generate new ideas	[AC51]
3	Define objectives	[AC24]
4	Formulate hypotheses	[AC23]
5	Perform tests and experiments	[AC140]
6	Adjust development strategies	[AC71]
7	Validate development progress	[AC5]
8	Evaluate system performance	[AC40, AC105, AC132]
9	Evaluate system maturity	[AC40, AC105, AC132]
10	Evaluate the success of the system	[AC40, AC105, AC132]
11	Identify usability issues	[AC4, AC40, AC42, AC84, AC92, AC100, AC102, AC118, AC130, AC134]
12	Evaluate how system changes affect user behavior	[AC120]
13	Evaluate how system changes affect overall product performance	[AC120]
14	Make decisions that are appropriate to the needs of the target audience	[AC2 e AC142]
15	Make design decisions focused on improving the user experience	[AC2, AC23, AC30, AC41, AC42, AC48, AC59, AC68, AC94, AC95, AC99, AC105, AC113, AC130, AC134]
16	Compare approaches to continually improve user experience	[AC38, AC52, AC60]
17	Improve user experience	[AC26, AC59, AC68, AC84, AC95]
18	Support the creation of more effective content strategies	[AC122, AC126, AC129]
19	Refine system navigation	[AC122, AC126, AC129]
20	Avoid assumptions in design choices	[AC37, AC92]
21	Optimize system usability	[AC1, AC6, AC29, AC59, AC68, AC109, AC123, AC130, AC131, AC134]
22	Validate decisions about system interactions	[AC4, AC48, AC76, AC81, AC87, AC89]
23	Validate system design decisions	[AC4, AC48, AC76, AC81, AC87, AC89]
24	Identify opportunities for system improvement	[AC2, AC22, AC63, AC91, AC98, AC103, AC117]
25	Measuring the effectiveness of different designs	[AC38, AC52, AC60]
26	Create experiences that improve user engagement	[AC1, AC6, AC29, AC59, AC68, AC109, AC120, AC123, AC126, AC130, AC131, AC134]

Table 37 – Sub-themes related to *objectives* information and its references in grey articles

purposes, and data. Each excerpt has been linked to a category and interconnected with the others, allowing individuals to engage with the taxonomy starting from any of the three categories and seamlessly explore the remaining ones (i.e., three-way access). The proposed taxonomy provides an abstraction of the key resources needed to leverage UX data effectively. By doing so, individuals can identify achievable *objectives*, grounded in actionable *purposes*, and supported by relevant *UX data*.

Figure 41 illustrates the structure of the information in the taxonomy. The figure illustrates the key elements of the taxonomy (i.e., rounded rectangles), the questions (i.e., rectangles in the middle of the figure) that explain the rationale behind the connections between the taxonomy's elements, and the three-way access (i.e., the different arrows). The three-way access provides the path options:

- ❑ (i) green solid arrows - the *objective* of performing an action to some UX-related activity (e.g., evaluate a system, understand the users' rates) is known: the path guides to the possible *purposes* for conducting such objective and the *UX data* that is required to be collected. Based on resources, they can determine if their *objective*

ID	Sub-themes	References
1	Understand how users interact with the system	[AC1, AC3, AC6, AC29, AC35, AC59, AC63, AC68, AC109, AC111, AC124, AC130, AC131, AC137, AC123, AC134]
2	Understanding user behavior	[AC9, AC18, AC26, AC76, AC111, AC131]
3	Understanding user preferences	[AC30, AC57, AC61, AC86, AC114, AC115]
4	Identify opportunities for improvement in functions of the system	[AC1, AC3, AC4, AC6, AC29, AC35, AC40, AC42, AC59, AC63, AC68, AC84, AC92, AC100, AC102, AC109, AC111, AC118, AC123, AC124, AC130, AC131, AC134, AC137]
6	Conduct tasks to collect data about the user	[AC30, AC57, AC61, AC86, AC114, AC115]
7	Generate insights for creating products based on user data	[AC27, AC30, AC57, AC61, AC75, AC86, AC91, AC92, AC114, AC115, AC121, AC138]
9	Determine relevant metrics	[AC27, AC75, AC91, AC92, AC121, AC138]
10	Provide data that supports design decisions	[AC27, AC75, AC91, AC92, AC121, AC138]
11	Create products that meet user expectations	[AC3, AC4, AC35, AC48, AC63, AC76, AC81, AC87, AC89, AC111, AC124, AC137]
12	Direct efforts to meet customer needs	[AC2, AC3, AC22, AC35, AC63, AC91, AC98, AC103, AC111, AC117, AC124, AC137]
13	Understand the system's performance in the market	[AC2, AC22, AC63, AC91, AC98, AC103, AC117]
14	Improve communication of goals among team members	[AC2, AC22, AC63, AC91, AC98, AC103, AC117]
15	Support team decision-making with user data	[AC85, AC96]
16	Understand user needs	[AC2, AC51, AC131, AC142, AC65]
17	Implement data-driven design	[AC2, AC5, AC23, AC41, AC50, AC59, AC69, AC92, AC94, AC98, AC105, AC113]
18	Analyze user behavior and understand how people interact with software	[AC5, AC51, AC53, A54, AC59, AC65, AC133]
19	Understand how users react to different elements of the software	[AC122, AC126, AC129]
20	Understand the user's subjective perception of the system	[AC25]

Table 38 – Sub-themes related to *purposes* information and its references in grey articles

First statement	Second statement	Result
I want [OBJECTIVE], but first I need [PURPOSE]	I need [PURPOSE] to achieve the objective of [OBJECTIVE]	baseline
I want [Evaluate the success of the system], but first I need [Understand how users interact with the system]	I need [Understand how users interact with the system] to achieve the objective of [Evaluate the success of the system]	OK
I want [Evaluate the success of the system], but first I need [Generate new ideas]	I need [Generate new ideas] to achieve the objective of [Evaluate the success of the system]	No, these sub-themes do not match to construct an approach

Table 39 – The validated statements and examples of application to determine connections between the sub-themes.

is achievable.

- (ii) blue dotted arrows - actions related to *purpose* are planned: the path suggests which *UX data* are required and the *objective* that can be achieved.

- (iii) orange dashed arrows - some *UX data* has been collected: the checklist of *purposes* that can be performed and the related *objectives* for that are make available.

Figure 42 illustrates an example in which the path starts by informing a known *objective* for conducting a UX activity, i.e., "Evaluate the success of the system". Based on this *objective*, the green arrows (see Figure 41) are followed to achieve the purposes of doing the activities and the UX data requested.

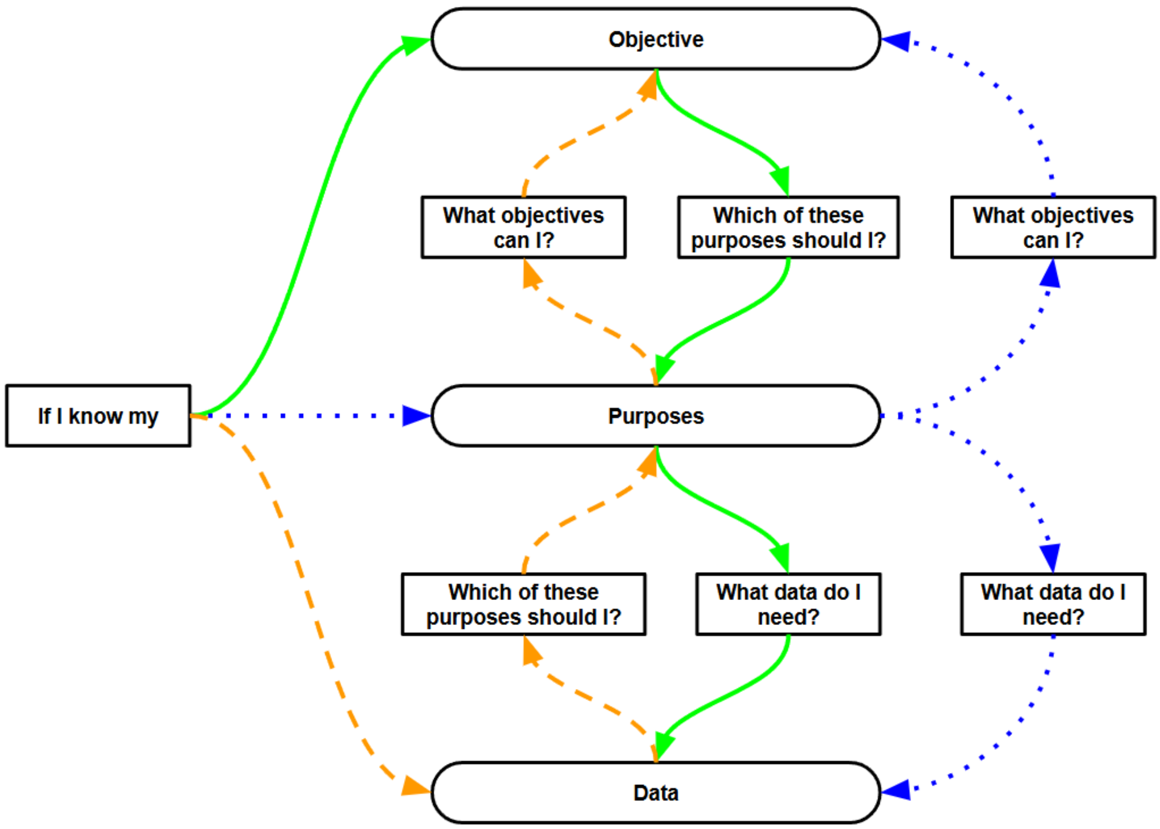


Figure 41 – The three-way access (arrows in different colors) of the information in the taxonomy.

We developed an interactive system (see Section 12.5) to facilitate access and dissemination of knowledge generated by the taxonomy. The system can be accessed here: UX Data Wiki (no login credentials are required). Two options for navigating the system are provided in the first user interface: explore the data in hands-free mode or guided mode. In guided mode, it requests information on which categories (i.e., objectives, purposes, or data) have already been defined, and the system presents the data related to the selected category. In hands-free mode, the user can interact with all the data and observe the connections between the categories of the taxonomy.

8.6 Discussion and Limitations

Our discussion is divided into three sections. First, we discuss the main findings of our work (see Subsection 8.6.1). Next, we highlight the main contributions and their similarities, as well as new findings in relation to related work (see Subsection 8.6.2). Finally, we pointed out the limitations of our work Subsection 8.6.3.

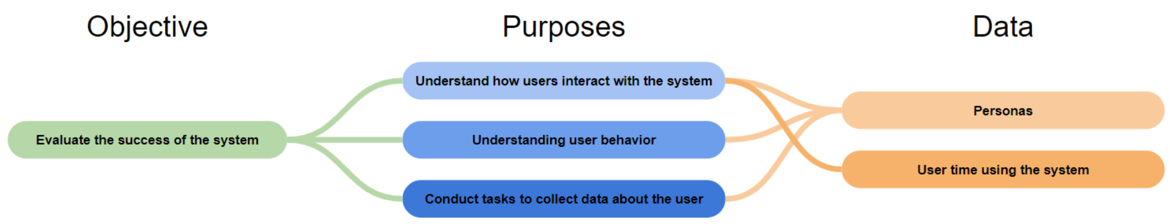


Figure 42 – Example of information gathered by following the green arrows path presented in Figure 41; the colors are applied to separate the categories.

8.6.1 Findings

In this work, we examined how UX data was managed in software development by conducting a GLR to explore knowledge from the practices available in articles. From an in-depth analysis of GLR articles. The GLR involved collecting 144 documents from online sources and performing a systematic extraction process. The findings successfully addressed our four research questions (see Subsection 8.3.2), revealing key practices, motivations, data characteristics, and visualization techniques related to UX work (i.e., UX design, research, evaluation). From an in-depth exploration of grey articles, we proposed a novel taxonomy that provides a structured set of categories to correlate UX data types and characteristics with their objectives, purposes, and the specific practices and visualizations used to derive insights for UX improvements.

Our findings indicate that key practices for integrating UX data into software development include defining UX goals, formulating hypotheses, conducting tests, and analyzing results to validate design decisions (see Subsection 8.4.1). These practices are driven by motivations such as understanding user behavior and closing the gap between expected and actual user experiences. However, a central challenge lies in the **fragmentation of responsibilities** across professional roles, which often disrupts the iterative flow of UX data (see Figure 36). When insights generated in early stages (e.g., research or analysis) are not effectively translated into decisions or implementation steps, the **data-driven process loses continuity** (see Subsection 8.4.1). This is particularly problematic when teams lack a shared understanding of UX goals and how to operationalize them.

The challenge of fragmentation of responsibilities becomes especially evident when examining the distribution of roles throughout the lifecycle (see Figure 36). UX researchers and designers typically lead data collection and analysis, while developers are more involved during the implementation phase. Product managers and analysts influence decision-making, but often rely on others for insight generation. As a result, misalignments in the activities of software development can occur; designers' early insights may not be carried through to implementation, and developers may lack the necessary context to apply UX findings meaningfully. We conclude that this disconnect between development activities highlights the need for **cross-role communication** and **more integrated workflows** to ensure that UX data not only informs but effectively shapes design and development decisions.

The characterization of UX data into behavioral (predominantly quantitative to offer objective insights) and attitudinal (primarily qualitative to capture subjective perceptions) is well-established (see Subsection 8.4.3). Our GLR findings highlight significant progress in supporting the integration of UX data, showcasing tools such as the HEART framework and methodologies, including thematic analysis, that extend UX data analysis beyond simple data collection (see Figure 38). These resources actively enable teams to move toward a synthesis of qualitative and quantitative insights. Considering the findings, we infer that, while the challenges teams face in achieving this synthesis are not always deeply explored, the presence of such **structured approaches** reflects an effort in the field to bridge the gap and operationalize UX data in practice.

In relation to data visualization, our review identified the use of conventional formats such as bubble charts, scatter plots, heat maps, dashboards, and UX scorecards to support comprehension and metric tracking (see Subsection 8.4.4). However, a key insight emerging from the grey literature is a growing emphasis on comprehensive visual storytelling, enabled by the integration of diverse data sources and a gradual shift toward real-time data integration and predictive analytics. This evolution reflects a **transition from static data representation to dynamic narrative-driven approaches** that enhance both decision-making and stakeholder engagement (see Figure 39).

According to our findings, dashboards that incorporate elements of data storytelling offer an accessible and interpretable format through which stakeholders can extract insights and contribute to decision-making processes throughout the development lifecycle. Furthermore, integrating both quantitative and qualitative data within visualizations provides a broader understanding of software performance from a user experience perspective. This dashboards not only strengthens the interpretability of UX outcomes but also fosters greater involvement from cross-functional teams by making insights more tangible and contextually relevant. Finally, the **adoption of interactive and real-time visualizations** represents a critical advancement. These tools allow stakeholders to explore multiple data streams simultaneously, gaining deeper insights into user behavior patterns and enabling proactive, rather than reactive, UX strategies.

Our findings converge into a novel UX data taxonomy. This taxonomy (see Section 8.5) **synthesizes the diverse practices, data characteristics, and visualization** techniques identified across our research questions (see Section 8.4). It provides a structured approach that correlates UX data types and their properties with the motivations for their use, and the specific practices and visualization methods employed. Organized as a practical, closed-loop cycle, the taxonomy aims to empower software professionals to effectively leverage user interaction data for enhanced overall UX.

8.6.2 Work Contributions

This work presents contributions that draw on the GRL and the proposed taxonomy. We understand that our contributions focus on discussing and providing evidence to support advances in structuring and systematizing UX data, thereby improving user-centered software development processes. Our findings clarified key practices and tools employed to integrate UX data throughout the software development lifecycle, along with the underlying motivations and benefits that drive its incorporation. Furthermore, we characterized UX data types and their leveraging in analysis, and mapped the visualization techniques commonly used to represent and analyze this data.

By comparing our taxonomy with related work (see Subsection 8.2.2, we could see that the taxonomy addresses the need for clear definitions and structured classifications to make UX data actionable, which was argued by (MENEWEGER et al., 2014) and (LEMON et al., 2020). While their works focus primarily on theoretical foundations and variability in interpretation, our taxonomy offers a practical guide for connecting UX data, objectives, and purposes in a sequential and relationship-based format.

In contrast to EUREKATAX (GERMANAKOS; FICHTE, 2020), which focuses only on qualitative usability feedback through hierarchical categorization, our taxonomy broadens the scope by integrating both qualitative and quantitative data. This integration provides a more comprehensive perspective that accommodates the diverse nature of UX data and its applications in various stages of the software lifecycle.

Furthermore, our taxonomy helps to fill the methodological gaps identified by (KALANTARI; LETHBRIDGE, 2022) by offering a nuanced categorization that facilitates the applicability of UX data exploration in different software tools. The inclusion of purposes as the pivot in our taxonomy ensures that objectives are directly linked to specific data requirements, which is a distinct feature compared to other approaches.

Our RQ findings (see Section 8.4) also support the insights of (LIN et al., 2023) by demonstrating how qualitative and quantitative data can be combined to address complex UX challenges. This integration enables teams to conduct richer analyses that go beyond surface-level metrics to understand user behavior, motivations, and outcomes holistically. This structured approach bridges the gap between theoretical insights and practical implementation, enhancing both the usability and effectiveness of UX-related initiatives.

We conclude that this work offers both theoretical and practical contributions to the field of Human-Computer Interaction (HCI). From a theoretical perspective, the taxonomy contributes to advancing conceptual models of data-driven design in the HCI field. From a practical standpoint, the taxonomy provides a systematic approach for software professionals to plan UX research by explicitly connecting UX data, objectives, and purposes. It can ensure that data collection is explicitly tied to clear objectives and goals (i.e., using the taxonomy as a checklist or planning tool). The emphasis on integrating both qualitative and quantitative data, as highlighted by our taxonomy and findings, enables practitioners to develop a broader understanding of user issues.

8.6.3 Limitations of the Work

This study presents limitations related to its methodological choices. First, the analysis relies on 144 grey literature sources, including industry blogs, company reports, practitioner forums, and non-peer-reviewed conferences. These sources vary in methodological rigor, objectivity, and depth. Although a systematic extraction protocol was applied, the heterogeneity of the material may have introduced bias—potentially amplifying emergent trends while underrepresenting established practices. To mitigate this, we prioritized sources that showed alignment with recognized UX principles (e.g., UX Collective, UX Plantet, UX Magazine, and NN Group).

The search strategy represents other limitations. It excluded non-English grey literature and focused specifically on the term "UX data". This decision aimed to reduce ambiguity and avoid an excess of irrelevant results commonly associated with broader terms such as "user data". However, this approach may have restricted the inclusion of practices and tools discussed in alternative terminologies, particularly in non-English contexts.

Additionally, while the proposed taxonomy offers a structured framework to classify and connect UX data, objectives, and purposes, it is not exhaustive. It is grounded in excerpts derived from the grey literature review and, despite their breadth, may not capture the full range of UX-related scenarios or data types. The effective application of the taxonomy also depends on the expertise of the user in identifying relevant data and articulating corresponding objectives and purposes. However, its practical utility could be enhanced by integrating real-time analytics and dynamic data visualization techniques, particularly in iterative software development environments.

8.7 Conclusion

In this paper, we presented a GLR conducted to identify and systematize the practices adopted by software professionals when working with UX data. A systematic process involving the analysis and categorization of 2262 grey literature articles led to the selection of 144 for in-depth analysis. From these selected articles, a total of 479 specific data extractions were performed, which formed the basis from which we distill the findings on how UX data is applied in software development practices. The GLR results revealed common practices, motivations for exploring UX data, methods for its characterization, and the visualization techniques employed to explore UX data. The extraction of the GLR supported a proposal for a comprehensive taxonomy to categorize UX data in practice.

The GLR findings stood out for the critical role of UX data in supporting practitioners by tracking metrics that facilitate usability issue diagnosis and enhancing data-driven decision-making for user-centered products. Using UX data, practitioners define quantitative metrics that represent UX goals; then, through testing to validate hypotheses, they determine which design choices align with user needs.

In addition to the GLR findings, we proposed a taxonomy to assist software professionals in deriving UX objectives and insights based on existing UX data. From a practical perspective, our taxonomy underscores the value of continuous UX data collection for identifying usability issues, diagnosing patterns, and informing product refinements. From a research standpoint, it serves as a tool that synthesizes the objectives of doing UX activities and the UX data related to. The taxonomy provides a foundation for developing new methodologies and studies on the relationship between UX data and software development practices, fostering further investigations into the execution of UX tasks by software professionals.

We consider the taxonomy a significant contribution, ensuring that practitioners do not miss opportunities to integrate UX tasks into the development lifecycle to enhance user experience. It also provides researchers with a systematic approach to UX data definition, enabling the development of new UX research methodologies. Moreover, the adoption of GLR as a research method can be regarded as a contribution for researchers in the HCI field; as far as we know, few works in the field embrace this method that can reveal findings from practice. Ultimately, we hope our contributions simplify, democratize, and encourage the adoption of UX data-driven practices within the software development lifecycle.

As future work, we plan to validate and extend the taxonomy through empirical studies, this involves: longitudinal case studies applying the taxonomy in real-world projects; controlled experiments comparing design outcomes with and without using the taxonomy; integrate new dimensions of UX data, such as AI-driven UX approches; develop an artificial intelligence-powered agent to provide recommendations and answer questions about UX data based on the taxonomy; and surveys to assess its perceived usability and utility among practitioners.

8.8 Final Considerations

This chapter presented a taxonomy for UX data exploration, developed through a systematic grey literature review of 144 articles. This work aims to structure and systematize UX data to improve user-centered software development processes. The taxonomy addresses a gap in clear definitions and structured classifications by offering a practical guide that connects UX data, objectives, and purposes. It expands beyond qualitative feedback by integrating both qualitative and quantitative data, offering a broader perspective on

UX data's diverse nature and applications. The findings underscore the critical role of UX data in identifying usability issues, diagnosing patterns, and informing product refinements, ultimately supporting data-driven decision-making for user-centered products. The taxonomy provides both theoretical contributions by advancing conceptual models of data-driven design and practical contributions by offering a systematic approach for software professionals to plan UX research. The work contributes to democratizing and encouraging the adoption of UX data-driven practices within the software development lifecycle.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 and grant 309497/2022-1, Consellho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - Brazil).

About the authors

Maylon Macedo is a Ph.D. candidate and UXLERIS member focusing on Interactive Systems Engineering. He has experience in web design and interface development, and teaches programming languages and competitive programming. His interests include software engineering, User Experience (UX), Information Visualization, and the application of GenAI as a research tool.

Luciana Zaina is a Full Professor at the Federal University of São Carlos, Brazil. Her expertise includes UX design in software development for different domains, e.g., startups and the automotive industry. She has led research projects and contributed as a reviewer, journal editor, and co-chair in Brazilian and international communities.

Chapter 9

UX Data Wiki

This chapter outlines the implementation of the taxonomy as an interactive web system, making it operational as Step E of the overall thesis methodology. The objective of this work is to demonstrate how the system enables professionals to interactively navigate and understand the taxonomy, connecting analysis objectives, purposes, and UX data types, with its online deployment ensuring accessibility from any device. The contributions of this chapter include providing a functional tool for exploring the taxonomy and detailing its evaluation, where software professionals used the system to assess potential insights derived from a given dataset.

9.1 Introduction

While a UX data taxonomy has been developed to categorize and structure information across analysis objectives, their corresponding purposes, and specific UX data characteristics, the relationships between the taxonomy dimensions can complicate practical application. The textual representation of the taxonomy, though comprehensive, requires software practitioners to manually trace relationships between these three dimensions, often hindering intuitive comprehension and efficient data utilization.

To address this challenge, an interactive web-based system was designed and implemented. This web system, called UX Data Wiki, operationalizes the conceptual taxonomy by providing a web interface that facilitates navigation and understanding of the taxonomy content. By enabling interactive exploration of the taxonomy, for instance, allowing users to click on an objective to immediately visualize its associated purposes and required UX data, the web system transforms a static knowledge base into a dynamic, and interactive resource. This chapter details the design, implementation, and evaluation, demonstrating the utility of the UX data taxonomy (i.e., used through the UX Data Wiki) for practitioners.

9.2 System Design and Implementation

The UX Data Wiki was developed with the primary goal of providing an interactive interface for the multifaceted UX data taxonomy. The design prioritized the navigation through the UX data taxonomy dimensions and relationships (i.e., objectives, purposes, and the required UX data). The system's architecture comprises a **frontend** for user interaction, a **backend** to retrieve information from database and load in the interface, and coupled with a **database** for storing the relationships between objectives, purposes, and UX data characteristics as defined in Chapter 8.

The **frontend** (i.e., web interface) was developed with HTML, CSS, and JavaScript. The **backend** was made in PHP. The **database** selected to store the structured knowledge of the UX data taxonomy was a MySQL database. The UX Data Wiki was implemented using only standard web technologies to ensure broad compatibility across various devices, and ensures that software professionals can access the system from any device with an internet connection, making its structured knowledge available for practical application.

The formal grammar and relational structure of the UX data taxonomy (see Subsection 8.5.1) were translated into the database schema. This translation ensures that the complex, interconnected nature of the taxonomy, where elements across objectives, purposes, and data are linked, is preserved and reflected within the system's data model, enabling querying and visualization of these relationships.

The core functionalities implemented facilitate two primary modes of interaction for users to explore the taxonomy: **Hands-free Mode** allows users to navigate through organized lists of objectives, purposes, and UX data definitions, providing a comprehensive overview of each dimension independently. **Guided Mode** enables users to click on a specific item (e.g., an objective) and instantly visualize all its directly associated information across the other dimensions (i.e., relevant purposes and required data). This direct interaction eliminates the need for manual cross-referencing between the taxonomy dimensions.Figure 43 shows the homepage of UX Data Wiki, where users can start the Guided Mode. If the user selects the objective dimension, a list of available objectives is displayed (see Figure 44). Figure 65 illustrates the information presented when a specific objective is selected from the list.

9.3 Evaluation

The primary objective of this evaluation was to determine whether consultation of the taxonomy, through the UX Data Taxonomy, would enhance the accuracy in identifying

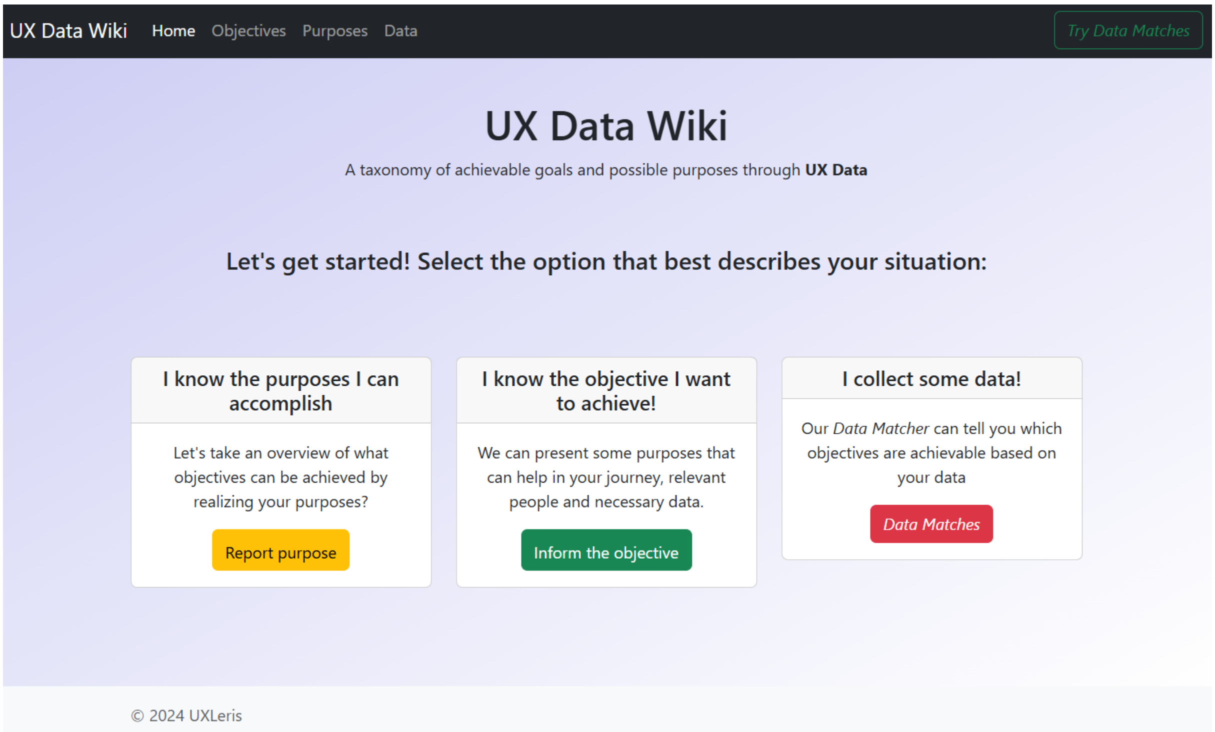


Figure 43 – Homepage of UX Data Wiki.

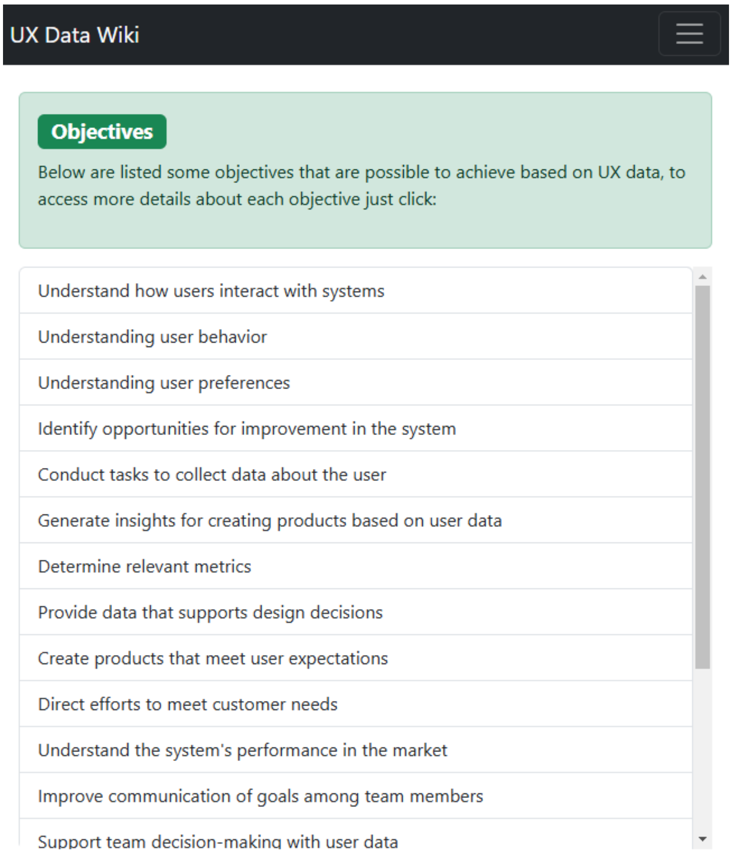


Figure 44 – Example of objectives available in UX Data Wiki.

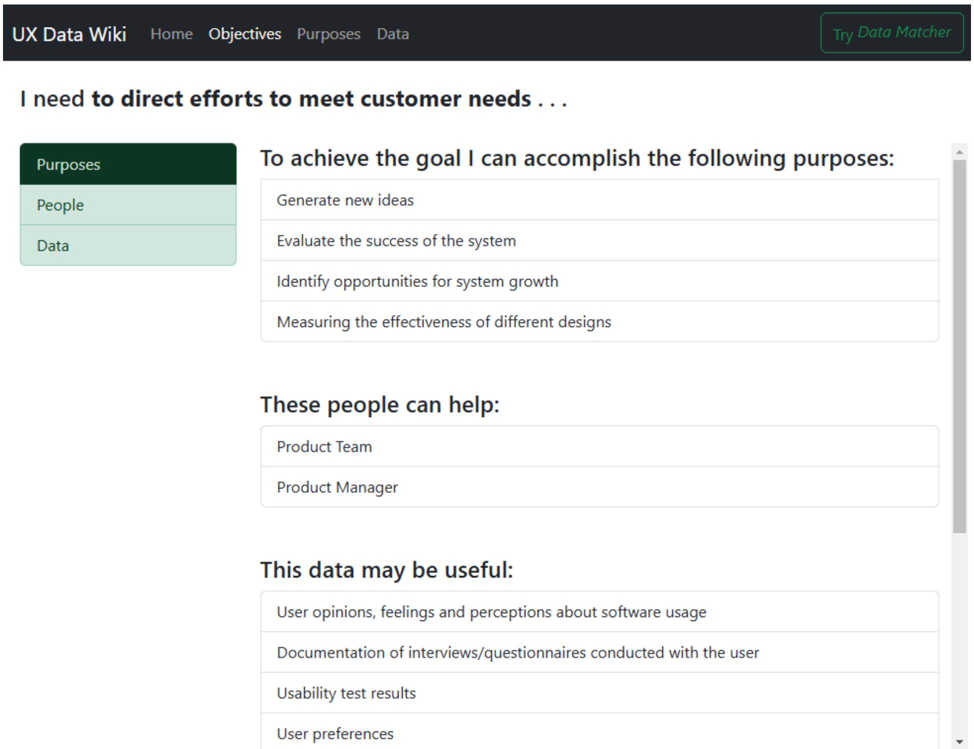


Figure 45 – Information showed when the user selected one objective.

achievable UX objectives from a given dataset and reduce the perceived cognitive load associated with this task. This led to the following research question: “Does consulting the UX Data Wiki improve the ability of software professionals to correctly identify achievable UX objectives from a clickstream dataset?”.

9.3.1 Study design

The evaluation was structured to assess the impact of the UX Data Wiki on objective identification and perceived workload. The procedure started with a recruitment phase, where potential participants were invited via professional social media platforms (i.e., Linkedin). Acceptance of the invitation was collected upon completing a profile questionnaire, which served to gather relevant professional experience details and facilitate subsequent scheduling of individual evaluation sessions. Following recruitment, participants were assigned to one of two distinct groups: Group A, which had access to the UX Data Wiki during the task, and Group B, which performed the task without this access. Each participant was then guided through a standardized task involving the analysis of a clickstream dataset to identify achievable UX objectives.

The evaluation procedure followed a standardized sequence: participants first explored the clickstream dataset and its data dictionary, and, when applicable (i.e., for Group A), the UX Data Wiki, before analyzing a list of 18 pre-defined UX objectives to determine which were achievable or not with the provided data. All apparatus could be accessed as many times as necessary.

The evaluation utilized the following apparatus:

- **Clickstream Dataset:** consisting of user interaction logs collected from an e-commerce platform. This dataset included detailed information about individual user sessions, such as page views, time spent on each page, actions performed, session duration, and metadata about the product. It served as the foundational data source for identifying achievable UX objectives, providing a behavioral snapshot of user navigation and interaction.

- **Data Dictionary:** document that provided explanations for each column available in the clickstream dataset, ensuring participants understood the data fields.

- **UX Data Wiki:** interactive web-based system developed as part of this research, which operationalizes the UX data taxonomy. This system provided access to the structured knowledge of objectives, purposes, and UX data.

- **Forms:** used as the interface for participants to respond to the task. It presented 18 pre-listed UX objectives, where participants were required to mark checkboxes indicating which objectives they believed were achievable using the provided dataset.

- **NASA-TLX (Task Load Index):** standardized assessment tool administered immediately following the objective identification task to measure participants' perceived cognitive workload (HART; STAVELAND, 1988).

To assess the accuracy of objective identification, a pre-determined list of objectives (see Table 40) that were achievable with the given clickstream dataset was established as a baseline. Participants' selections in the Form were compared against this baseline. To measure the perceived cognitive load, the NASA Task Load Index (TLX) was administered to all participants.

9.3.2 Execution

A total of six software professionals participated in the study. All participants were recruited via invitations extended on professional social media platforms (e.g., LinkedIn). The recruitment targeted individuals who actively participated in discussion groups and online communities focused on User Experience (UX), ensuring that the sample included professionals engaged with UX-related topics and practices. Their ages ranged from 22 to 41 years. Of the six participants, five possessed more than two years of professional experience in roles focused on the design of interactive software systems, while one was a trainee. This division aimed to assess whether less experienced participants could reach conclusions comparable to those of the more experienced ones, using the latter's responses

ID	UX objectives
UX1	Understand how users interact with systems
UX2	Understanding user behavior
UX3	Understanding user preferences
UX4	Identify opportunities for improvement in the system
UX5	Conduct tasks to collect data about the user
UX6	Generate insights for creating products based on user data
UX7	Determine relevant metrics
UX8	Provide data that supports design decisions
UX9	Create products that meet user expectations
UX10	Direct efforts to meet customer needs
UX11	Understand the system's performance in the market
UX12	Improve communication of goals among team members
UX13	Support team decision-making with user data
UX14	Understand user needs
UX15	Implement data-driven design
UX16	Analyze user behavior and understand how people interact with software
UX17	Understand how users react to different elements of the software
UX18	Understand the user's subjective perception of the system

Table 40 – The 18 pre-determined UX objectives.

as a proxy for expert answers. All participants were employed at well-established companies, with those in Group B holding specific responsibilities related to UX data within their respective organizations. The evaluation was conducted individually via Google Meet video calls, scheduled by participants at their preferred dates after receiving the study invitation. Each participant was allotted one hour to complete the task. Table 41 presents the participants individual profile.

ID	Age	Role	Experience	InfoVis	UX	Analytics
P1	35	UX Researcher	+ 2 years	Strong expertise	Reasonable expertise	Strong expertise
P2	37	UX Researcher	+ 4 years	Strong expertise	Strong expertise	Strong expertise
P3	23	Developer	+ 1 year	Reasonable expertise	Reasonable expertise	Reasonable expertise
P4	22	Developer	+ 2 years	Little expertise	Little expertise	Little expertise
P5	41	Software Engineer	+ 4 years	Strong expertise	Strong expertise	Strong expertise
P6	23	Developer	+ 2 years	Strong expertise	Little expertise	Reasonable expertise

Table 41 – Profile of the participants

9.3.3 Results

The analysis focuses on the accuracy of UX objective identification by participants across the two groups (i.e. A and B), and explores specific instances where the operationalized taxonomy may have contributed to a more objective understanding. The accuracy of objective identification was calculated as the percentage of correctly identified objectives out of the total 18 pre-defined UX objectives. Figure 46 presents the individual responses of the participants, categorized by their respective study groups (i.e., Group A: with UX Data Wiki access; Group B: without UX Data Wiki access), the final row represents the ground truth for each objective's achievability as determined for the study.

Considering the average performance across groups, Group A (i.e., with UX Data Wiki access) achieved a mean accuracy of 67.78%, while Group B (i.e., without UX Data Wiki

ID	Group	UX1	UX2	UX3	UX4	UX5	UX6	UX7	UX8	UX9	UX10	UX11	UX12	UX13	UX14	UX15	UX16	UX17	UX18
P1	B	x	x	x	x				x			x				x	x	x	
P2	B	x					x	x	x	x				x		x	x	x	
P3	A	x		x	x		x	x											
P4	A	x	x	x	x	x		x	x		x	x	x	x			x		
P5	A	x		x			x			x				x					
P6	A	x		x				x			x	x	x	x	x		x		
K	-	x		x	x		x	x	x		x	x	x	x		x	x		

Figure 46 – Responses across 18 UX objectives. “X” indicates affirmative responses. “K” indicates baseline response.

access) achieved a mean accuracy of 55.56%. Group A demonstrated accuracies ranging from 61.11% to 72.22%. Group B showed accuracies ranging from 50.00% to 61.11%.

ID	Group	Correct Identifications	Incorrect Identifications	False Positives	False Negatives	Total Marked	Accuracy
P1	B	11	7	7	0	11	61.11
P2	B	9	9	5	4	9	50.00
P3	A	13	5	1	4	9	72.22
P4	A	13	5	3	2	11	72.22
P5	A	11	7	1	6	7	61.11
P6	A	12	6	2	4	10	66.67

Table 42 – Individual Participant Analysis of UX Objective Identification.

A key observation from the analysis pertains to the distinct response patterns exhibited by the two groups, reflecting differing approaches to objective identification.

Group B displayed a more comprehensive approach, tending to mark a larger number of objectives as achievable. This breadth in selection indirectly contributed to their inclusion of correct objectives, even if it came with a higher rate of false positives. Specifically, Group B exhibited an average of 6 false positives per participant (i.e., P1 marked 7 and P2 marked 5), indicating a propensity to incorrectly identify objectives as achievable when they were not supported by the dataset according to the baseline. This suggests that, without the structured guidance of the taxonomy, participants in Group B adopted a broader strategy, potentially driven by their experience or an inclination to cover all perceived possibilities.

In contrast, Group A adopted a more conservative approach. On average, participants in Group A marked fewer objectives as possible. This group demonstrated precision, resulting in a significantly lower average of 1.75 false positives per participant (see Table 42). This pattern, characterized by fewer incorrect inclusions, suggests that the UX Data Wiki may have facilitated a more objective understanding of the relationship between the dataset and the achievable objectives. By providing a structured knowledge, the taxonomy likely helped participants in Group A to discern what could and could not be inferred from the given clickstream data, thus leading to more targeted and precise selections.

An analysis of individual objectives illuminates where the taxonomy provided a more objective vision for Group A, particularly in discerning the capabilities and limitations of the clickstream dataset.

For the objective "Understand how users react to different elements of the software" (Baseline: Not Possible), Group A demonstrated 100% correctness by accurately identifying it as not achievable with a clickstream dataset. In stark contrast, Group B showed 0% correctness, indicating that all participants in this group incorrectly judged it as possible. This highlights a insight that the taxonomy appears to have provided Group A with the necessary clarity to recognize the limitations of clickstream data for capturing subjective user reactions.

For objectives such as "Understanding user preferences" (Baseline: Possible), Group A achieved 100% correctness and Group B only 50%. Similarly, for "Determine relevant metrics" and "Support team decision-making with user data" (both Baseline: Possible), Group A surpassed Group B (75% vs. 50% for both). This suggests that the structured knowledge of the taxonomy guided Group A participants more effectively in making the correct connections between the clickstream data's capabilities and these specific objectives, promoting a more objective interpretation.

Conversely, for some objectives (e.g., "Implement data-driven design", "Provide data that supports design decisions", and "Direct efforts to meet customer needs"), Group B showed 100% correctness while Group A performed lower. The higher professional experience of Group B participants may also have contributed to their performance on these specific objectives, allowing them to accurately assess achievability even without explicit taxonomic guidance.

The TLX was administered to assess the perceived workload associated with identifying UX objectives from the clickstream dataset. The analysis of perceived workload provides insights into the difficulty of analyzing UX data without the structured assistance of a taxonomy. The individual and average group scores for total workload and its six dimensions are presented in Table 43 and Table 44, respectively.

ID	Group	Total TLX	Mental	Physical	Temporal	Performance	Effort	Frustration
P1	B	67.00	213.00	0.00	284.00	91.00	62.00	355.00
P2	B	63.00	430.00	0.00	51.00	32.00	364.00	63.00
P3	A	87.00	500.00	0.00	132.00	66.00	400.00	213.00
P4	A	46.00	228.00	0.00	304.00	30.00	122.00	1.00
P5	A	41.00	355.00	0.00	108.00	48.00	93.00	11.00
P6	A	44.00	300.00	0.00	46.00	33.00	198.00	80.00

Table 43 – The individual scores for total workload

Group	Total TLX	Mental	Physical	Temporal	Performance	Effort	Frustration
A	54.50	345.75	0.00	147.50	44.25	203.25	76.25
B	65.00	321.50	0.00	167.50	61.50	213.00	209.00

Table 44 – The group average scores for total workload

On average, Group B reported a higher total perceived workload (i.e., TLX = 65.00) compared to Group A (i.e., TLX = 54.50). This suggests that the absence of a guiding framework, such as the UX Data Wiki, increased the overall cognitive burden on participants attempting to identify achievable objectives from the raw clickstream data.

While **mental demand** shows Group A with a slightly higher mean (345.75 vs 321.50), the **frustration level** stands out. Group B reported a notably higher average **frustration level** (i.e., 209.00) compared to Group A (i.e., 76.25). This significant difference indicates that navigating the unstructured dataset and attempting to map it to objectives without the taxonomy's assistance led to considerably more feelings of annoyance, discouragement, and stress among participants in Group B. The lack of a clear, interactive guide likely made the task more ambiguous and prone to dead ends, thereby increasing frustration.

Although **effort** shows Group B with a slightly higher mean (213.00 vs 203.25), the higher frustration in Group B, combined with their lower accuracy in objective identification, strongly implies that the effort exerted without the taxonomy was less efficient or effective. The perceived **performance** (i.e., lower scores mean better perceived performance) was also slightly worse for Group B (i.e., 61.50) than Group A (i.e., 44.25), further supporting the notion that working without the taxonomy led to a less satisfactory outcome despite the effort.

These findings strongly suggest that the UX data taxonomy, operationalized through the interactive web system, significantly contributes to reducing the perceived difficulty and frustration associated with analyzing complex UX datasets for objective identification. This highlights the practical value of such a tool in making data analysis tasks more manageable and less burdensome for software professionals, directly addressing the challenge of working with unstructured data.

9.3.4 Discussion

This study aimed to investigate whether the operationalization of the UX data taxonomy through an interactive web system (i.e., UX Data Wiki) would improve the accuracy of identifying achievable UX objectives and reduce perceived workload for software professionals. The results provide valuable insights into the utility of structured knowledge in navigating complex data analysis tasks.

The primary hypothesis, suggesting that utilizing the taxonomy would lead to higher accuracy and lower perceived workload, was partially supported. While Group A (i.e., with UX Data Wiki access) achieved a higher average accuracy (i.e., 67.78%) compared to Group B (i.e., 55.56%), the most compelling insights emerged from the qualitative differences in response patterns and the workload assessment.

Group B tendency toward a more comprehensive approach, marking a greater number of objectives as possible, led to more false positives (i.e., an average of 6.00 per participant). This suggests that without the structured guidance of the UX Data Wiki, professionals might adopt a less precise strategy, overestimating the scope of insights derivable from the dataset. For this study, the higher level of professional experience of Group B, did not inherently translate into a more precise mapping of objectives to data capabilities in this specific context.

Conversely, Group A exhibited a more conservative and precise approach, reflected in a significantly lower rate of false positives (i.e., an average of 1.75 per participant). This indicates that the UX Data Wiki likely fostered a more objective understanding of the relationship between the clickstream dataset and the achievable UX objectives. The taxonomy structured knowledge appears to have aided participants in discerning the specific boundaries of what could and could not be inferred from the data, leading to more targeted selections.

This is particularly evident in objectives like "Understand how users react to different elements of the software", where Group A had 100% accuracy in identifying it as not achievable contrasted with Group B 0% correctness. This specific finding strongly underscores the taxonomy ability to clarify data limitations, a key aspect of UX data analysis.

Beyond accuracy, the perceived workload provided a strong indication of the taxonomy value. Group B reported a higher overall workload and, notably, significantly higher levels of frustration compared to Group A. This suggests that performing the task without the interactive taxonomy was a less satisfying experience. The effort required to trace relationships directly from raw data likely contributed to increased cognitive load and feelings of discouragement. The UX Data Wiki, by providing an intuitive and accessible interface, appears to have mitigated these challenges, making the task of identifying relevant objectives more manageable and less frustrating.

In summary, while the overall accuracy difference was present, the more significant implications lie in the reduction of cognitive burden and frustration. The UX data taxonomy, operationalized through the UX Data Wiki, effectively bridges the gap between complex theoretical knowledge and practical application. It enables a more objective process for software professionals seeking to leverage UX data for informed decision-making. The UX Data Wiki serves not merely as a data repository, but as a tool that guides the analytical process of understanding what data is needed to achieve a predetermined objective.

9.4 Study Validity

We considered the conclusion, construction, internal, and external threats to discuss our study validity according to Wohlin et al. (2012) recommendations. To mitigate threats to conclusion validity, the baseline (i.e., expected answers) used for accuracy calculation was established by expert consensus, ensuring measurement reliability. Furthermore, the standardized NASA-TLX instrument was employed to reliably measure perceived cognitive workload, and a one-hour time limit was set for all participants to prevent time-related confounding variables.

To mitigate construction validity problems, all participants interacted with the same standardized artifacts, including the clickstream dataset and the data dictionary. The clear operationalization of the core construct (i.e., Taxonomy Access) by providing only Group A with the interactive UX Data Wiki, ensured that differences in performance could be attributed to the taxonomy's structured knowledge. Fatigue was minimized by limiting the voluntary task to one hour at most, ensuring motivation was driven by intrinsic interest in the topic. Finally, concerning external validity, while the sample size was small, the inclusion of professionals in relevant roles provides ecological validity for the domain studied. Moreover, the task of mapping objectives to an unfamiliar dataset was new for all participants. The novelty of the task, supported by evidence from Salman, Misirli e Juristo (2015) enhances the confidence that the observed effects are attributable to the taxonomy's utility rather than pre-existing data-mapping experience.

9.5 Final Considerations

This chapter's study investigated the impact of operationalizing the UX data taxonomy via an interactive web system, the UX Data Wiki, on software professionals' ability to identify achievable UX objectives and manage their workload. While the group with access to the UX Data Wiki showed higher average accuracy, the more significant findings emerged from qualitative differences in response patterns and workload assessments. The UX Data Wiki appeared to foster a more precise and conservative approach, significantly reducing false positives and aiding participants in objectively understanding data limitations. Crucially, the system also led to a reduction in perceived workload and frustration, indicating that the interactive taxonomy effectively bridges the gap between complex theoretical knowledge and its practical application. The UX Data Wiki functions as more than just a data repository; it guides the analytical process, enabling more objective decision-making by clarifying what data is necessary for specific objectives.

Chapter 10

Future perspectives of using GenAI in UX data analysis

This chapter presents the article "Can GenAI Provide Insights Like a UX Professional? A comparison of Humans and GenAI rationale about UX data", authored by Maylon Macedo and Luciana Zaina. This article was presented at the Workshop on Engineering Interactive Systems Embedding AI Technologies (EISEAIT) on the ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS) in 2025. This work constitutes Step H of the overall thesis methodology. Its objective is to investigate the future perspectives of using Generative AI (GenAI) in UX data analysis, specifically by comparing the capabilities of GenAI with those of UX professionals in data analysis. The contributions of this work include highlighting the continued relevance of the research in a practical application scenario for GenAI and underscoring the importance of structured knowledge bases, even for advanced AI systems. In this chapter, references to **ChatGPT** specifically denote the use of the ChatGPT-4o model, guided by the custom prompts developed for this study. Therefore, the reported results pertain to this controlled context rather than to general responses from ChatGPT.

10.1 Introduction

Generative AI (GenAI), such as large language models (LLMs), is reshaping practices across digital product development cycles (KHALIFA; ALBADAWY, 2024). Among these transformations, GenAI presents an emerging opportunity to support user experience (UX) research, particularly in tasks involving the interpretation of user interaction data (e.g., navigation logs or clickstream events) (CHENG; LI; BING, 2023; HOU et al., 2024). At the same time, professionals who perform research on UX rely not only on technical knowledge, but also on interpretive and context-aware judgment regarding user behavior, business goals, and technological constraints (HASSENZAHL, 2018; MARTINELLI; LOPES; ZAINA, 2022). This context raises a critical question: To what extent can GenAI simulate the rationale of human experts when analyzing user behavior data? However, while these technologies demonstrate efficiency in repetitive tasks (e.g., pattern detection), their potential to replicate the judgment with background in contextual information of UX Research professionals remains underexplored (HOU et al., 2024). This debate is critical because analyzing interaction data (e.g., navigation logs) in UX transcends technical processing: it requires interpreting behavioral nuances, aligning with business goals, and creative hypothesis formulation (MACEDO; ZAINA, 2024; LUTHER; TIBERIUS; BREM, 2020).

Prior research has shown that automatically generated personas can appear authentic yet still face issues of credibility and coherence, reinforcing the need to refine AI-driven user representations (SALMINEN et al., 2019). Other studies explored GenAI for both creating and interacting with personas, calling for deeper investigation into its influence on user modeling (SCHMIDT et al., 2024a). Additionally, interaction data has proven valuable for informing system design, and GenAI has shown promise in helping interpret such data to support user-centered decisions (GARDEY et al., 2022).

This paper explores how well GenAI can replicate expert UX rationale in assessing which research objectives related to UX (i.e., UX objectives) a dataset can support, based on a comparison with six professionals analyzing e-commerce clickstream data. All participants (i.e., human and ChatGPT) were asked the same 18 yes-or-no questions, each aligned with a specific UX objective (e.g., “Can this dataset help understand user behavior?”). These questions were designed to probe the perceived applicability of the clickstream dataset for various UX insights. To simulate human rationale, the GenAI was configured using customized prompts that incorporated persona profiles matching each participant. This paper contributes by discussing the role of GenAI as a complementary tool in UX research practice, its current limitations, and the potential risks of delegating UX interpretive tasks to artificial agents.

10.2 Methodology

This research adopts an exploratory approach to evaluate the potential of generative artificial intelligence (ChatGPT) to simulate the rationale of UX professionals in interpreting user interaction data. Specifically, it investigates whether ChatGPT can reproduce expert judgment when determining which UX research goals can be supported by a dataset of user behavior (i.e., clickstream data). This study was approved by the Research Ethics Committee of the UFSCar, under protocol number 68524023.0.0000.5504. All data collected from human participants were anonymized. No personally identifiable information (e.g., name and email) was provided to ChatGPT at any stage of the experiment.

Apparatus: three main materials were used in this study. The clickstream dataset[1] consisted of user interaction logs collected from an e-commerce platform. It included detailed information about individual user sessions, such as page views, time spent on each page, actions performed, session duration, and metadata about the product. The dataset provided a behavioral snapshot of how users navigated and interacted with the platform, offering a foundation for evaluating whether specific UX objectives could be addressed using this type of data. A UX data taxonomy (MACEDO; ZAINA, 2025) developed to support the interpretation and categorization of UX research goals. The taxonomy provides a mapping between *UX objectives*, *purposes*, the *data needed* to achieve them (e.g., behavioral logs, user profiles, survey responses), and the *professional roles* typically involved in such tasks. This taxonomy was delivered to human participants through an interactive web interface[2], allowing for dynamic exploration of the information. For ChatGPT, the taxonomy was provided in CSV format as part of the prompting process. Finally, 18 UX objectives were made available to the participants. They represent common goals in UX research. The UX objectives reflect a broad spectrum of design and research intentions, ranging from exploratory analysis to strategic software development. Table 45 presents the complete list of the UX objectives used in this study.

ID	UX objectives
UX1	Understand how users interact with systems
UX2	Understanding user behavior
UX3	Understanding user preferences
UX4	Identify opportunities for improvement in the system
UX5	Conduct tasks to collect data about the user
UX6	Generate insights for creating products based on user data
UX7	Determine relevant metrics
UX8	Provide data that supports design decisions
UX9	Create products that meet user expectations
UX10	Direct efforts to meet customer needs
UX11	Understand the system's performance in the market
UX12	Improve communication of goals among team members
UX13	Support team decision-making with user data
UX14	Understand user needs
UX15	Implement data-driven design
UX16	Analyze user behavior and understand how people interact with software
UX17	Understand how users react to different elements of the software
UX18	Understand the user's subjective perception of the system

Table 45 – The 18 UX objectives derived from the UX data taxonomy.

Human participants: recruited via invitations on professional social media platforms (i.e., Linkedin), who were aged 22-37 years; 5 had 2+ years professional experience and 1 was a trainee (i.e., 6 participants in total). All participants were employed at well-established companies. Their professional roles focused on the design of interactive software systems. All participants worked at companies based in the same country. Their

1 <https://archive.ics.uci.edu/dataset/553>
2 <https://uxleris.net/uxdatawiki/>

responsibilities included tasks such as journey mapping, creating product/channel indicators, analyzing behavioral and attitudinal user data, conducting interviews and usability testing, gathering data directly from community members, field research, discovery activities, and analytics. The tools used in their daily work included Maze, Looker Studio, Power BI, Hotjar, spreadsheets, whiteboards, MaxQDA, NVivo, R, Python, heat maps, interaction replay tools, in-app event tracking, and Google Analytics. These participants were selected due to their direct involvement in UX activities that require critical interpretation of user data and the formulation of research insights. Their experience with tools and methodologies commonly used in UX research positioned them as suitable benchmarks for evaluating the rationale performance of ChatGPT. The professional profiles are available at Table 46.

Table 46 – Professional profiles. Column *XP* means the years of experience on the professional role. Columns *InfoVis*, *UX*, *Analytics tools*, and *Charts* means the participant interest in the topic ranging from little interest to much interest.

ID	Age	Role	Area	XP	InfoVis	UX	Analytics tools	Charts
P1	35	UX Researcher	Financial	2 to 4 years	Much	Reasonable	Much	Much
P2	37	UX Researcher	Design	+ 4 years	Much	Much	Much	Reasonable
P3	23	Developer	Software	< 1 year	Reasonable	Reasonable	Reasonable	Reasonable
P4	22	Developer	Software	2 to 4 years	Little	Little	Little interest	Little
P5	41	Software Engineer	ERP	+ 4 years	Much	Much	Much	Much
P6	23	Developer	Software	2 to 4 years	Much	Little	Reasonable	Reasonable

Data Collection with Human Participants: conducted individually via Google Meet video calls, scheduled by participants at their preferred dates after receiving the study invitation. During the call, each participant was presented with the Informed Consent Form (ICF), which they read and agreed to before proceeding. After consenting, participants completed an online questionnaire to provide information about their professional background. Next, they received access to an online spreadsheet containing the clickstream dataset, along with a PDF file describing the dataset columns. Participants were given time to explore the dataset and could ask clarifying questions to the researcher. Once their doubts were resolved, they were asked to review 18 predefined UX objectives and indicate which ones could reasonably be addressed using the data provided. Four of the six participants received the UX data taxonomy (MACEDO; ZAINA, 2025) to use as a reference in the task. The two most experienced participants, on the other hand, did not have access to the taxonomy during the study.

Data Collection with ChatGPT: Through successive Chain-of-Thought[3] interactions, we refined a standardized four-step process for engaging with ChatGPT: **Initial greeting and overview** – The session began with a message, introducing the task and outlining the upcoming steps (see Figure 47 (A)). **Persona configuration** – A prompt was provided to instruct the model to adopt a professional profile that reflected the characteristics of one of the human participants (see Figure 47 (B)). The prompts developed

[3] This refers to a prompting technique that encourages the model to generate intermediate reasoning steps before providing a final answer.

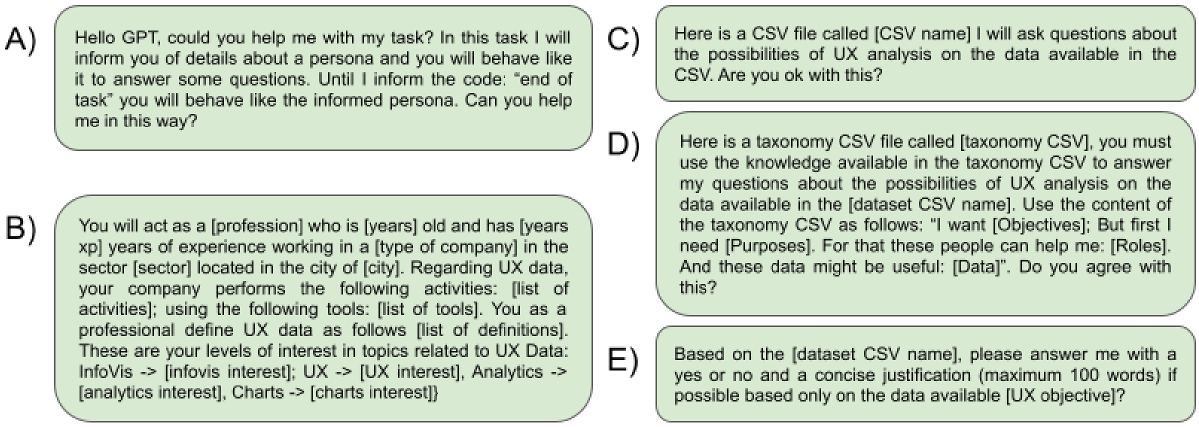


Figure 47 – Prompts used in the study.

using participants profile (see Table 46) are available here. **Dataset input** – The clickstream dataset was uploaded and introduced (see Figure 47 (C)). **Questioning** – The model was questioned about the same 18 UX objectives (see Table 45) that were posed to the human participants, with answers requested in a consistent Yes/No format followed by a brief justification (see Figure 47 (E)). After completing this initial protocol, the same process was repeated with one key variation: in the **Dataset input** step, the UX data taxonomy was also uploaded (see Figure 47 (D)). A template was included to inform ChatGPT of the structure and content of the taxonomy CSV. All ChatGPT responses were recorded in a spreadsheet for comparison with the corresponding human participants. In most cases, the AI simulation was configured inversely (i.e., when the human had access to the taxonomy, the AI did not). To enrich the analysis, AI simulations for Personas 1 (second most experienced) and 3 (least experienced) were run both with and without the taxonomy. In contrast, the AI simulation for Persona 2 (most experienced) was run only without the taxonomy to ensure an equal condition comparison.

10.3 Results

The complete set of responses provided by both human participants and ChatGPT is presented in Figure 48. ChatGPT consistently showed a more "positivist" pattern, selecting more UX objectives than human participants across conditions. However, for Persona 1 without access to the taxonomy, the AI selected fewer. This trend suggests a form of overconfidence, with ChatGPT perceiving broader potential in the data than human experts. ChatGPT responses for Personas 4, 5, and 6 converged toward similar patterns, possibly due to the participants' comparable professional profiles. In contrast, their human counterparts showed more variation despite similar backgrounds. For example, Personas 3 and 5 agreed with about 30% of the UX objectives, while Personas 4 and

Participant	Agent	Taxonomy?	UX1	UX2	UX3	UX4	UX5	UX6	UX7	UX8	UX9	UX10	UX11	UX12	UX13	UX14	UX15	UX16	UX17	UX18
Persona 1	Human	No	x	x	x	x				x			x				x	x	x	
	AI	No	x	x		x			x	x					x		x	x		
	AI	Yes	x	x	x	x		x	x	x					x		x	x	x	
Persona 2	Human	No	x					x	x	x	x				x		x	x	x	
	AI	No	x	x	x	x		x	x	x	x	x		x	x	x	x	x	x	
Persona 3	Human	Yes	x		x	x		x	x											
	AI	Yes	x	x	x	x		x	x	x	x	x			x	x	x	x	x	
	AI	No	x	x	x	x		x	x	x	x	x			x	x	x	x	x	
Persona 4	Human	Yes	x	x	x	x	x		x	x		x	x	x	x			x		
	AI	No	x	x	x	x		x	x	x	x	x			x	x	x	x	x	
Persona 5	Human	Yes	x		x			x			x				x					
	AI	No	x	x	x	x		x	x	x	x	x			x	x	x	x	x	
Persona 6	Human	Yes	x		x				x			x	x	x	x	x		x		
	AI	No	x	x	x	x		x	x	x	x	x			x	x	x	x	x	

Figure 48 – Responses from human participants and ChatGPT instances across 18 UX objectives. "X" indicates affirmative responses.

6 endorsed between 50% and 70%. This suggests that while ChatGPT generalized across similar profiles, humans applied more individualized judgment.

Introducing the UX data taxonomy had opposite effects: it expanded human judgments but constrained ChatGPT responses, reducing affirmative selections from 77% to 59% overall. However, this pattern did not hold for Persona 1, where ChatGPT selected more objectives with the taxonomy (61%) than without (44%). This suggests that while humans use structured knowledge to broaden interpretation, ChatGPT may respond to it either as a constraint or a clarifier, depending on context. Both humans and ChatGPT agreed on the usefulness of clickstream data for understanding users (UX objectives 1 to 3), with ChatGPT being only about 10% more affirmative. However, for product creation objectives (UX objectives 7 to 9), ChatGPT showed around 30% more positive responses, highlighting a tendency to overestimate the data's potential in this area.

10.4 Discussion

Observing the consensus patterns across all participants reveals a distinct distribution of agreement levels, with objectives clustering predominantly at high consensus (i.e., >75%) and low consensus (i.e., <35%) thresholds. This polarization suggests that participants (i.e., both human and GenAI) operate with clear demarcation boundaries when evaluating the validity of e-commerce data applications in UX contexts.

Universal agreement (i.e., 100%) was achieved exclusively for *understand how users interact with systems* (see UX1 in Table 45), representing the direct application of behavioral data in UX research. Near-universal consensus (i.e., 87.5%) emerged for *understanding user preferences* (see UX3 in Table 45), with only one participant dissenting across all conditions. These high-consensus objectives share common characteristics: they primarily involve the direct observation of user behaviors without requiring complex inferential leaps or contextual interpretation beyond immediate data patterns.

Conversely, zero consensus (i.e., 0%) was observed for *understand the user's subjective perception of the system* (see UX18 in Table 45), indicating unanimous recognition that the e-commerce interaction data (i.e., the dataset used in the study) cannot adequately capture internal psychological states or subjective experiences. Similarly, minimal consensus (i.e., 16.7%) was recorded for *implement data-driven design* (see UX15 in Table 45), suggesting widespread skepticism about the sufficiency of clickstream data for comprehensive design decision-making. The moderate consensus range (i.e., 40 to 60%) encompassed objectives requiring intermediate-level inference, such as *generate insights for creating products based on user data* (see UX18 in Table 45) and *create products that meet user expectations* (see UX9 in Table 45). These objectives represent a contested middle ground where participants diverged on whether behavioral data provides a sufficient foundation for more complex UX applications.

This overall polarization pattern remained consistent across human and GenAI participants, suggesting both operate with shared operational boundaries that clearly distinguish between empirically grounded applications and those requiring substantial interpretive inference. However, within these boundaries, our analysis reveals critical differences in rationale, particularly in GenAI's approach to data applicability.

The consistent overconfidence (i.e., positivism) observed in ChatGPT responses, manifesting as slightly higher endorsement rates across most consensus categories, may stem from its tendency to generalize patterns learned during training, whereas human experts rely on contextual filters grounded in domain-specific experience. This difference is starkly highlighted in the moderate consensus range: ChatGPT demonstrated a tendency to infer applicability even in borderline or less contextually appropriate scenarios, particularly in objectives related to product creation, where it was roughly 30% more positive than human participants. In contrast, for high-consensus goals centered on fundamental user understanding, ChatGPT and human responses were more aligned, with only a 10% difference. This pattern suggests ChatGPT may overestimate the utility of behavioral data in areas where contextual nuance is crucial.

Furthermore, the contrasting effects of the taxonomy reinforce this distinction: for humans, the taxonomy served as a scaffold that broadened their perspective, encouraging affirmation within the moderate consensus range; for ChatGPT, it primarily acted as a constraint, reducing the number of affirmative responses overall. An exception was observed for Persona 1, where access to the taxonomy led to more positive responses, possibly due to a narrow definition of what the AI considered UX data. Interestingly, ChatGPT outputs became increasingly similar for Personas 4, 5, and 6, likely due to the homogeneous nature of their professional profiles, a finding that contrasts with the expected variability of individual human judgment.

These findings suggest an opportunity for hybrid workflows: ChatGPT can expand the scope of exploration, surfacing overlooked hypotheses, while human researchers apply critical judgment to filter results, balancing creativity with practical relevance. Nevertheless, the documented differences (i.e., GenAI's tendencies and its distinct interaction with structured knowledge) remain fundamental distinctions from human rationale that must be explicitly acknowledged and managed when integrating ChatGPT into professional UX research practices.

10.5 Conclusion

We compared the judgment of UX experts and ChatGPT regarding the applicability of interaction data for UX objectives. Our findings reveal differences in rationale, with significant implications for AI integration in UX research workflows. ChatGPT *positivism*, consistently identifying more potential applications than humans, indicates overconfidence requiring careful management, particularly in product creation objectives. The taxonomy expanded human interpretation but constrained ChatGPT responses, demonstrating a distinct functioning of conceptual frameworks in human versus artificial cognition. The higher convergence among ChatGPT responses for similar profiles, contrasting with human individuality, highlights distinct generalization mechanisms.

This study highlights the value of integrating GenAI in the field of interactive systems engineering. Understanding how human and AI agents differ in interpreting UX data, particularly when guided by structured knowledge, can inform the design of hybrid workflows aligned with the principles of Human-Centered Artificial Intelligence (HCAI). A complementary model, in which GenAI amplifies rather than replaces humans, offers a promising direction. For system engineers, this underscores the importance of designing processes that balance GenAI's exploratory power with human contextual judgment. Future research should further investigate how different configurations of human-AI collaboration can enhance the analysis and development of user-centered interactive systems.

10.6 Final Considerations

This chapter concluded its investigation into the comparative judgments of UX experts and ChatGPT concerning the applicability of interaction data for UX objectives, revealing key differences in their rationales. The study found that ChatGPT exhibited a consistent "positivism", identifying more potential applications than humans, particularly in product creation objectives, highlighting a need for careful management of its overconfidence. The taxonomy, while expanding human interpretation, constrained ChatGPT's responses, indicating distinct operational mechanisms of conceptual frameworks between human and artificial cognition. The research suggests an opportunity for hybrid workflows where Generative AI (GenAI) can broaden the scope of exploration by surfacing overlooked hypotheses, while human experts apply critical judgment to filter results and ensure practical relevance. This work emphasizes the importance of designing processes that balance GenAI's exploratory power with human contextual judgment, informing the future design of Human-Centered Artificial Intelligence (HCAI) workflows in interactive systems engineering.

Acknowledgments

We thank the partial financial support of the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - Brazil) grant 309497/2022-1.

Chapter 11

UX Data Wiki AI

This chapter presents the development and evaluation of a tool (i.e., UX Data Wiki AI) using GenAI to access th UX data taxonomy content using natural language. Our primary objective is to empower software professionals to explore the defined UX data taxonomy and extract meaningful insights from UX data, even without specialized knowledge in visualization or databases. This tool, developed using a RAG approach and adhering to the Model Context Protocol (MCP), only leverages the proposed UX data taxonomy as its knowledge base to answer user questions through natural language, thereby simplifying complex data exploration. A version of the work described in this chapter is currently under preparation for submission to a HCI conference.

This work directly contributes partially to answering our research questions. It primarily addresses RQ3: "How can software professionals improve their understanding of UX data, and effectively integrate and operationalize this knowledge into their software development practices?" by demonstrating how the UX Data Wiki AI serves as an accessible and effective mechanism for interacting with structured UX knowledge, enabling professionals to identify relevant data that helps in find design problems. Furthermore, it directly responds to RQ4: "What are the emerging opportunities for UX data analysis, particularly concerning the role of GenAI?" by presenting a robust and evaluated practical application of generative AI to enhance UX data accessibility and analysis, showcasing the potential of RAG-based systems in specialized domains.

The remainder of this chapter is structured as follows: Section 11.1 details the system's design and implementation. Section 11.2 presents a pilot evaluation, comparing the tool's performance against a generic conversational AI. Section 11.3 then describes a broader evaluation with software professionals, focusing on the relevance and utility of the insights provided by our system in a data exploration context. Finally, Section 11.5 offers a discussion of our findings.

11.1 System Design and Implementation

This section provides a detailed account of how the tool was developed, outlining its architecture and the specific technical choices made. The system's design is rooted in the principles of the Model Context Protocol (MCP), which emphasizes a standardized approach for language models to effectively understand and interact with external data sources and services. This protocol ensures that the AI component can take meaningful action beyond mere text generation, specifically by using a rich, domain-specific knowledge base. The core of the UX Data Wiki AI is a RAG architecture, designed to provide accurate, contextually grounded answers solely from the established UX data taxonomy. The application was hosted on Hugging Face Spaces, implemented with Python for its backend logic and Gradio[1] for a streamlined frontend interface.

11.1.1 Architectural Overview

The tool is structured as a multi-component system designed to facilitate seamless natural language interaction with the UX data taxonomy. As depicted in Figure 50, the architecture comprises five components that orchestrate the flow from user query to AI-generated response:

1. **Chat Interface:** Developed using Gradio, this component serves as the primary point of interaction for the user. It captures natural language queries and displays the generated responses.

2. **Backend:** Hosted on Hugging Face Spaces[2], this Python-based backend acts as the central orchestrator. It receives user queries from the UI, manages the RAG workflow, and integrates with the external LLM (i.e., UX Data Wiki AI used ChatGPT and Gemini).

3. **Question Embedder:** The ChatGPT API is utilized as the LLM responsible for embedding natural language queries (i.e., usign "text-embedding-3-small"[3]) to submit then to semantic search in the vector database.

[1] <https://www.gradio.app/>

[2] Hugging Face Spaces is a platform within the Hugging Face ecosystem that allows users to host and share machine learning models as interactive web applications. <https://huggingface.co/docs/hub/spaces>

[3] An embedding model transforms discrete data (like words or images) into dense, continuous numerical vectors (embeddings) in a high-dimensional space. In this space, items with similar meanings or characteristics are positioned closer together, enabling computational analysis of relationships and similarities that are not explicitly present in the original data.

4. **Vector Database:** The MongoDB Atlas Vector Search[4] database stores the vectorized representation of the UX data taxonomy, serving as the system's exclusive knowledge base. It is optimized for efficient semantic similarity searches.

5. **Response generator:** Integrated via its API, the Gemini model is utilized as the LLM responsible for generating natural language responses. Its role is defined through a specific prompt context to act as a senior UX teacher to better explain the UX data taxonomy information (see prompt in Figure 49).

> You are a senior UX teacher, answer in detail to clear students' doubts. You are talking with brazilians, answer in Portuguese. Answer the question based only on the following context: [information_retrieved_from_vector_database]. Question: [question_from_user]

Figure 49 – Prompt used in Response generator

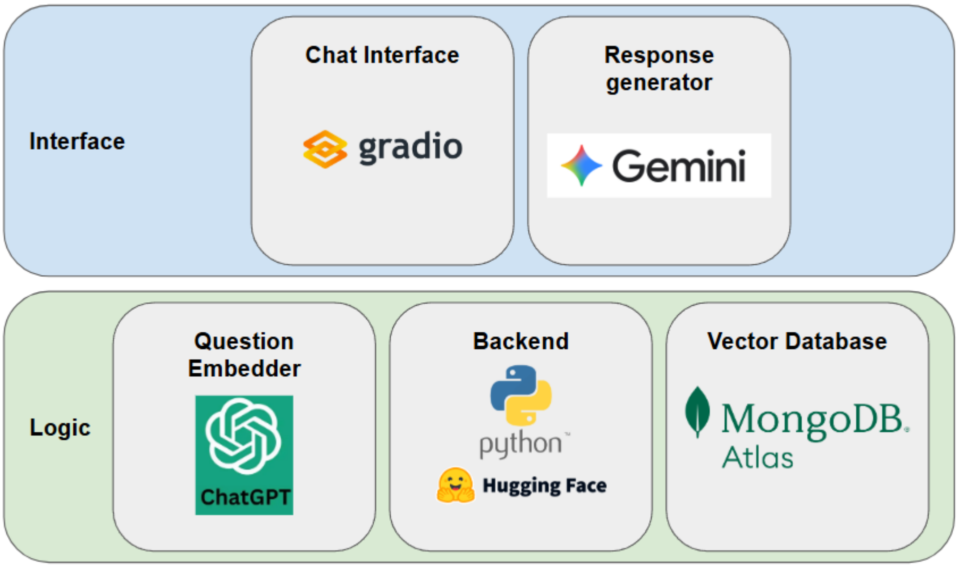


Figure 50 – Technologies used in the operation of UX Data Wiki AI

The interaction flow begins when a user submits a natural language question via the **Chat Interface**. This query is transmitted to the **Backend** and the user's question is vectorized using the **Question Embedder**. This vectorized query is then submitted to the **Vector Database** for a semantic similarity search. The search retrieves the 10 most semantically similar pieces of information from the vectorized UX data taxonomy. These retrieved taxonomic segments, along with the user's original question, are then passed to the **Response generator**.

[4] <https://www.mongodb.com/products/platform/atlas-vector-search>

The **Response generator** is instructed with a specific role, acting as a UX teacher, and is strictly guided to elaborate its response only based on the provided context from the **Vector Database**. This adherence to the MCP ensures that responses are grounded in the taxonomy and relevant to the UX domain. Finally, the generated answer is sent back through the **Chat Interface**. Figure 51 summarizes the inner workings of UX Data Wiki AI during user interaction.

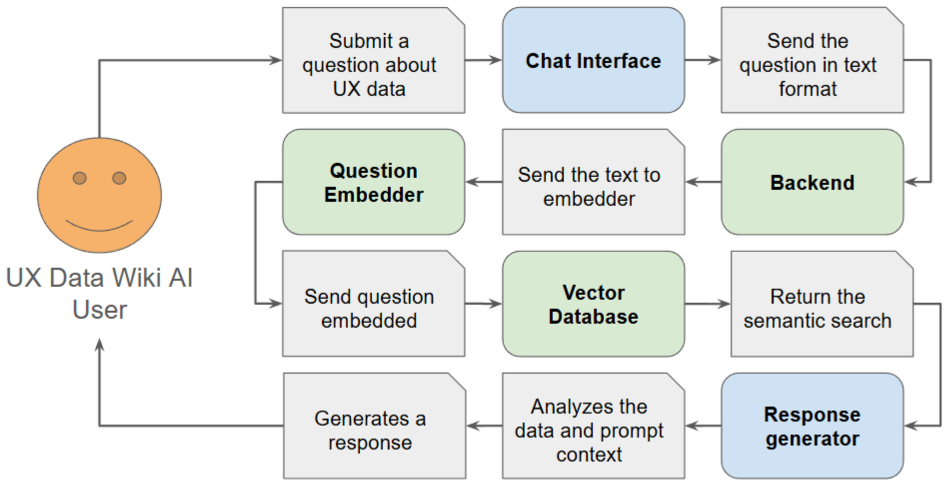


Figure 51 – Interaction flow of a prompt between UX Data Wiki AI technologies

11.1.2 Taxonomy Integration

The foundational element of the UX Data Taxonomy AI is the UX data taxonomy, which serves as the sole knowledge source for the RAG system. To move beyond arbitrary connections and ensure the logical integrity of our taxonomy, this research adopted a approach inspired by the concept of a controlled vocabulary. In essence, a controlled vocabulary is a selected list of terms used to ensure uniformity and clarity in information organization within a specific domain, thereby minimizing ambiguity and ensuring consistent understanding (ZHANG et al., 2015).

Just as a controlled vocabulary brings precision to terminology, the research applied its underlying principles to defining the relationships between taxonomy categories (i.e., Objectives, Purposes, and Data characteristics). Initially the content of the UX data taxonomy was organized using this specific grammar: "*I want [OBJECTIVE], but first I need [PURPOSE] and for that I need the following data [Data characteristics]*".

The process of populating the knowledge base involved two steps, **content vectorization** and **vector database indexing**. These steps were performed only once to populate the database. During **content vectorization** each segment of the organized taxonomy content was converted into a high-dimensional numerical vector (i.e., embedding process, same as **Question Embedder** does). As mentioned in last section, the text-embedding-

3-small model was employed for this purpose, chosen for its efficiency and effectiveness in capturing semantic meaning.

Then, in **vector database indexing**, these generated embeddings and their corresponding original taxonomic content were then ingested and indexed into MongoDB Atlas Vector Search. This specialized vector database enables semantic searches (i.e., query information by similarities) which is crucial for the RAG mechanism. Figure 52 illustrates the process of selecting an excerpt from the UX data taxonomy, submitting it to the embedding process, and indexing it into the vector database.

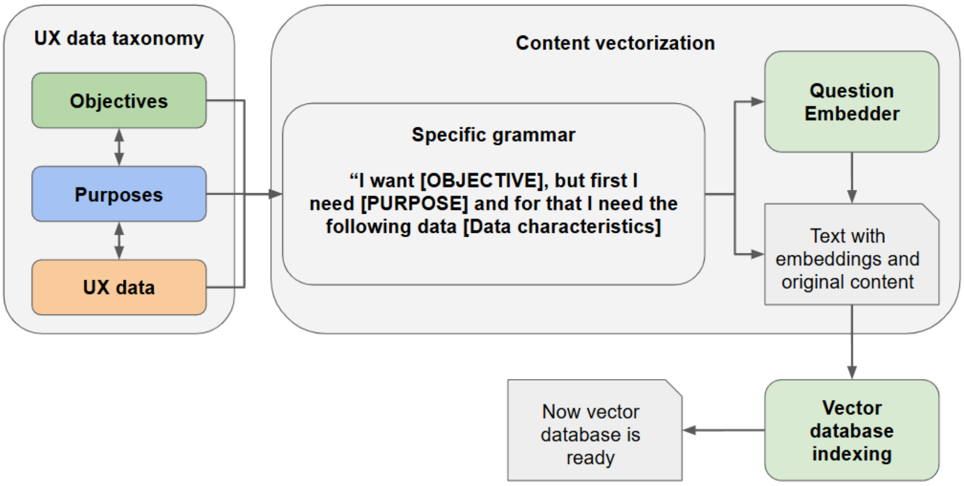


Figure 52 – The process of vectorizing taxonomy information. This process was repeated for each objective.

When a user's query enters the system, it undergoes the same vectorization process made in **content vectorization** step. The resulting query embedding is then used to perform a semantic search within the MongoDB Atlas Vector Search. This retrieval mechanism is configured to identify and return the top 10 most relevant segments of the UX data taxonomy that semantically align with the user's input. This contextually relevant information forms the basis upon which the Gemini agent formulates its grounded answers, adhering strictly to the Model Context Protocol by using the retrieved data as its authoritative source.

11.2 Pilot Evaluation

This section details the pilot evaluation conducted to assess the foundational capabilities of the UX Data Wiki AI against ChatGPT.

11.2.1 Objectives and Participants

The primary objective of this study was to evaluate the effectiveness of the UX Data Wiki AI in answering user questions about UX data, specifically comparing its performance when relying solely on LLM pre-trained knowledge. To achieve this, we compared three distinct tasks: No Prompt, Prompt Engineering, and UX Data Wiki AI.

A total of 10 software professionals with a background in UX participated in this evaluation conducted in a controlled setting. All participants had previously collaborated with the UXLeris research group and were invited to take part in the study due to their familiarity with the research theme and its related activities. However, it is not a controlled evaluations since the sample size does not allow for generalization. Their professional roles included Software Engineers, UX Researchers, and Designers, reflecting a diverse representation of the target user group. Participant ages ranged from 21 to 54 years. Notably, 70% of the sample had more than four years of professional experience, indicating a seasoned group of participants. A significant portion of the participants (six individuals) expressed a strong expertise in the fields of UX, Analytics, and Information Visualization, suggesting a high level of engagement with the study's subject matter. Table 47 presents the profile of each participant.

ID	**Age**	**Role**	**Experience**	**UX**	**Analytics**	**InfoVis**
PP1	28	Product Manager	+ 4 years	Strong expertise	Strong expertise	Strong expertise
PP2	38	Tech Lead	+ 4 years	Strong expertise	Little expertise	Reasonable expertise
PP3	42	Software engineer	+ 4 years	Strong expertise	Reasonable expertise	Reasonable expertise
PP4	32	UX Researcher	+ 2 years	Strong expertise	Strong expertise	Strong expertise
PP5	54	UX Researcher	+ 4 years	Strong expertise	Strong expertise	Strong expertise
PP6	21	Developer	+ 1 year	Reasonable expertise	Reasonable expertise	Strong expertise
PP7	27	Developer	+ 1 year	Little expertise	Strong expertise	Strong expertise
PP8	26	Developer	+ 4 years	Strong expertise	Reasonable expertise	Strong expertise
PP9	36	Developer	+ 4 years	Little expertise	Little expertise	Little expertise
PP10	44	Developer	+ 4 years	Reasonable expertise	Strong expertise	Reasonable expertise

Table 47 – Profile of pilot evaluation participants

11.2.2 Study design

Participants engaged with the LLM in a structured online testing environment, where each condition (i.e., No Prompt, Prompt Engineering, and RAG) was tested with the same set of three predefined tasks (i.e., nine tasks per participant). This setup allowed for an assessment of response quality under varying levels of LLM guidance and knowledge augmentation.

In tasks using **No Prompt** the LLM generated responses without any specific instructional prompting, representing its general knowledge capabilities. In **Prompt Engineering** the LLM was guided by a crafted prompt designed to steer its answers towards relevant information. Finally, in our **RAG** approach the LLM's response generation was augmented by retrieving relevant information exclusively from our UX data taxonomy, embodying the core of our proposed tool.

The quality of each LLM response was evaluated by the participants using a 5-point Likert scale across two dimensions: Relevance and Clarity. **Relevance** evaluated if the response correctly addressed the question's intent (1 - Not relevant, 5 - Very relevant). **Clarity** determined if the response was easy to understand (1 - Confusing, 5 - Very clear).

The three tasks (see Table 48) were designed each one focusing on specific aspects of UX data understanding and application. Table 48 presents the tasks, their objective and the **prompt** that participants should ask to ChatGPT.

Task	Objective	Prompt to use
1	Test the LLM's ability to accurately define a specific UX data term (i.e., bounce rate)	What is bounce rate?
2	Assess the LLM's capacity to differentiate between related UX data concepts	What is the difference between bounce rate and error rate?
3	Explore the LLM's ability to identify appropriate UX data types based on a described scenario	Users are frequently abandoning their shopping carts before completing a purchase. Which UX data should I use to evaluate this situation?

Table 48 – Pilot evaluation tasks and prompts used.

Furthermore, upon completion of all tasks across all conditions, participants were asked a concluding open-ended question: "Would you like to comment on the answers provided by the tools used?". This qualitative data provided insights into user perception and preference regarding the different LLM interaction strategies.

11.2.3 Execution and Analysis

Our study adhered to the ethical guidelines outlined in Regulation Document 510/2016 of the Health National Board in Brazil. Participants received an invitation message stating the researchers' identities, the study's aim, and the data collection methodology, emphasizing that all data would be used exclusively for scientific purposes. Each participant was provided with a Term of Informed Consent (ICF), developed in accordance with Regulation Document 510/2016. Only those who read and provided their consent proceeded to complete the profile questionnaire and participate in the study. We stated that our evaluation focused on the tool's performance, not on individual participant expertise. For transparency, authors' contact details were provided, allowing participants to request their data at any time. Regarding the well-being of participants, we ensured that participation was entirely voluntary, with the right to withdraw at any time without penalty. To address concerns about transparency, auditability, dissemination, and the use of research products, the ICF stipulated that all data used in publications would be anonymized and made publicly available to facilitate replication of analyses. Raw data, accessible only to researchers, would have all identifying information permanently erased upon completion of data collection.

The pilot study was conducted remotely and synchronously, with each participant connecting via Google Meet monitored by the researcher. Each session lasted approximately 30 minutes per participant. The synchronous setup allowed the researcher to be available throughout the session to address any participant queries promptly, ensuring a smooth testing experience.

Participants initiated their involvement by accessing a Google Forms link, which outlined all the tasks and instructions for the study. Additionally, a link to a Google Drive text document was provided, enabling participants to copy and paste the LLM-generated responses for later analysis.

The tasks were presented to participants in a fixed and sequential order: from Task1 to Task3. Tasks were performed in this order to mitigate inherent AI learning (i.e., that cannot be avoided) and to prevent biased responses for tasks not utilizing prompt engineering. For each specific task, participants engaged with the LLM under all three experimental conditions (i.e., No Prompt, Prompt Engineering, and RAG) in a predetermined sequence for every task.

This structured approach ensured that each participant experienced all interaction methods for each type of question. Immediately after completing a given task across all three conditions, participants proceeded to evaluate the relevance and clarity of the responses for each interaction method using the specified Likert scales within the Google Forms. To analyze the pilot evaluation data, we calculated the mean and standard deviation (i.e., Std. Dev.) for the **Relevance** and **Clarity** scores across each task and LLM condition.

11.2.4 Results

The results provide insights into the performance of the three approaches. Our analysis of the Likert scale responses across the three tasks and conditions provides a quantitative overview of the participants' perceptions.

For **defining UX data (Task 1)**, the RAG approach consistently outperformed both No Prompt and Prompt Engineering, achieving the highest mean Relevance (i.e., 4.70, Std. Dev. = 0.48) and Clarity (i.e., 4.90, Std. Dev. = 0.32). This suggests that grounding responses in the UX taxonomy leads to clearer and more pertinent definitions compared to general LLM knowledge or simple prompting.

When **distinguishing between UX data concepts (Task 2)**, the RAG approach again demonstrated superior performance, with the highest mean Relevance (i.e., 4.80, Std. Dev. = 0.42) and Clarity (i.e., 4.90, Std. Dev. = 0.32). The lower standard deviations for RAG also indicate a more consistent quality of responses, reinforcing its effectiveness in providing precise and understandable comparisons.

In **identifying relevant UX data from scenarios (Task 3)**, the RAG approach continued to show strong performance, securing the highest mean Relevance (i.e., 4.60, Std. Dev. = 0.70) and Clarity (i.e., 4.70, Std. Dev. = 0.67). While all approaches performed well in this task, RAG maintained a slight edge, indicating its value in context-specific data recommendations.

Overall, the RAG approach consistently achieved the highest mean scores for both Relevance and Clarity across all three tasks, often with smaller standard deviations, indicating more consistent high ratings from participants. This suggests that grounding the LLM's responses in the predefined UX taxonomy significantly improved the perceived quality of the answers, particularly in terms of their direct applicability and comprehensibility. Participants' comments on Task 1, which focused on defining UX data, provided valuable qualitative insights into their perceptions of each approach.

Several participants highlighted the conciseness and directness of the UX Data Wiki AI (RAG) tool. For instance, PP2 noted, "The second tool [implied RAG based on typical ordering] is more objective and goes straight to the point, without unnecessary content that might not be relevant in a first interaction."Similarly, PP9 observed, "The UXDataWikiAi tool seemed more concise in its response than ChatGPT; it also seemed more specific than ChatGPT, which was more generic. In the second interaction with ChatGPT, it was also very verbose". This feedback corroborates the quantitative findings of high clarity for the RAG approach.

Conversely, some participants found the ChatGPT (No Prompt/Prompt Engineering) responses to be overly verbose or general. PP8 commented on the task 1 no prompt approach, stating, "there is an extrapolation of the question's content; it brings too much content that ends up complicating understanding". This participant also noted that while the prompt-engineered version reduced verbosity, it sometimes lost relevant information, suggesting a trade-off.

The relevance and comprehensiveness of the UX Data Wiki AI's answers were also praised. PP7 found that "The UX Data WIKI AI answer was more complete right from the first interaction," and importantly, it "brought an additional answer" that helped them understand the utility of system logs as UX data, an insight not found in ChatGPT's response. This indicates the RAG tool's ability to surface relevant information from the taxonomy that might be missed by general LLMs.

However, some comments also pointed out areas for potential refinement. PP4, expecting a more elaborate explanation from UX Data Wiki AI, stated, "As I had previous experience with ChatGPT asking the same question, I found the UXDataWikiAI response to be much more concise and basic. I expected a slightly longer explanation". PP5 noted a minor contextual issue, "In UXData, it starts the answer by saying 'In the context you provided...' however, I only inserted the question without describing the context". Finally, PP10 offered a suggestion for improvement applicable to all tools, wishing they "could at least provide a simple statistical formula to explain with an example how this rate works in a certain population" suggesting a desire for more detail.

For Task 2, which focused on distinguishing between UX data concepts, participants' qualitative feedback largely reinforced the patterns observed in Task 1, while also introducing new insights regarding the tools' ability to provide nuanced and domain-specific information.

Participants frequently lauded the UX Data Wiki AI (RAG) for its directness and precision. PP9 highlighted this, stating, “Again, the UXDataWikiAI tool is more focused and less verbose than ChatGTP. It’s also interesting to see that the tool uses analogies in its explanation, which helps with understanding. Furthermore, in the second interaction, a big difference between the answers was observed: while ChatGPT went for a tool-oriented answer for measurement, UXDataWikiAI actually explained how the measurements can be made”. This suggests the RAG tool’s strength in delivering focused answers and its potential to use illustrative examples, enhancing comprehension.

PP7 further emphasized the RAG tool’s advantage: “Again, the UX Data Wiki AI’s answer is much more explanatory and concise. And, because it’s a tool with a limited scope, it ‘hits’ the topic on the first try, whereas ChatGPT didn’t understand, in the first interaction, that error rate was about UX, not data science.” This crucial feedback underscores the RAG tool’s ability to remain within the UX domain, preventing irrelevant or out-of-scope information often provided by generic LLMs.

Several participants found the UX Data Wiki AI’s responses more insightful or practical. PP4 noted, “I found the UXDataWikiAI answer more interesting than ChatGPT’s in these tasks. It returned a longer explanation than in the first task and, even though it was a shorter text than ChatGPT’s, it brought more practical examples to justify the answer”. This indicates an improvement in the RAG tool’s verbosity for this task, coupled with valuable practical examples.

However, the general knowledge of ChatGPT (No Prompt/Prompt Engineering) sometimes offered a broader perspective, albeit at the cost of domain specificity. PP2 found it interesting that “what GPT brought is that the bounce rate can be positive depending on the context, but the second tool didn’t bring that information”. This suggests that while RAG excels in domain-specific clarity, it might, in some instances, miss broader contextual nuances that a general LLM could infer.

Finally, the suggestion from PP10, similar to Task 1, for “formulas explaining their parameters and bringing examples comparing these parameters in a correlated application” reiterates the desire for more practical, quantifiable details in the explanations, especially from professionals with an “exact sciences” background.

For Task 3, which involved identifying relevant UX data from described scenarios, participants’ qualitative feedback consistently highlighted the strengths of the RAG-based UX Data Wiki AI in providing focused, and contextually relevant suggestions, while also shedding light on the varying levels of detail and generalizability across the approaches.

The UX Data Wiki AI (RAG) was frequently praised for its conciseness, directness, and strong focus on UX. PP3 explicitly stated, “The UXDataWikiAI tool’s responses were more assertive, direct, with a focus on UX”. Similarly, PP4 noted, “Again, I found the UXDataWikiAI response to be of significant length, able to cover several items, and more direct than ChatGPT". PP7 further emphasized, "Again, the UX Data WIKI AI was more concise and brought more information, without the need for additional interactions". This feedback aligns with the quantitative results, indicating that the RAG tool effectively delivered relevant information without excessive verbosity for scenario-based queries.

Participants also appreciated the UX Data Wiki AI's practical approach and contextual understanding. PP6 observed that "Both tools responded correctly, but UXDataWikiAI raised several 'questions to itself' that offer a different (and more interesting) way to try to diagnose the possible reason for the high cart abandonment rate compared to the diagnostic tool suggestions offered by ChatGPT". This highlights the RAG tool's ability to offer a more analytical and diagnostic perspective, moving beyond simple suggestions. PP9 also noted that the UX Data Wiki AI was "more concise and embedded within the context of the question, even recommending which professional could help in data analysis", showcasing its deeper contextual understanding. PP5 succinctly summarized this by stating, "It seems that the UXData tool provides more specific information for the UX area while ChatGPT is more generalist".

Conversely, some participants found the ChatGPT (No Prompt/Prompt Engineering) responses to be either too general, technologically biased, or lacking actionable steps. PP8 commented on the task 3 response, "Is the answer easy to understand? It's understandable, but it's not possible to comprehend the necessary actions to carry out the proposed 'solutions"'. This participant also noted that with Prompt Engineering the LLM "brought examples tied to technology, regardless of the context I provided, it implicitly assumed I was from the technology area", highlighting a potential bias. In contrast, they found that UX Data Wiki AI "brought a clearer answer with accessible examples for anyone, with or without technology experience".

A key insight emerged regarding the target audience and level of detail. PP1 observed, "I found the response a bit less structured to understand in groups and themes which data I should use to perform this evaluation", suggesting a desire for more organized output from the RAG tool for this specific task. PP10 provided a comparative perspective, stating, "The responses in the developed tool [UX Data Wiki AI] are more objective and concise, giving me the impression that it's focused on more experienced professionals. On the other hand, ChatGPT provides longer and more explanatory answers, making it easier for those who don't have as much experience". This suggests a potential trade-off between conciseness (RAG) and broad explanatory detail (ChatGPT), which might cater to different user experience levels.

Across all three tasks (i.e., defining UX data (Task 1), distinguishing between UX data concepts (Task 2), and identifying UX data from scenarios (Task 3)) the RAG (Retrieval-Augmented Generation) approach, implemented as the UX Data Wiki AI tool, consistently demonstrated superior performance in both quantitative and qualitative assessments. Quantitatively, the RAG approach achieved the highest mean scores for both

Relevance and Clarity across every task. These consistently higher means, often accompanied by smaller standard deviations, suggest that grounding the LLM's responses in a predefined UX taxonomy leads to more pertinent and understandable answers compared to generic LLM knowledge (No Prompt) or mere prompt engineering.

Qualitatively, participant feedback strongly corroborated these numerical findings. Users frequently praised the UX Data Wiki AI for its conciseness, directness, and strong focus on the UX domain. Comments indicated that the RAG tool provided more specific and actionable insights, avoiding the verbosity or generic nature sometimes observed in the ChatGPT (No Prompt/Prompt Engineering) responses. Participants highlighted the RAG tool's ability to **go straight to the point** and offer **more specific information**, sometimes even providing novel insights directly from the taxonomy that a general LLM missed. The use of analogies and practical examples within the RAG-generated responses was also noted as particularly helpful for understanding complex distinctions.

However, the evaluation also surfaced important nuances. While RAG excelled in domain-specific clarity and precision, some participants occasionally desired more elaborate explanations or broader contextual information that a general LLM might provide. There were also isolated instances where the RAG tool included extraneous introductory phrases or minor irrelevancies, suggesting ongoing opportunities for prompt refinement.

In conclusion, this pilot evaluation supports the efficacy of a RAG-based conversational interface, explicitly grounded in a domain-specific taxonomy, for improving the accessibility and utility of UX data insights. The findings underscore the critical role of a curated knowledge base in enabling AI to deliver highly relevant and concise information tailored to specialized professional needs.

11.3 Evaluation with Software Professional

This section details the second evaluation aimed at evaluating the practical utility and impact of the UX data taxonomy in real-world UX data exploration scenarios with software professionals using our RAG approach.

11.3.1 Objectives and Participants

The primary objective of this evaluation was to evaluate how the content returned by the RAG approach impacts the ability of software professionals to improve user experience design when interacting with UX data. We hypothesized that the use of the taxonomy, accessed via the UX Data Wiki AI tool, would enhance software professionals' capacity for UX improvement work.

A total of 23 software professionals participated in this study. The participant group was diverse, comprising UX Researchers, Software Engineers, Testers, and Software Developers, reflecting various roles involved in the software development lifecycle. Their ages ranged from 23 to 66 years. Notably, 18 of the 23 participants had more than four years of professional experience, and none had less than two years of experience, indicating a seasoned and relevant participant pool. Table 49 presents the profile of each participant.

ID	Age	Role	Experience	UX	Analytics	InfoVis
PS1	28	Developer	+ 2 years	Strong expertise	Reasonable expertise	Reasonable expertise
PS2	26	Developer	+ 2 years	Reasonable expertise	Reasonable expertise	Strong expertise
PS3	48	Tester	+ 4 years	Reasonable expertise	Reasonable expertise	Reasonable expertise
PS4	33	Developer	+ 4 years	Little expertise	Little expertise	Little expertise
PS5	49	Developer	+ 4 years	Reasonable expertise	Strong expertise	Strong expertise
PS6	50	Analyst	+ 4 years	Strong expertise	Strong expertise	Strong expertise
PS7	23	Developer	+ 2 years	Strong expertise	Reasonable expertise	Reasonable expertise
PS8	50	Analyst	+ 4 years	Strong expertise	Strong expertise	Strong expertise
PS9	42	Tester	+ 4 years	Reasonable expertise	Strong expertise	Strong expertise
PS10	40	UX Researcher	+ 4 years	Strong expertise	Strong expertise	Strong expertise
PS11	32	Developer	+ 4 years	Reasonable expertise	Reasonable expertise	Strong expertise
PS12	25	Developer	+ 2 years	Little expertise	Strong expertise	Strong expertise
PS13	29	Developer	+ 2 years	Strong expertise	Strong expertise	Strong expertise
PS14	51	Developer	+ 4 years	Reasonable expertise	Reasonable expertise	Reasonable expertise
PS15	40	Tech lead	+ 4 years	Reasonable expertise	Strong expertise	Strong expertise
PS16	32	Developer	+ 4 years	Little expertise	Reasonable expertise	Strong expertise
PS17	48	Analyst	+ 4 years	Strong expertise	Reasonable expertise	Strong expertise
PS18	66	Developer	+ 4 years	Strong expertise	Strong expertise	Strong expertise
PS19	39	Developer	+ 4 years	Strong expertise	Strong expertise	Strong expertise
PS20	31	Developer	+ 4 years	Reasonable expertise	Reasonable expertise	Reasonable expertise
PS21	63	Developer	+ 4 years	Reasonable expertise	Strong expertise	Strong expertise
PS22	32	Developer	+ 4 years	Strong expertise	Reasonable expertise	Strong expertise
PS23	39	Developer	+ 4 years	Strong expertise	Strong expertise	Strong expertise

Table 49 – Profile of participants in evaluation with software professionals

11.3.2 Study design

This evaluation was conducted as an evaluation with a controlled setting structured as a hybrid activity, combining an introductory lecture with a hands-on practical session. The total duration of the activity was approximately two hours.

The core of the practical activity revolved around the DUX: A dataset of User Interactions and User Emotions[5]. This dataset, originally collected from a dedicated user study involving 50 participants interacting with a test application for travel expense reports, contains rich information on keyboard and mouse interactions, specifically designed to capture predominantly negative emotions through integrated emotional triggers (LEPPICH et al., 2023). Due to its massive size, a reduced version of the DUX dataset was provided to participants in an electronic spreadsheet format, accompanied by a data dictionary. This subset contained 10,000 lines, maintaining the original columns but reducing the total number of entries to be manageable within the study's timeframe. The two-hour session was structured as follows:

- An overview of UX and Information Visualization, and a discussion on the importance of UX data (15 min).

[5] <https://zenodo.org/records/7778612>

- ❏ Explanations about the challenges of interpreting UX data, selecting appropriate visualizations, and an introduction to the role of AI tools in addressing these issues (15 min).
- ❏ Presentation of the UX Data Wiki AI tool (15 min).
- ❏ Slot for questions (15 min).
- ❏ Each participant received the tasks for analyzing the provided DUX dataset using the UX Data Wiki AI tool (60 min).

Participants engaged with three specific tasks (see Table 50), each designed to elicit different types of UX data analysis and insights. These tasks were designed to simulate real-world scenarios, progressing from initial interface diagnosis to strategic data analysis and flow redesign. Each task focused on specific objectives, presenting a unique scenario for participants to analyze using the provided DUX dataset and the AI tool. Table 50 presents the objectives, scenarios, and expected outcomes for each task.

Task	Objective	Scenario	Expected Response
1	Diagnosis without access to the dataset	Based on the interfaces, are there parts of the form that might cause a negative experience during travel expense report completion?	A initial interpretation of the interface based on expert knowledge
2	Planning data collection and analysis	Imagine you can access user interaction and emotion data. Which data would you analyze to better understand the critical points of the form?	Interpretation of the potential of the data, considering different data points
3	Redesign	Based on the interactions, do any show signs of difficulty in use? How would you redesign this interaction?	Identification of a specific interaction (e.g., button click, value selection, text input) exhibiting patterns of confusion, and proposal for its redesign

Table 50 – Task objectives, scenarios, and expected outcomes for evaluation with software professionals.

To assess the impact of the UX Data Wiki AI tool, data were collected through a combination of open-ended task-based responses and a standardized questionnaire. Participants provided written answers to questions from the tasks (i.e., tasks 1 to 3) and the insights gained with and without the AI tool's assistance. These included questions like "Which data did the AI recommend analyzing to better understand the critical points of the form?" and "What insights did the AI suggest according to the available data?" following their interactions with the UX Data Wiki AI.

Complementing these qualitative insights, participants completed a questionnaire based on the TAM3 (VENKATESH; BALA, 2008) (see Table 51). This questionnaire, using a 7-point Likert scale (i.e., 1 - strongly disagree to 7 - strongly agree), captured their perceptions across various constructs including: **Perceived Usefulness (PU)** (e.g., how the system aided in identifying design problems and proposing relevant solutions), **Perceived Ease of Use (PEOU)** (e.g., ease of learning and clarity of interaction), **Result**

Demonstrability (RES) (e.g., evident benefits and clear contribution to data analysis), **Computer Self-Efficacy (CSE)** (confidence in using the system independently), and **Perceived Enjoyment (ENJ)** (how the system made tasks more interesting), along with questions on **Output Quality (OUT)** (relevance, usefulness, and clarity of information provided by the system) (VENKATESH; BALA, 2008).

#	Question	Construct
Q1	Using the system improves the identification of design problems in interfaces	**PU**
Q2	Using the system enhances my redesign proposals	
Q3	My interaction with the system is clear and understandable	**PEOU**
Q4	I find the system to be easy to use	
Q5	The results of using the system are apparent to me	**RES**
Q6	It was easy to see how the system contributed to data analysis.	
Q7	Even without help, I felt confident using the system for tasks	**CSE**
Q8	I find using the system to be enjoyable	**ENJ**
Q9	The system's responses helped to understand the data clearly	**OUT**
Q10	The information provided by the system was relevant and useful	

Table 51 – Questions used from TAM and its associated constructs.

11.3.3 Execution and Analysis

Our study adhered to the ethical guidelines outlined in Regulation Document 510/2016 of the Health National Board in Brazil. Each participant was provided with a Term of Informed Consent (ICF), developed in accordance with Regulation Document 510/2016. Only those who read and provided their consent proceeded to complete the profile questionnaire and participate in the study. We explicitly stated that our evaluation focused on the tool's performance, not on individual participant expertise. Regarding the well-being of participants, we ensured that participation was entirely voluntary, with the explicit right to withdraw at any time without penalty. To address concerns about transparency, auditability, dissemination, and the use of research products, the ICF stipulated that all data used in publications would be anonymized and made publicly available to facilitate replication of analyses. Raw data, accessible only to researchers, would have all identifying information permanently erased upon completion of data collection.

The study was conducted in-person in a meeting room environment at CPQD[6]. All participants utilized their own computers for the activity. Although all participants work for the same company, they are not necessarily part of the same development team. Due to the company's large size, multiple products are developed simultaneously across different teams and projects. The session began with a talk about the problem (i.e., UX data analysis), delivered synchronously to all participants. During the problem presentation, the UX Data Wiki AI tool was formally introduced and explained. While the DUX dataset and its structure were explained, no specific training on how to use the UX Data

6 CPQD is a Brazilian research and development center focused on Information and Communication Technologies (ICTs). <https://www.cpqd.com.br/>

Wiki AI tool was provided beyond the initial introduction, allowing for an assessment of its intuitive usability.

Following the initial talk, participants proceeded to the practical activity, working individually on the tasks (see Table 50). The tasks required participants to interact directly with the UX Data Wiki AI (RAG approach) to interpret the provided DUX dataset. The entire session, including presentations and practical activity, lasted approximately two hours. The researcher were present throughout the session to facilitate the activity and address any technical issues, though their role was primarily supportive and non-interfering with the task execution.

For the TAM3 questionnaire responses, a descriptive statistical approach was employed and the participants' responses were categorized into three acceptance groups: Strong Acceptance (scores 5-7), Neutral Acceptance (score 4), and Weak Acceptance/Rejection (scores 1-3). This allowed for a understanding of the distribution of user perceptions across various constructs like PU, PEOU, and OUT. The qualitative data from the open-ended task responses (i.e., both unassisted and AI-assisted) were subjected to a thematic analysis. Each participant's written feedback for Tasks 1 to 3 was reviewed to identify recurring themes, patterns, and salient observations regarding the tool strengths and weaknesses. The aim was to triangulate findings from both quantitative and qualitative data to provide a understanding of the tool's perceived value and its influence on UX data exploration.

11.3.4 Results

The evaluation with 23 software professionals aimed to assess the perceived impact and acceptance of the content proposed by UX Data Wiki AI tool responses, particularly focusing on how its content aids in UX data exploration and design improvement. The TAM3 questionnaire, utilizing a 7-point Likert scale (where 1 indicates strong disagreement and 7 indicates strong agreement), was used to capture participant perceptions across various constructs. To interpret these perceptions, responses were categorized into three groups: Strong Acceptance (scores of 5, 6, or 7), Neutral Acceptance (a score of 4), and Weak Acceptance/Rejection (scores of 1, 2, or 3).

Perceived Usefulness: Both items related to Perceived Usefulness, "The system helped me better identify design problems in the interfaces" and "The use of the system contributed to proposing more relevant design solutions", showed strong endorsement. For both, 78.3% of participants demonstrated strong acceptance, with only 13.0% indicating weak acceptance. This suggests that a large majority of professionals found the tool effective in aiding their design-related tasks. Figure 53 and Figure 54 presents the number of participants responses on PU questions.

Perceived Ease of Use: Perceptions of ease of use were also high. "Learning to use the system was easy" garnered strong acceptance from 73.9% of participants, while "Du-

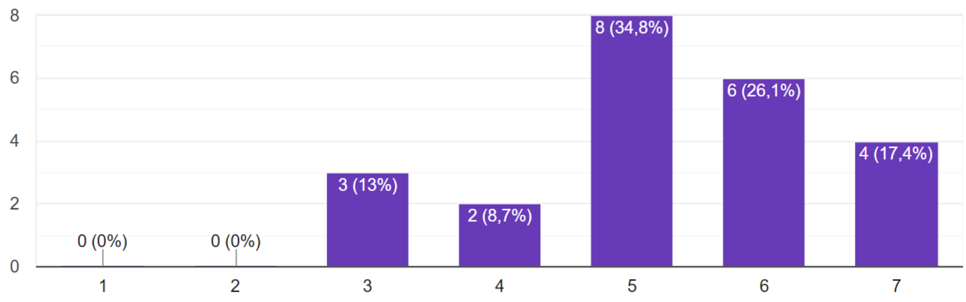


Figure 53 – Number of responses by question Q1 (see Table 51) of the Perceived Usefulness construct.

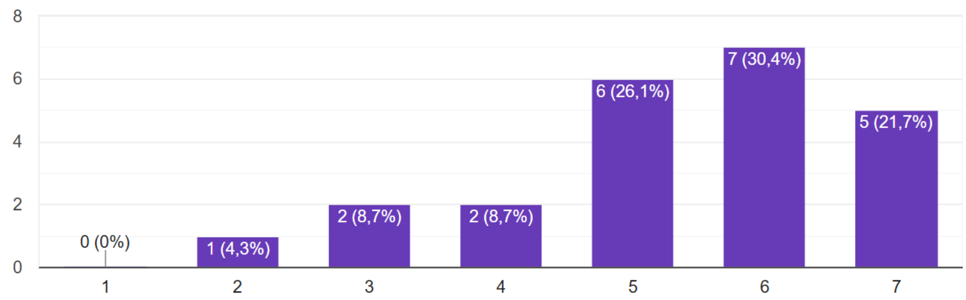


Figure 54 – Number of responses by question Q2 (see Table 51) of the Perceived Usefulness construct.

ring the tasks, interacting with the system was clear and comprehensible" was strongly accepted by 69.6%. The remaining participants were mostly neutral or showed weak acceptance, indicating a generally accessible and understandable user experience. Figure 55 and Figure 56 presents the number of participants responses on PEOU questions.

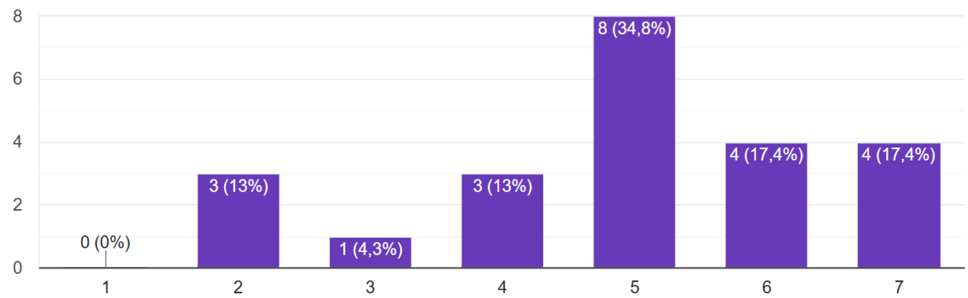


Figure 55 – Number of responses by question Q3 (see Table 51) of the Perceived Ease of Use construct.

Result Demonstrability: For "The benefits of using the system were evident in the tasks", 65.2% of participants showed strong acceptance, with a notable 21.7% in the neutral category. "It was easy to perceive how the system contributed to data analysis" saw strong acceptance from 73.9%. These results suggest that while the benefits were clear for most, a minority remained somewhat undecided about how overtly the system's

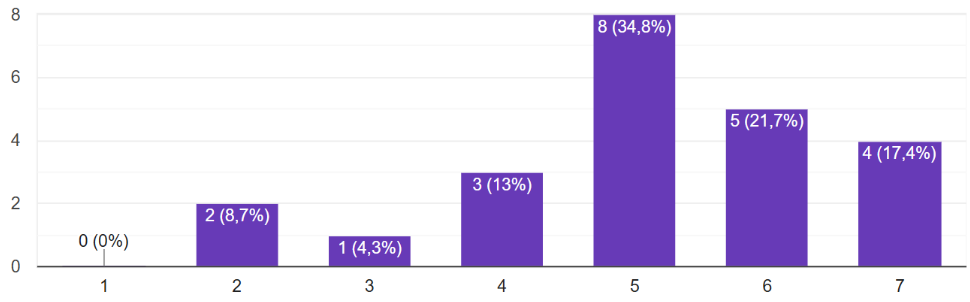


Figure 56 – Number of responses by question Q4 (see Table 51) of the Perceived Ease of Use construct.

contribution manifested. Figure 57 and Figure 58 presents the number of participants responses on RES questions.

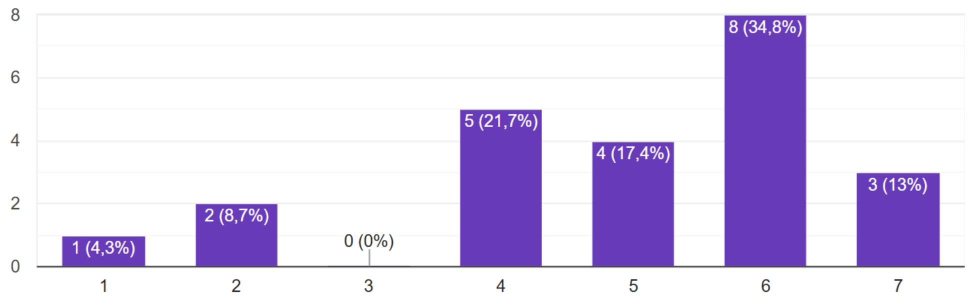


Figure 57 – Number of responses by question Q5 (see Table 51) of the Result Demonstrability construct.

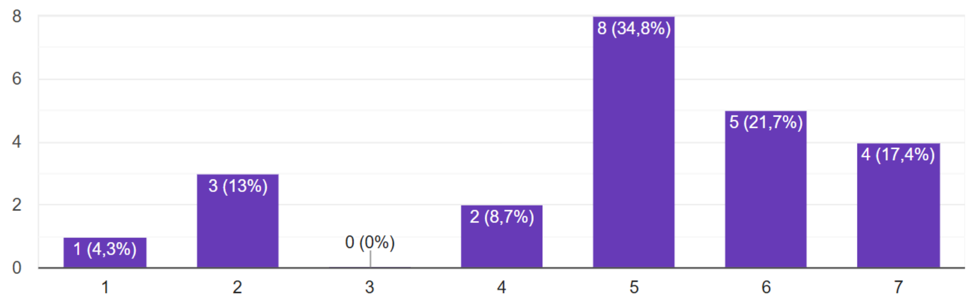


Figure 58 – Number of responses by question Q6 (see Table 51) of the Result Demonstrability construct.

Computer Self-Efficacy: "Even without help, I felt confident using the system for the tasks" had 65.2% strong acceptance, but also the highest proportion of weak acceptance at 21.7%. This indicates that while many felt capable, a significant segment of participants might have felt less confident using the system without external assistance. Figure 59 presents the number of participants responses on CSE questions.

Perceived Enjoyment: "Using the system made the tasks more interesting" received strong acceptance from 73.9% of participants, indicating that the conversational interface

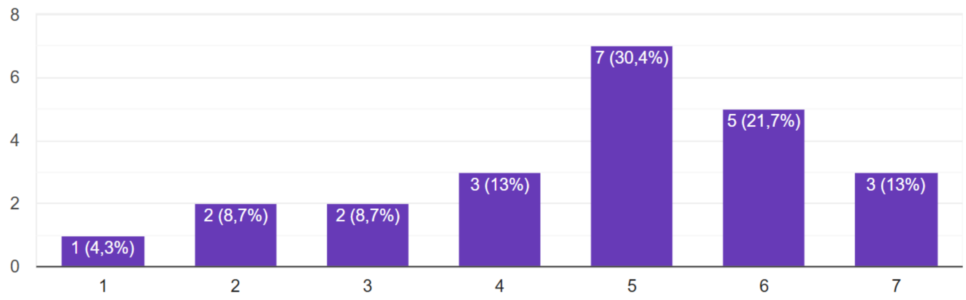


Figure 59 – Number of responses by question Q7 (see Table 51) of the Computer Self-Efficacy construct.

positively engaged users and enhanced their interest in data exploration tasks. Figure 60 presents the number of participants responses on ENJ questions.

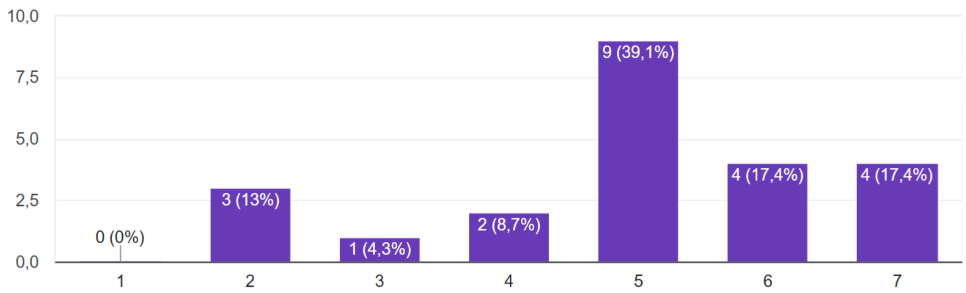


Figure 60 – Number of responses by question Q8 (see Table 51) of the Perceived Enjoyment construct.

Output Quality: Both items measuring Output Quality showed very strong endorsement. “The information provided by the system was relevant and useful” had the highest strong acceptance across all items at 78.3%, with only 4.3% indicating weak acceptance. Similarly, “The system’s answers helped to comprehend the data clearly” was strongly accepted by 73.9%. These findings are particularly crucial, reinforcing the effectiveness of the RAG approach in delivering high-quality, relevant, and comprehensible information from the UX taxonomy. Figure 61 and Figure 62 presents the number of participants responses on PU questions.

In summary, the TAM3 results strongly indicate that the UX Data Wiki AI tool is highly accepted by software professionals, particularly in terms of its **perceived usefulness** and the **quality of its output**. While perceived ease of use and enjoyment were also high, there’s a slight indication that confidence in using the system independently could be an area for further consideration. Overall, the data clearly supports the tool’s potential to be effectively integrated into UX improvement workflows.

For Task 1, where participants were asked to diagnose potential negative experiences solely based on provided interfaces (without LLM assistance), their responses converged on several key themes, revealing common pain points and usability concerns. The feed-

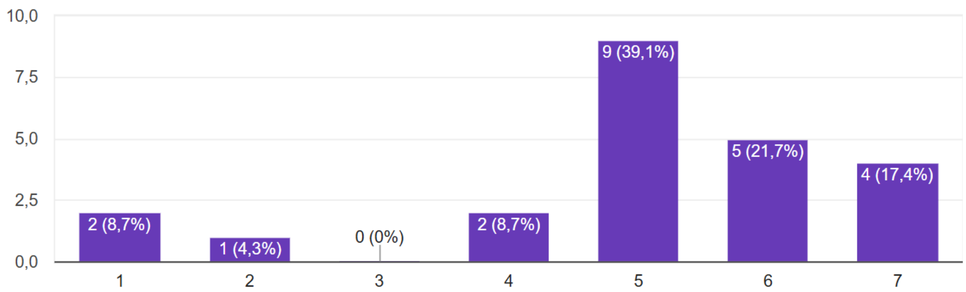


Figure 61 – Number of responses by question Q9 (see Table 51) of the Output Quality construct.

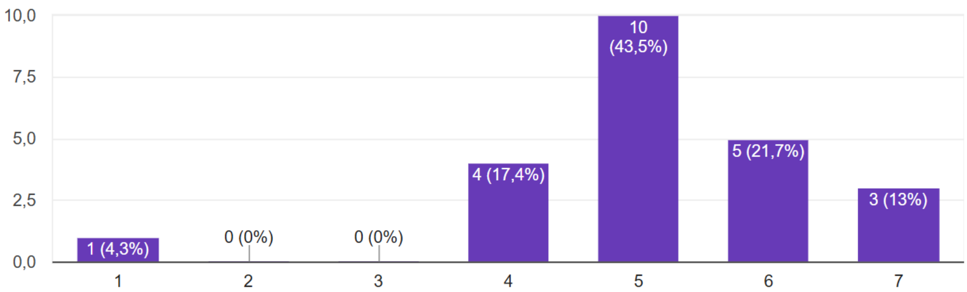


Figure 62 – Number of responses by question Q10 (see Table 51) of the Output Quality construct.

back was largely consistent across participants, highlighting pervasive issues in the form's design.

The most frequently cited problem was poor interface design and visual inconsistencies. Participants identified "misaligned fields" (PS1, PS23), "discrepancies in font sizes" (PS2, PS5, PS12, PS19), "inconsistent text formatting" (PS5), and a general "lack of standardization" (PS5, PS20). The visual appearance was described as "poor" (PS5), "ugly" (PS20), and "not aesthetically pleasing" (PS20).

Another major concern revolved around excessive information and complexity. Many participants noted "too much information on some screens" (PS1, PS12, PS14, PS21), "overly extensive questionnaires" (PS20), and a process that was "not objective" (PS19) or "extremely annoying to fill out, very long" (PS8). This verbosity and length contributed to a feeling that the system was designed to "discourage the user from requesting a refund" (PS20).

A critical and recurring theme was unclear or unhelpful error messages. Participants repeatedly pointed out "error messages that do not say how to solve the problem" (PS2), messages that were "not user-friendly and even ambiguous" (PS3), "unhandled errors" (PS4, PS16), "technical, low-level information (hexadecimal code)" (PS8, PS22), and a general lack of clarity about "what went wrong" (PS5). The frustration of "losing everything at the end due to session drop or critical error" (PS17) due to such messages was also highlighted.

Furthermore, participants identified issues related to usability and user guidance: *lack of input assistance* (PS10), *absence of masking for date fields* (PS5, PS17); *ambiguous or mandatory fields* (PS12); *questions not related to the software's objective* (PS4); and the potential to *lose everything at the end* (PS17). These findings from Task 1 provide a overview of the primary usability challenges observed by professionals when evaluating the interfaces directly.

Task 2 prompted participants to imagine access to user interaction and emotion data, asking them to identify which data they would analyze to understand critical points in the form. Their responses, both unassisted and with the aid of the UX Data Wiki AI, revealed differences in scope, specificity, and depth of proposed data analysis.

Participants, when relying on their own knowledge, generally proposed broad categories of data focusing on common usability metrics. Recurring themes included: **Behavioral data:** "Time spent on each screen", "number of clicks and click locations", "typing speed" (PS2), "completion time" (PS6), and "errors that occur most frequently" (PS7). PS8 mentioned "Time to fill, number of errors encountered"; **User Feedback/Emotion:** "If the user liked the steps followed... and if it met their objective" (PS1), "if it made them irritated", "how annoying they found it" (PS5), "emotions when seeing an unhelpful error message" (PS15), and "user emotions at the critical error at the end" (PS17); **Process Efficiency:** "If they clicked multiple times on a field; if they stopped at a certain point... and stayed longer; if they went back to previous pages" (PS10); **Data Content Focus:** PS3 mentioned "repetitive data entry", while PS18 cited "financial data, reimbursement deadline". While these unassisted responses indicated a good general understanding of relevant UX metrics, they often lacked the structured detail or the breadth of specific data types that a taxonomy might provide.

Responses generated with the assistance of the UX Data Wiki AI (RAG) tool demonstrated a significantly more comprehensive and granular approach to data analysis, often providing a structured list of specific UX data points directly from the taxonomy. This contrast was evident in how the AI systematically broke down the problem into data categories: **Quantitative Behavioral Metrics**: The AI frequently suggested "Abandonment Rate" (PS2, PS4, PS5, PS10, PS12, PS13, PS16, PS20, PS23), "Error Rate" (PS2, PS4, PS5, PS10, PS12, PS13, PS16, PS20), "Time Spent per Field/Interface" (PS2, PS5, PS7, PS10, PS11, PS12, PS16, PS17, PS19, PS20, PS23), and "Click Rate/Interactions" (PS2, PS4, PS5, PS7, PS10, PS12, PS13, PS17, PS20, PS22). The AI's responses consistently provided these specific metrics, underscoring its utility in recalling precise data points.

Beyond quantitative metrics, the AI also systematically listed **Qualitative & Contextual Data**, such as: "User Opinions, Feelings, and Perceptions" (PS2, PS4, PS5, PS8, PS12, PS13, PS16, PS17, PS20, PS22, PS23), "Usability Test Results" (PS2, PS4, PS5, PS8, PS12, PS13, PS16, PS17, PS20, PS22, PS23), "User Behavior Patterns" (PS2, PS4, PS5, PS8, PS10, PS17, PS20, PS23), and “Interfaces Visited by the User” (PS4, PS5, PS8, PS10, PS17, PS20, PS22). This broad coverage indicates the RAG tool’s ability to draw from a rich taxonomy that includes both quantitative and qualitative data types.

The AI-generated responses often provided **why specific data should be analyzed**, connecting the data point directly to a potential problem (e.g., “Fields that take too long can be complex or confusing” (PS2)). PS5’s response from the AI went further, offering “Improvements and recommendations for a more engaging form” based on the data types, including simplification, clarity, real-time validation, and contextual help.

Some AI responses were quite extensive lists of data points (e.g., PS8, PS17, PS20, PS22), reflecting the comprehensiveness of the underlying taxonomy. PS21’s response was an LLM output indicating an irrelevant self-description by the AI, suggesting some instances where the AI’s internal role prompt might have interfered with the direct answer. PS1 also provided an LLM error message, similar to PS21’s response, confirming an interaction issue for that participant. PS15’s AI response was noted as “generic recommendations about how to create ’efficient’ forms”, indicating that while generally useful, some RAG outputs might still lack the precise granularity desired by the user. For Task 2, the UX Data Wiki AI tool significantly enhanced participants’ ability to identify a comprehensive and relevant set of UX data points for strategic analysis, moving beyond general ideas to specific derived from the taxonomy.

Task 3 challenged participants to identify signs of difficulty in interactions based on provided data and propose redesigns, with separate responses for unassisted analysis and AI-assisted insights. The qualitative data revealed varying depths of analysis and the supportive role of the AI.

Participants, relying on their own interpretation of the provided data (likely the spreadsheet) and interfaces, often identified specific areas of friction and proposed general design improvements. Common themes included: **Specific Problem Areas:** “Time field, variable number of fields, error messages” (PS2), “re-entering the same travel expense note, not suggesting a type or pre-registration, and the completion and saving message” (PS3), “absence of an exit button from a modal. Finalize button for the form. No auto-fill or select options” (PS10), and “dropdowns often lead the user to choose an option that does not adequately represent them” (PS22).

As **proposed redesigns** participants suggested “automatic field formatting, prefillable fields, formatting error messages so the user can solve their problems” (PS2), “putting what’s necessary on a single screen, for one interaction and with little information to fill in” (PS5), “simplifying the use and interface, reducing the number of steps and time to fill” (PS9), and “removing checkboxes and adding a scroll bar for easier use, and for costs, a single line for expenses and value, allowing image uploads” (PS21). PS18 suggested “no confirmation on each page, friendly error messages”.

Some participants explicitly **struggled without the AI**. PS8 stated, “I didn’t understand anything looking at the spreadsheets. Nothing is clear, and I'm forced to assume analyses". PS16 noted, "It's very difficult to analyze the spreadsheet without AI or putting it into a [relational] database". PS17 also reported, "I couldn't analyze the data. Difficult". and "Not defined, only guided how to identify". These comments highlight the inherent challenge of raw data interpretation for some professionals, underscoring the need for assistive tools. PS15 mentioned "Impossible to do this" indicating a practical constraint.

The insights provided by the UX Data Wiki AI were generally more structured, comprehensive, and detailed, offering a systematic approach to identifying difficulties and proposing redesigns. The AI's outputs often highlighted key metrics for identifying problems, such as "High Abandonment Rate", "Elevated Error Rate", "Low Click-Through Rate on Essential Elements", "Excessive Time to Complete Tasks", "Negative Opinions, Feelings, and Perceptions", and "Unexpected Behavior Patterns" (PS11, PS21). This systematic breakdown provides a clear framework for diagnosis.

The AI's suggestions for redesign were also very organized, covering broad principles and specific actions. Themes included "Clarity and Simplicity" (PS2), "Minimizing Effort and Errors" (PS2), "Motivation and Engagement" (PS2), and "Prioritization" (PS2). Specific recommendations included "Descriptive Labels", "Real-Time Validation", "Automatic Formatting", "Autocomplete", "Progress Bars", and "Clear Error Messages" (PS2).

The AI consistently stressed the importance of interdisciplinary collaboration, advising users to "Work in collaboration with the designer and product team" (PS5) and highlighting that "Designer" and "Product Team" are key resources (PS10, PS21). The AI often linked observed difficulties directly to the types of data that would reveal them. For example, PS5's AI response detailed how to use "Sequence Analysis", "Time Spent", "Negative Emotional Patterns", and "Entered Text/Selected Value" from the DUX dataset to identify problems, followed by corresponding redesign approaches.

In summary, for Task 3, the UX Data Wiki AI tool significantly enhanced participants' capacity to identify usability issues from complex datasets and conceptualize comprehensive redesign strategies. It provided a structured, and UX-focused analytical framework that many participants struggled to formulate independently from raw data, thus fulfilling its objective of enabling professionals to interact more effectively with UX data for design improvement.

This evaluation aimed to validate the real-world impact and utility of the UX Data Wiki AI tool in addressing the challenges of UX data exploration and design improvement. The qualitative insights gleaned from Task 1, Task 2, and Task 3 consistently highlight the tool's effectiveness in enhancing professionals' analytical capabilities and providing actionable guidance.

In Task 1 (i.e., diagnosis without access to dataset), participants, without AI assistance, largely focused on surface-level usability flaws like visual inconsistencies, excessive information, and unclear error messages. Their observations, while valid, were primarily descriptive. This set the stage by underscoring the inherent difficulties professionals face in translating raw interface observations into systematic UX data analysis.

Task 2 (i.e., diagnosis without access to dataset) revealed a significant impact of the UX Data Wiki AI. When prompted to identify data for analysis, participants' unassisted responses were generally broad, focusing on common metrics. However, with the AI's assistance, participants received significantly more comprehensive, structured, and granular lists of specific UX data points directly from the taxonomy. This demonstrated the tool's ability to not only recall precise behavioral and qualitative metrics (e.g., abandonment rates, error rates, user sentiments) but also to provide the why behind their relevance. The AI moved users from general ideas to specific categories of data for strategic analysis, thereby broadening and deepening their understanding of how to approach UX data.

Finally, in Task 3, the UX Data Wiki AI further proved its value by offering structured, and UX-focused analytical frameworks for identifying usability issues and proposing redesigns. While some participants struggled to interpret the raw DUX dataset independently, the AI-assisted responses consistently provided systematic approaches, linking observed difficulties to data-driven insights and offering redesign strategies. The tool's emphasis on detailed explanations, and even its implicit encouragement for interdisciplinary collaboration, underscored its potential to bridge the gap between data interpretation and practical design improvement. Some feedback also highlighted that the AI's concise and focused output might be particularly valuable for experienced professionals navigating complex data, though a balance with more detailed explanations could cater to broader user expertise.

Collectively, the findings from this second evaluation strongly support the hypothesis that a RAG-based conversational interface, grounded in a UX data taxonomy, significantly enhances software professionals' capacity for UX improvement work. The tool empowers users to move beyond superficial observations, strategically identify relevant UX data, and formulate data-driven design solutions, ultimately making UX data exploration more efficient and impactful for real-world application.

11.4 Study Validity

The study's validity was assessed by considering conclusion, construct, internal, and external threats, drawing upon the recommendations of Wohlin et al. (2012).

To address conclusion threats across both evaluations, the measurement and analysis processes were designed. In the pilot evaluation, where subjective ratings were crucial, the use of Likert scales provided standardized metrics. For the second evaluation, the application of a structured TAM3 questionnaire with a 7-point Likert scale and a subsequent categorization into acceptance groups (i.e., strong, neutral, weak) ensured a uniform interpretation of user perceptions. This minimized bias from differing interpretations of the tools' output or user experience.

To mitigate construct threats, efforts focused on ensuring that measurements reflected their intended constructs. In both studies, participants interacted with the UX Data Wiki AI under defined conditions. The pilot study utilized predefined tasks that targeted the LLM's ability to define, distinguish, and identify UX data, ensuring that the "effectiveness" construct was measured by the quality of the AI's responses.

For the second evaluation, the tasks were grounded in a DUX dataset, and participants were asked to interpret data and suggest redesigns, reflecting the construct of "impact on improving design through UX data interaction". Specific training on the UX Data Wiki AI tool was not provided, allowing for an assessment of its intuitive usability and immediate perceived value from professionals with diverse backgrounds. This approach aimed to mitigate the impact of participants' prior knowledge by focusing on their ability to use the new tool to achieve study objectives.

Internal threats were primarily related to participant engagement and potential biases. Both evaluations involved voluntary participation without financial compensation, suggesting that participants were motivated by genuine interest in the topics of UX, AI, and data analysis. The hybrid format of the second study, which included an talk before the practical activity, aimed to sustain participant interest and clarify foundational concepts. The tasks in both studies were designed to be lightweight, with the practical activity in the second study lasting approximately 20 minutes per task, which minimized the risk of participant fatigue impacting the quality of responses.

Concerning external threats, related to the generalizability of the findings. In the context of this study, the activity of exploring UX data using an AI-powered conversational interface for UX data analysis was novel for all participants, regardless of their professional status. Furthermore, the diverse roles of participants in the second study (i.e., UX Researchers, Software Engineers, Testers, and Software Developers) enhance the generalizability of the findings to a broader spectrum of software development practitioners. The use of a DUX dataset further strengthened the ecological validity of the second evaluation, suggesting that the tool's benefits could translate to similar real-world UX data challenges.

11.5 Discussion

This chapter detailed the development and evaluation of the UX Data Wiki AI, a novel conversational interface designed to empower software professionals in exploring our UX data taxonomy and extracting data-driven insights. The findings from two distinct evaluations collectively underscore the significant value of a Retrieval-Augmented Generation (RAG) approach grounded in a domain-specific taxonomy.

The pilot evaluation served as a crucial initial validation, directly comparing the UX Data Wiki AI (RAG) against general LLM capabilities (No Prompt and Prompt Engineering) in answering UX-related questions. Quantitatively, the RAG approach consistently achieved the highest mean scores for **Relevance** and **Clarity** across all tasks, defining, distinguishing, and identifying UX data from scenarios. Qualitatively, participants frequently praised the UX Data Wiki AI for its conciseness, directness, and domain-specificity. They highlighted how the RAG tool provided more specific and valuable insights, avoiding the verbosity or generic nature often seen in general LLM responses. This initial study firmly established that grounding an LLM with a curated UX data taxonomy significantly enhances the quality and focus of its output for specialized queries.

Building on these foundational findings, the second evaluation with 23 software professionals validated the tool's impact in a more practical, data exploration context. The TAM3 questionnaire results demonstrated high user acceptance, particularly regarding the **Perceived Usefulness** and **Output Quality** of the UX Data Wiki AI. A large majority of participants strongly agreed that the system helped them identify design problems, propose relevant solutions, and that the information provided was useful and clear. This strong acceptance from real users, including UX Researchers and Software Developers, is a testament to the tool's practical value and its ability to integrate into professional workflows.

The qualitative insights from the second study further reinforced these positive perceptions. When tasked with analyzing and redesigning based on a complex UX dataset, participants who used the UX Data Wiki AI moved beyond superficial observations. The AI-assisted responses consistently offered more comprehensive, and structured frameworks, linking observed difficulties directly to specific data types and proposing systematic redesign strategies. Crucially, many participants explicitly articulated the challenge of interpreting raw data without such a tool, highlighting the UX Data Wiki AI's capacity to bridge the gap between raw data and design insights. While the RAG tool's conciseness was highly valued, some feedback also suggested a potential trade-off between brevity and detailed explanations, hinting at different needs for users with varying levels of experience.

Collectively, these evaluations confirm that the UX Data Wiki AI successfully enhances the usability of our UX data taxonomy by making them accessible through natural language, even for professionals without specialized data visualization or database knowledge. The RAG approach, with its exclusive reliance on the taxonomy, proved effective in contributing to addressing the complexities of leveraging UX data for interactive system improvements. This work significantly advances how specialized knowledge can be operationalized and understood by a broader professional audience via AI-powered conversational interfaces, representing a tangible step forward in integrating UX data analysis more deeply into software development practices.

11.6 Final Considerations

This chapter details the development and evaluation of the UX Data Wiki AI, an interactive conversational interface designed to operationalize the UX data taxonomy. Through a two-phase evaluation, including a pilot study and a professional assessment with 23 software professionals, the research demonstrated the effectiveness of a Retrieval-Augmented Generation (RAG) approach grounded in a domain-specific knowledge base. The pilot study quantitatively showed the RAG approach's superior performance in Relevance (e.g., mean 4.70 for defining UX data) and Clarity (e.g., mean 4.90 for distinguishing concepts) compared to general Large Language Models (LLMs), effectively delivering concise and domain-specific insights. The subsequent evaluation with professionals further confirmed high user acceptance across Perceived Usefulness and Output Quality, with 78.3% of participants strongly agreeing on the tool's effectiveness in identifying design problems and proposing relevant solutions. Qualitatively, participants noted the tool's ability to provide structured insights for UX data analysis and redesign, bridging the gap between raw data interpretation and practical design improvement. This work underscores how AI-powered conversational interfaces, when grounded in a curated taxonomy, can significantly enhance the accessibility and utility of specialized knowledge for software professionals, thereby deepening the integration of UX data analysis into development workflows.

Chapter 12

Conclusions

This chapter summarizes the thesis, its key findings, and contributions. The work addresses the unclear definition and inconsistent use of UX data that hinders software professionals from applying user interaction insights to enhance interactive systems. Early research, including initial visualization studies, underscored the challenges of unstructured UX data, leading to a pivotal shift in focus: from developing new visualizations to first establishing a clear understanding of UX data itself. This thesis formalizes a UX data taxonomy to provide structure and clarifies understanding of UX data types, purposes, and objectives. It also presents empirical insights and practical tools, including an interactive web interface and a Generative AI RAG tool, that operationalize this taxonomy. The research evaluates the impact and acceptance of these approaches, including the role of AI, on professional practice in UX data analysis.

12.1 Project Findings and Contributions

This thesis addressed the research problem: "The unclear definition of what constitutes relevant UX data and the complexities of its implementation, which prevents software professionals from adequately applying insights from user interaction data to guide and support the enhancement of interactive systems". To work on this, the primary objective was to develop a structured understanding and provide practical tools to empower software professionals in effectively interpreting and leveraging UX data. To achieve this, a multi-stage research methodology was adopted, combining systematic literature mapping, empirical studies, a grey literature review, artifact development (e.g., the taxonomy), and comparative evaluations with both human participants and AI models.

This thesis provides answers to its guiding research questions, drawing upon the findings and contributions presented throughout the work:

RQ1 - What is the current state of UX data visualization in literature, and what are the practical challenges of utilizing UX data in software development? This question was answered by a systematic mapping of the literature (see Chapter 4). The analysis revealed a predominant focus on evaluating software and game user experiences, highlighting challenges such as sparse experimentation with diverse chart formats, limited exploration of various data sources (e.g., demographic, acoustic, UX metrics), and a notable absence of overview functions in existing visualizations. This established the foundational understanding of the state of the art and identified critical gaps.

RQ2 - What kind of user interaction information can be explored to support software practitioners in discovering opportunities to improve interactive systems? This question was answered in Chapter 5 and Chapter 6. Initial investigations into visualizing Lean Personas data (see Chapter 5) demonstrated the complexities of working with unstructured UX data. Subsequent work exploring visualizations for mobile application problems through grey literature analysis (see Chapter 8) reinforced the multi-purpose nature of visualizations and the critical need for a structured context to correlate data effectively. Collectively, these chapters demonstrated that a wide array of user interaction information, including behavioral data, user actions, and even qualitative emotional responses, can be explored, provided there is a clear conceptual framework.

RQ3 - How can software professionals improve their understanding of UX data, and effectively integrate and operationalize this knowledge into their software development practices? This question was answered through Chapter 7, Chapter 8, and Chapter 9. An investigation into technology acceptance of UX data visualization tools (see Chapter 7) highlighted that beyond tool features, a standardized understanding of UX data itself is paramount for effective adoption. To address this, a UX data taxonomy was proposed (see Chapter 8), providing a structured framework to clarify definitions and connect UX data, objectives, and purposes. This taxonomy was then operationalized as an interactive web system (i.e., UX Data Wiki) (see Chapter 9), which demonstrated its ability to improve the accuracy of identifying achievable UX objectives and significantly reduce the perceived workload and frustration for software professionals.

RQ4 - What are the emerging opportunities for UX data analysis, particularly concerning the role of GenAI? This question was answered through Chapter 10 and Chapter 11. A comparative study between human experts and Generative AI (GenAI) (see Chapter 10) revealed distinct rationales, with GenAI exhibiting "positivism" and the taxonomy having a different impact on human versus AI cognition. This suggested promising avenues for hybrid human-AI workflows. Further, the evaluation of a Retrieval-Augmented Generation (RAG) based AI tool, UX Data Wiki AI (see Chapter 11), demonstrated its superior performance in delivering relevant and clear UX-specific insights when grounded in the taxonomy, highlighting its potential to make specialized knowledge more accessible and actionable for professionals.

12.2 Comparison with Literature

This thesis advances existing literature by providing theoretical and practical artifacts that address gaps in applying UX data in software development. While previous studies have focused on making UX data accessible through visualizations from usability tests (FRANCO et al., 2019b; MÓRO; DARÁZ; BIELIKOVÁ, 2014; DITTRICH; HOF; WIETHOFF, 2019a; BERNHAUPT et al., 2020b) or user navigation (BUONO et al., 2020a; RODDEN, 2014; BAKIU; GUZMAN, 2017; YANG; CHOU; CHEN, 2020; KARAPANOS; MARTENS; HASSENZAHL, 2012a), this work goes further by offering solutions for conceptual clarity and practical operationalization of diverse UX data.

Initial systematic mapping (see Chapter 4) identified key underserved areas in UX data visualization research, such as limited data sources and missing overview functions, establishing the empirical foundation for the structured solutions later developed. This contrasts with existing literature, which often highlights the potential of visual interfaces (e.g., UXmood dashboard for emotional states (FRANCO et al., 2019b) or visualizations for website navigation (BUONO et al., 2020a)) but does not consistently pinpoint specific practical deficiencies.

Further investigation into Lean Personas (see Chapter 5) revealed the inherent complexities of unstructured UX data, highlighting the critical need for a guiding conceptual framework to select and interpret visualizations. This problem was not fully addressed by existing literature on direct visualization techniques or general usability evaluation visualizations (DITTRICH; HOF; WIETHOFF, 2019a; BERNHAUPT et al., 2020b), which often focused on the directness of representation rather than the deeper challenge of data correlation. Empirical validation through grey literature review (see Chapter 8) underscored the necessity of a structured context to effectively correlate UX data, emphasizing conceptual clarity over mere technical presentation in visualization literature.

The UX data taxonomy was introduced as a central contribution. Unlike previous frameworks or hierarchical categorizations (e.g., EUREKATAX), this taxonomy offers a practical, sequential, and relationship-based guide that explicitly connects UX data, objectives, and purposes. It integrates both qualitative and quantitative data, providing a comprehensive approach absent in prior work, such as specialized tools focused solely on organizing retrospective user study data (KARAPANOS; MARTENS; HASSENZAHL, 2012a).

Operationalizing this taxonomy, the UX Data Wiki interactive web system was developed. This directly bridges the gap between theoretical artifacts and practical utility by empirically demonstrating improved accuracy and significantly reduced cognitive load for professionals interpreting UX data. This offers a concrete solution for knowledge application, a missing link in proposals of standalone taxonomies or theoretical frameworks.

Furthermore, this thesis contributed to Human-Centered AI (HCAI) by comparing human and Generative AI (GenAI) rationale. It revealed how structured knowledge (i.e., the taxonomy) distinctively impacts human versus AI cognition, providing specific insights for designing effective hybrid human-AI workflows in interactive systems engineering. Finally, the robust empirical validation of the UX Data Wiki AI (a RAG-based tool) consistently showed superior relevance and clarity in UX-specific queries compared to general LLMs. This offers a concrete solution to the challenge of applying generic AI models effectively in specialized HCI contexts, a limitation often faced when applying general AI models to specific HCI contexts.

12.3 Project Limitations

While this thesis offers contributions, it is important to acknowledge limitations that delimit the scope of its findings and propose directions for future work.

First, the scope of data sources for the systematic mapping (see Chapter 4) and grey literature review (see Chapter 8) primarily focused on published academic literature and specific industry-recognized grey literature sources. This approach, while systematic, might not have captured all nuanced industry practices or emerging trends discussed in less formal, proprietary, or non-English channels.

Second, the generalizability of empirical studies (see Chapter 5, 6, 7, 9, and 10) is bound by the characteristics of their participant pools. While diverse, these groups of software professionals may not fully represent the entire spectrum of practitioners across all industries, company sizes, or geographical locations. Future work could aim for broader participant recruitment to enhance generalizability.

Third, the exhaustiveness of the proposed taxonomy (see Chapter 8), while comprehensive based on the analyzed grey literature, is not absolute. The field of UX and user interaction data is dynamic, constantly evolving with new technologies (e.g., more complex sensor data, AI-driven insights, multimodal data). Consequently, the taxonomy may require future extensions to encompass an even wider range of UX-related scenarios or data types.

Fourth, the technology-specific nature of some evaluations (see Chapter 7, 9, 10, and 11) means the findings are partially dependent on the capabilities and versions of the specific tools (e.g., Looker, Amplitude, ChatGPT) available at the time of the study. These tools undergo rapid changes, which could alter outcomes in future assessments. The UX Data Wiki AI, while effective, is a research prototype and would necessitate further development and validation for large-scale production use.

Fifth, while measures were taken to mitigate it, the inherent researcher bias in the initial systematic mapping's selection and coding (see Chapter 4) and the interpretation of grey literature (see Chapter 8) should be acknowledged. Although supervised and made publicly available, a multi-researcher approach for these phases could further reduce potential subjectivity in future similar studies.

Finally, the generalization of the findings presented herein faces inherent challenges, primarily due to the high rate of change in the topics investigated. Although this work focuses on defining UX data, the very nature of such data was in constant evolution and, notably, expansion throughout the course of this research, driven by new data collection and analysis technologies. An analogous challenge was encountered in the application of Artificial Intelligence, where the continuous release of new models and the availability of new functionalities created a dynamic research landscape. Consequently, the reproducibility of the results must be approached with caution, taking into account the specific models and features available at the time the studies were conducted. Of particular importance are the prompts developed for the tasks, as a modification to a prompt can significantly alter the outcome, even when all other variables are held constant. To ensure transparency, all data and resources utilized are listed throughout the text, and owing to space limitations, several external links have been included. Special attention should be paid to the text highlighted in blue, which indicates a link to external material.

12.4 Future Work

Future research should advance both theoretical understanding and practical application in UX data utilization. A primary direction involves refining the UX data taxonomy by integrating not only feedback from practitioners and emerging UX data types, but also by exploring how the qualitative analysis of user interaction can be systematically incorporated. Furthermore, future work should analyze the taxonomy's connection with different temporal dimensions of user experience (e.g., anticipated, cumulative, and longitudinal UX) to broaden its theoretical scope and applicability. This effort aims to enhance its ecological validity and adaptability to evolving user interaction landscapes.

From an engineering perspective, a next step is to develop more integrated tools that operationalize the taxonomy's conceptual models. This includes researching automated data mapping and leveraging Generative AI to produce UX data visualizations. Such systems could utilize the knowledge embedded within the taxonomy to intelligently assess the analytical possibilities of a given dataset and autonomously generate appropriate and insightful visual representations. Such developments would explore the efficacy of intelligent systems in automating complex analytical tasks, moving towards dynamic and context-aware artifacts.

Furthermore, conducting extensive empirical evaluations of the taxonomy and its tools across diverse organizational and industry contexts is essential. This will strengthen the generalizability and external validity of the thesis's contributions. Specifically, future studies should not only apply the taxonomy to identify further weaknesses in the UX data analysis process but also establish the role of the people who will use the taxonomy. Evaluating its utility within varying professional experience levels (e.g., junior designers vs. senior researchers) and diverse organizational structures would shed light on the specific contexts where the taxonomy provides the most significant benefit.

Finally, this research must investigate how to balance automated insights with indispensable human oversight, ensuring interpretability and accountability. The central challenge is to design AI systems that augment human capabilities in nuanced UX research without diminishing the value of practitioner intuition and critical judgment. This direction contributes directly to the principles of Human-Centered AI (HCAI) by focusing on creating synergistic partnerships between humans and intelligent systems for more effective and responsible UX data analysis.

12.5 Final Considerations

This thesis proposes a conceptual and methodological framework to bridge the gap between raw user interaction data and its practical use in identifying opportunities to improve interactive systems. By systematically defining UX data, categorizing its multifaceted nature, and operationalizing this knowledge through intelligent systems, this work empowers software professionals to move beyond ambiguous definitions and toward data-driven design decisions. The contributions herein underscore the enduring importance of structured knowledge bases, even in the era of advanced AI, demonstrating how taxonomies can enhance the relevance and clarity of AI-generated insights. Ultimately, this research provides a clear pathway for integrating UX data analysis more deeply and effectively into the software development lifecycle, fostering a more user-centered approach to system enhancement.

References

ADAMS, R. J.; SMART, P.; HUFF, A. S. Shades of grey: Guidelines for working with the grey literature in systematic reviews for management and organizational studies. **International Journal of Management Reviews**, v. 19, n. 4, p. 432–454, 2017.

ALBERT, B.; TULLIS, T. **Measuring the User Experience: Collecting, Analyzing, and Presenting UX Metrics**. [S.l.]: Elsevier Science, 2022. (Interactive Technologies). ISBN 9780128180808.

ALBERT, W.; TULLIS, T. **Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics**. [S.l.]: Elsevier Science, 2013. (Interactive Technologies). ISBN 9780124157927.

AMARAL, L. K. d.; MACEDO, M.; ZAINA, L. What is the state of the art on ux data visualization? a systematic mapping of the literature. **Journal on Interactive Systems**, v. 16, n. 1, p. 267–287, Apr. 2025. Disponível em: <https://journals-sol.sbc.org.br/index.php/jis/article/view/4487>.

ANDERSEN, E. et al. Gameplay Analysis through State Projection. In: **Proceedings of the Fifth International Conference on the Foundations of Digital Games**. New York, NY, USA: Association for Computing Machinery, 2010. (FDG ’10), p. 1–8. ISBN 9781605589374.

AWS, A. **What is RAG (retrieval augmented generation)?** 2025. Disponível em: <https://aws.amazon.com/pt/what-is/retrieval-augmented-generation/>.

BAKIU, E.; GUZMAN, E. Which feature is unusable? detecting usability and user experience issues from user reviews. In: **2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)**. [S.l.: s.n.], 2017. p. 182–187.

BASKERVILLE, R. What design science is not. **European Journal of Information Systems**, Taylor & Francis, v. 17, n. 5, p. 441–443, 2008.

BATCH, A. et al. uxsense: Supporting user experience analysis with visualization and computer vision. **IEEE Transactions on Visualization and Computer Graphics**, p. 1–15, 2023.

BELCIC, I. **What is RAG (retrieval augmented generation)?** 2025. Disponível em: <https://www.ibm.com/think/topics/retrieval-augmented-generation>.

BERNHAUPT, R. et al. A Generic Visualization Approach Supporting Task-Based Evaluation of Usability and User Experience. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 12481 LNCS, p. 24–44, 2020.

_____. A generic visualization approach supporting task-based evaluation of usability and user experience. In: SPRINGER. **Human-Centered Software Engineering: 8th IFIP WG 13.2 International Working Conference, HCSE 2020, Eindhoven, The Netherlands, November 30–December 2, 2020, Proceedings 8**. [S.l.], 2020. p. 24–44.

_____. Enriching task models with usability and user experience evaluation data. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 11262 LNCS, p. 146–163, 2019.

BILLESTRUP, J. et al. Persona usage in software development: Advantages and obstacles. In: **ACHI 2014**. [S.l.: s.n.], 2014.

BLASCHECK, T. et al. Visualization of eye tracking data: A taxonomy and survey. **Computer Graphics Forum**, v. 36, n. 8, p. 260–284, 2017.

BOGAARD, T. et al. Understanding User Behavior in Digital Libraries Using the MAGUS Session Visualization Tool. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 12246 LNCS, p. 171–184, 2020.

BROUGHTON, V. **Essential Classification**. [S.l.]: Facet Publishing, 2019.

BROWN, T. et al. Language models are few-shot learners. **Advances in neural information processing systems**, v. 33, p. 1877–1901, 2020.

BUCHINGER, D.; CAVALCANTI, G.; HOUNSELL, M. Mecanismos de busca acadêmica: uma análise quantitativa. **Revista Brasileira de Computação Aplicada**, v. 6, n. 1, p. 108–120, abr. 2014.

BUONO, P. et al. Towards the detection of ux smells: The support of visualizations. **IEEE Access**, v. 8, p. 6901–6914, 2020.

_____. Towards the Detection of UX Smells: The Support of Visualizations. **IEEE Access**, v. 8, p. 6901–6914, 2020.

BÜSCHEL, W.; LEHMANN, A.; DACHSELT, R. MIRIA: A Mixed Reality Toolkit for the In-Situ Visualization and Analysis of Spatio-Temporal Interaction Data. In: **Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2021. (CHI '21), p. 1–15. ISBN 9781450380966.

Campos Filho, A. S.; NOVAES, M. A.; GOMES, A. S. A 3D visualization framework to social network monitoring and analysis. **Computers in Human Behavior**, Elsevier Ltd, v. 49, p. 623–634, ago. 2015. ISSN 07475632.

CARD, S.; JACKO, J. A. **Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, Third Edition**. 3rd. ed. Boca Raton, FL, USA: CRC Press, Inc., 2012. ISBN 1439829438, 9781439829431.

CARD, S.; MACKINLAY, J.; SHNEIDERMAN, B. **Readings in Information Visualization: Using Vision To Think**. [S.l.]: Academic Press, 1999. ISBN 978-1-55860-533-6.

CARRASCO, M.; KOH, E.; MALIK, S. PopHistory: Animated Visualization of Personal Web Browsing History. In: **Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2017. (CHI EA '17), p. 2429–2436. ISBN 9781450346566.

CERNEA, D. et al. An interactive visualization for tabbed browsing behavior analysis. **Communications in Computer and Information Science**, v. 458, p. 69–84, 2014.

______. Emotion-prints: Interaction-driven emotion visualization on multi-touch interfaces. In: **Proceedings of SPIE - The International Society for Optical Engineering**. [S.l.: s.n.], 2015. v. 9397.

CHANG, Y. et al. A survey on evaluation of large language models. **ACM transactions on intelligent systems and technology**, ACM New York, NY, v. 15, n. 3, p. 1–45, 2024.

CHEN, A. T. et al. Mapping User Trajectories to Examine Behavior and Outcomes in Digital Health Intervention Data. In: **2019 IEEE Workshop on Visual Analytics in Healthcare (VAHC)**. [S.l.: s.n.], 2019. p. 1–8.

CHEN, C. et al. Is your prompt detailed enough? exploring the effects of prompt coaching on users' perceptions, engagement, and trust in text-to-image generative ai tools. In: **Proceedings of the Second International Symposium on Trustworthy Autonomous Systems**. [S.l.: s.n.], 2024. p. 1–12.

CHENG, L.; LI, X.; BING, L. Is GPT-4 a good data analyst? In: **The 2023 Conference on Empirical Methods in Natural Language Processing**. [s.n.], 2023. Disponível em: <https://doi.org/10.48550/arXiv.2305.15038>.

CHOMA, J. et al. Influences of UX factors in the agile UX context of software startups. **Inf. Softw. Technol.**, v. 152, p. 107041, 2022.

CHRISTENSEN, S. Y. et al. Define ux design. In: **2020 Intermountain Engineering, Technology and Computing (IETC)**. [S.l.: s.n.], 2020. p. 1–5.

COCKBURN, A.; KARLSON, A.; BEDERSON, B. B. A Review of Overview+detail, Zooming, and Focus+context Interfaces. **ACM Comput. Surv.**, Association for Computing Machinery, New York, NY, USA, v. 41, n. 1, jan. 2009. ISSN 0360-0300.

CONDORI-FERNáNDEZ, N. Happyness: an emotion-aware qos assurance framework for enhancing user experience. p. 235–237, 2017.

CONVERTINO, G.; FRISHBERG, N. Why agile teams fail without ux research. **Commun. ACM**, Association for Computing Machinery, New York, NY, USA, v. 60, n. 9, p. 35–37, 2017. ISSN 0001-0782. Disponível em: <https://doi.org/10.1145/3126156>.

COSTA, I. et al. EmojiText: An Information Visualization Technique for Analyzing Phrases and Sentiments. In: **2021 25th International Conference Information Visualisation (IV)**. [S.l.: s.n.], 2021. p. 114–119.

CRUZ, M. et al. Uso do TAM — Technology Acceptance Model — no Ciclo de Design de Aplicações Computacionais. In: SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. **Anais Estendidos do XXI Simpósio Brasileiro de Fatores Humanos em Sistemas Computacionais (IHC 2022)**. [S.l.], 2022. p. 3–4.

CRUZES, D. S.; DYBA, T. Recommended steps for thematic synthesis in software engineering. In: **2011 International Symposium on Empirical Software Engineering and Measurement**. [S.l.: s.n.], 2011. p. 275–284.

CRUZES, D. S. et al. Case studies synthesis: a thematic, cross-case, and narrative synthesis worked example. **Empirical Softw. Engg.**, Kluwer Academic Publishers, USA, v. 20, n. 6, p. 1634–1665, dez. 2015. ISSN 1382-3256. Disponível em: <https://doi.org/10.1007/s10664-014-9326-8>.

Da Silva Franco, R. Y. et al. UXmood - A Tool to Investigate the User Experience (UX) Based on Multimodal Sentiment Analysis and Information Visualization (InfoVis). In: **2019 23rd International Conference Information Visualisation (IV)**. [S.l.: s.n.], 2019. p. 175–180.

DAVILA, F.; PAZ, F.; MOQUILLAZA, A. Usage and application of heatmap visualizations on usability user testing: A systematic literature review. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 14032 LNCS, p. 3–17, 2023.

DESAI, R. **4 Ways to Implement Data For A Better User Experience Design**. 2019. Disponível em: <https://www.freecodecamp.org/news/4-ways-to-implement-data-for-a-better-user-experience-design/>.

DIAS, R.; FONSECA, M. J.; GONÇALVES, D. Interactive Exploration of Music Listening Histories. In: **Proceedings of the International Working Conference on Advanced Visual Interfaces**. New York, NY, USA: Association for Computing Machinery, 2012. (AVI '12), p. 415–422. ISBN 9781450312875.

DITTRICH, S.; HOF, F.; WIETHOFF, A. Interacdiff: Visualizing and interacting with ux-data. In: **Proceedings of Mensch Und Computer 2019**. New York, NY, USA: Association for Computing Machinery, 2019. (MuC'19), p. 583–587. ISBN 9781450371988. Disponível em: <https://doi.org/10.1145/3340764.3344463>.

_____. InteracDiff: Visualizing and Interacting with UX-Data. In: **Proceedings of Mensch Und Computer 2019**. New York, NY, USA: Association for Computing Machinery, 2019. (MuC'19), p. 583–587. ISBN 9781450371988.

DONG, M. M.; STRATOPOULOS, T. C.; WANG, V. X. A scoping review of chatgpt research in accounting and finance. **International Journal of Accounting Information Systems**, Elsevier, v. 55, p. 100715, 2024.

DRACHEN, A.; CANOSSA, A. Analyzing Spatial User Behavior in Computer Games Using Geographic Information Systems. In: **Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era**. New York, NY, USA: Association for Computing Machinery, 2009. (MindTrek '09), p. 182–189. ISBN 9781605586335.

______. Towards Gameplay Analysis via Gameplay Metrics. In: **Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era**. New York, NY, USA: Association for Computing Machinery, 2009. (MindTrek '09), p. 202–209. ISBN 9781605586335.

DU, W. et al. Personal Web Library: organizing and visualizing Web browsing history. **International Journal of Web Information Systems**, v. 14, n. 2, p. 212–232, 2018.

EBEL, P.; LINGENFELDER, C.; VOGELSANG, A. Visualizing event sequence data for user behavior evaluation of in-vehicle information systems. 2021.

______. Visualizing Event Sequence Data for User Behavior Evaluation of In-Vehicle Information Systems . In: **13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications**. New York, NY, USA: Association for Computing Machinery, 2021. (AutomotiveUI '21), p. 219–229. ISBN 9781450380638.

FABIJAN, A.; OLSSON, H. H.; BOSCH, J. The lack of sharing of customer data in large software organizations: Challenges and implications. In: SHARP, H.; HALL, T. (Ed.). **Agile Processes, in Software Engineering, and Extreme Programming**. [S.l.]: Springer International Publishing, 2016. p. 39–52. ISBN 978-3-319-33515-5.

FAILY, S.; LYLE, J. Guidelines for integrating personas into software engineering tools. In: **Proceedings of the 5th ACM SIGCHI Symposium on Engineering Interactive Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2013. (EICS '13), p. 69–74. ISBN 9781450321389. Disponível em: <https://doi.org/10.1145/2494603.2480318>.

FEITOSA, V. R. M. et al. GameVis: Game Data Visualization for the Web. In: **2015 14th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)**. [S.l.: s.n.], 2015. p. 70–79.

FERREIRA, J. J. et al. Abstraction Levels as Support for UX Design of User's Interaction Logs . In: ANTONA, M.; STEPHANIDIS, C. (Ed.). **Universal Access in Human–Computer Interaction. Design and Development Approaches and Methods**. Cham: Springer International Publishing, 2017. p. 369–382. ISBN 978-3-319-58706-6.

FRANCO, R. Y. D. S. et al. Uxmood - a tool to investigate the user experience (ux) based on multimodal sentiment analysis and information visualization (infovis). In: **2019 23rd International Conference Information Visualisation (IV)**. [S.l.: s.n.], 2019. p. 175–180.

FRANCO, R. Y. da S. et al. Uxmood—a sentiment analysis and information visualization tool to support the evaluation of usability and user experience. **Information**, MDPI, v. 10, n. 12, p. 366, 2019.

FRITZ, M.; BERGER, P. D. **Improving the User Experience through Practical Data Analytics: Gain Meaningful Insight and Increase Your Bottom Line**. [S.l.]: Elsevier Science, 2015. ISBN 9780128006351.

GARDEY, J. C. et al. Predicting interaction effort in web interface widgets. **International Journal of Human-Computer Studies**, v. 168, p. 102919, 2022. ISSN 1071-5819. Disponível em: <https://doi.org/10.1016/j.ijhcs.2022.102919>.

GAROUSI, V.; FELDERER, M.; MäNTYLä, M. V. The need for multivocal literature reviews in software engineering: Complementing systematic literature reviews with grey literature. In: **Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering**. New York, NY, USA: Association for Computing Machinery, 2016. (EASE '16). Disponível em: <https://doi.org/10.1145/2915970.2916008>.

GAROUSI, V.; FELDERER, M.; MäNTYLä, M. V. Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. **Information and Software Technology**, v. 106, p. 101–121, 2019. ISSN 0950-5849. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0950584918301939>.

GERMANAKOS, P.; FICHTE, L. Eurekatax: A taxonomy for the representation and analysis of qualitative usability test data. **Multimodal Technologies and Interaction**, v. 4, p. 22, 05 2020.

GIBBONS, S. **UX research cheat sheet**. 2020. Available at: https://bit.ly/2LESwR7.

GILARDI, F.; ALIZADEH, M.; KUBLI, M. Chatgpt outperforms crowd workers for text-annotation tasks. **Proceedings of the National Academy of Sciences**, National Academy of Sciences, v. 120, n. 30, p. e2305016120, 2023.

GLASS, R.; DEMARCO, T. **Software Creativity 2.0**. [S.l.]: Developer.* Books, 2006. ISBN 9780977213313.

GLASS, R.; VESSEY, I. Contemporary application-domain taxonomies. **IEEE Software**, v. 12, n. 4, p. 63–76, 1995.

GÖBEL, F.; KIEFER, P.; RAUBAL, M. FeaturEyeTrack: automatic matching of eye tracking data with map features on interactive maps . **GeoInformatica**, v. 23, n. 4, p. 663–687, 2019.

GOODELL, H. et al. Collecting and Harnessing Rich Session Histories. In: **Tenth International Conference on Information Visualisation (IV'06)**. [S.l.: s.n.], 2006. p. 117–123.

GOODWIN, S. et al. VETA: Visual eye-tracking analytics for the exploration of gaze patterns and behaviours. **Visual Informatics**, v. 6, n. 2, p. 1–13, 2022. ISSN 2468-502X.

GOOGLE. **What is Recovery Enhanced Generation (RAG)?** 2025. Disponível em: <https://cloud.google.com/use-cases/retrieval-augmented-generation>.

GORLENKO, L.; ENGLEFIED, P. Usability error classification: Qualitative data analysis for ux practitioners. In: **CHI '06 Extended Abstracts on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2006. (CHI EA '06), p. 803–808. ISBN 1595932984. Disponível em: <https://doi.org/10.1145/1125451.1125610>.

GOTHELF, J. Using proto-personas for executive alignment. **UX Magazine**, v. 1, 2012.

GU, Z. et al. Visualizing group user behaviors for social network interaction design iteration. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 9188, p. 36–45, 2015.

HALABI, N.; WALLNER, G.; MIRZA-BABAEI, P. Assessing the Impact of Visual Design on the Interpretation of Aggregated Playtesting Data Visualization. In: **Proceedings of the Annual Symposium on Computer-Human Interaction in Play**. New York, NY, USA: Association for Computing Machinery, 2019. (CHI PLAY '19), p. 639–650. ISBN 9781450366885.

HART, S. G.; STAVELAND, L. E. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In: HANCOCK, P. A.; MESHKATI, N. (Ed.). **Human Mental Workload**. North-Holland, 1988, (Advances in Psychology, v. 52). p. 139–183. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0166411508623869>.

HARTSON, R.; PYLA, P. **The UX Book: Agile UX Design for a Quality User Experience**. [S.l.]: Elsevier Science, 2018. ISBN 978-0-12-805342-3.

HASSENZAHL, M. **Experience Design: Technology for All the Right Reasons**. [S.l.]: Morgan & Claypool Publishers, 2010. (Synthesis Lectures on Human-Centered Informatics). ISBN 9781608450480.

_____. The thing and i (summer of '17 remix). In: _____. **Funology 2: From Usability to Enjoyment**. Cham: Springer International Publishing, 2018. p. 17–31. ISBN 978-3-319-68213-6. Disponível em: <https://doi.org/10.1007/978-3-319-68213-6_2>.

HASSENZAHL, M.; TRACTINSKY, N. User experience - a research agenda. **Behaviour & Information Technology**, Taylor & Francis, v. 25, n. 2, p. 91–97, 2006.

HEVNER, A. A three cycle view of design science research. **Scandinavian Journal of Information Systems**, v. 19, 01 2007.

HOKKANEN, L.; KUUSINEN, K.; VÄÄNÄNEN, K. Minimum viable user experience: A framework for supporting product design in startups. In: SHARP, H.; HALL, T. (Ed.). **Agile Processes, in Software Engineering, and Extreme Programming**. [S.l.]: Springer International Publishing, 2016. p. 66–78. ISBN 978-3-319-33515-5.

HOKKANEN, L.; VÄÄNÄNEN-VAINIO-MATTILA, K. Ux work in startups: current practices and future needs. In: SPRINGER. **International Conference on Agile Software Development**. [S.l.], 2015. p. 81–92.

HORNBÆK, K.; HERTZUM, M. The notion of overview in information visualization. **International Journal of Human-Computer Studies**, v. 69, n. 7, p. 509–525, 2011. ISSN 1071-5819.

HOSSEINI, M.; HORBACH, S. P. Fighting reviewer fatigue or amplifying bias? considerations and recommendations for use of chatgpt and other large language models in scholarly peer review. **Research integrity and peer review**, Springer, v. 8, n. 1, p. 4, 2023.

HOU, X. et al. Large language models for software engineering: A systematic literature review. **ACM Transactions on Software Engineering and Methodology**, ACM New York, NY, v. 33, n. 8, p. 1–79, 2024. Disponível em: <https://doi.org/10.1145/3695988>.

HUSSAIN, J. et al. A multimodal deep log-based user experience (ux) platform for ux evaluation. **Sensors**, v. 18, p. 1622, 05 2018.

IIVARI, J.; VENABLE, J. Action research and design science research - seemingly similar but decisively dissimilar. In: . [S.l.: s.n.], 2009. p. 1642–1653.

ISHERWOOD, M. **Guide to Audience Data for UX Design**. 2020. Disponível em: <https://uxdesigntemplates.com/evidence/method/audience-data>.

JASIM, M. et al. Supporting Serendipitous Discovery and Balanced Analysis of Online Product Reviews with Interaction-Driven Metrics and Bias-Mitigating Suggestions. In: **Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2022. (CHI '22), p. 1–24. ISBN 9781450391573.

KALANTARI, R.; LETHBRIDGE, T. Characterizing ux evaluation in software modeling tools: a literature review. **Ieee Access**, v. 10, p. 131509–131527, 2022.

KARAPANOS, E.; MARTENS, J.-b.; HASSENZAHL, M. Reconstructing experiences with iscale. **International Journal of Human-Computer Studies**, v. 70, p. 849–865, 11 2012.

KARAPANOS, E.; MARTENS, J.-B.; HASSENZAHL, M. Reconstructing experiences with iScale. **International Journal of Human-Computer Studies**, v. 70, n. 11, p. 849–865, 2012. ISSN 1071-5819.

KASHFI, P.; FELDT, R.; NILSSON, A. Integrating ux principles and practices into software development organizations: A case study of influencing events. **J. Syst. Softw.**, Elsevier Science Inc., USA, v. 154, n. C, p. 37–58, aug 2019. ISSN 0164-1212. Disponível em: <https://doi.org/10.1016/j.jss.2019.03.066>.

KEPPLINGER, D. et al. See, Feel, Move: Player Behaviour Analysis through Combined Visualization of Gaze, Emotions, and Movement. In: **Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2020. (CHI '20), p. 1–14. ISBN 9781450367080.

KHALIFA, M.; ALBADAWY, M. Using artificial intelligence in academic writing and research: An essential productivity tool. **Computer Methods and Programs in Biomedicine Update**, Elsevier, p. 100145, 2024. Disponível em: <https://doi.org/10.1016/j.cmpbup.2024.100145>.

KIEFFER., S. et al. Specification of a ux process reference model towards the strategic planning of ux activities. In: INSTICC. **Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: HUCAPP,**. [S.l.]: SciTePress, 2019. p. 74–85. ISBN 978-989-758-354-4.

KITCHENHAM, B. A.; CHARTERS, S. **Guidelines for performing Systematic Literature Reviews in Software Engineering**. [S.l.], 2007.

KOESTEN, L.; SIMPERL, E. Ux of data: Making data available doesn't make it usable. **Interactions**, Association for Computing Machinery, New York, NY, USA, v. 28, n. 2, p. 97–99, mar 2021. ISSN 1072-5520.

KOTAPALLE, G. R.; KANDALA, H.; GADE, K. S. Extracting relationship between browser history items for improved client-side analytics and recommendations. In: **2018 3rd International Conference on Contemporary Computing and Informatics (IC3I)**. [S.l.: s.n.], 2018. p. 141–146.

KUCHER, K.; PARADIS, C.; KERREN, A. The state of the art in sentiment visualization. **Computer Graphics Forum**, v. 37, n. 1, p. 71–96, 2018.

KUJALA, S. et al. UX Curve: A method for evaluating long-term user experience. **Interacting with Computers**, v. 23, n. 5, p. 473–483, 2011. ISSN 0953-5438. Feminism and HCI: New Perspectives.

LACHNER, F. et al. Quantified ux: Towards a common organizational understanding of user experience. In: **Proceedings of the 9th Nordic Conference on Human-Computer Interaction**. New York, NY, USA: Association for Computing Machinery, 2016. (NordiCHI '16). ISBN 9781450347631. Disponível em: <https://doi.org/10.1145/2971485.2971501>.

______. Quantified UX: Towards a Common Organizational Understanding of User Experience. In: **Proceedings of the 9th Nordic Conference on Human-Computer Interaction**. New York, NY, USA: Association for Computing Machinery, 2016. (NordiCHI '16), p. 1–10. ISBN 9781450347631.

LAM, H. et al. Empirical studies in information visualization: Seven scenarios. **IEEE Transactions on Visualization and Computer Graphics**, v. 18, n. 9, p. 1520–1536, 2012.

LAZAR, J.; FENG, J. H.; HOCHHEISER, H. **Research Methods in Human Computer Interaction**. Boston, NY, USA: Morgan Kaufmann, 2017.

LEMON, C. et al. A narrative review of methods for applying user experience in the design and assessment of mental health smartphone interventions. **International Journal of Technology Assessment in Health Care**, v. 36, p. 64–70, 2020.

LEPPICH, D. et al. Dux: a dataset of user interactions and user emotions. **i-com**, v. 22, n. 2, p. 101–123, 2023. Disponível em: <https://doi.org/10.1515/icom-2023-0014>.

LI, P.; LIU, Z. Y. E-customized product: User-centered co-design experiences. **KSII Transactions on Internet and Information Systems**, v. 14, n. 9, p. 3680–3692, 2020.

LI, Q. et al. A Visual Analytics Approach for Understanding Reasons behind Snowballing and Comeback in MOBA Games. **IEEE Transactions on Visualization and Computer Graphics**, v. 23, n. 1, p. 211–220, 2017.

LIN, Z. et al. Using pleasurability to compare wristwatches and iot smartwatches: providing novel insights into ux design. **Proceedings of the Design Society**, v. 3, p. 3761–3770, 2023.

LIU, J.; EAGAN, J. ADQDA: A Cross-Device Affinity Diagramming Tool for Fluid and Holistic Qualitative Data Analysis. **Proceedings of the ACM on Human-Computer Interaction**, Association for Computing Machinery, New York, NY, USA, v. 5, n. ISS, p. 1–19, nov. 2021.

LIU, V.; CHILTON, L. B. Design guidelines for prompt engineering text-to-image generative models. In: **Proceedings of the 2022 CHI conference on human factors in computing systems**. [S.l.: s.n.], 2022. p. 1–23.

LUTHER, L.; TIBERIUS, V.; BREM, A. User experience (ux) in business, management, and psychology: A bibliometric mapping of the current state of research. **Multimodal Technologies and Interaction**, MDPI AG, v. 4, n. 2, p. 18, May 2020. ISSN 2414-4088. Disponível em: <http://dx.doi.org/10.3390/mti4020018>.

MACCORMICK, D.; ZAMAN, L. Echo: Analyzing Gameplay Sessions by Reconstructing Them From Recorded Data. In: **Proceedings of the Annual Symposium on Computer-Human Interaction in Play**. New York, NY, USA: Association for Computing Machinery, 2020. (CHI PLAY '20), p. 281–293. ISBN 9781450380744.

______. Echoing the Gameplay: Analyzing Gameplay Sessions across Genres by Reconstructing Them from Recorded Data . **International Journal of Human–Computer Interaction**, Taylor & Francis, v. 39, n. 1, p. 52–84, 2023.

MACEDO, M. et al. Supporting user-centered requirements elicitation from lean personas: A ux data visualization-based approach. In: INSTICC. **Proceedings of the 26th International Conference on Enterprise Information Systems - Volume 2: ICEIS**. [S.l.]: SciTePress, 2024. p. 461–472. ISBN 978-989-758-692-7.

MACEDO, M.; ZAINA, L. Beyond raw data: A taxonomy for ux data exploration to support the design of interactive systems. **International Journal of Human–Computer Interaction**, Taylor & Francis, v. 0, n. 0, p. 1–27, 2025. Disponível em: <https://doi.org/10.1080/10447318.2025.2532711>.

MACEDO, M. P.; ZAINA, L. Mind the gap between ux data and visualization proposals: An approach emerged from the grey literature to support the analysis of user dissatisfaction. In: **Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2024. (IHC '24). ISBN 9798400712241. Disponível em: <https://doi.org/10.1145/3702038.3702072>.

MARTINELLI, S.; LOPES, L.; ZAINA, L. Ux research in the software industry: an investigation of long-term ux practices. In: **Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2022. (IHC '22). ISBN 9781450395069.

MCKENNA, S. et al. Design activity framework for visualization design. **IEEE Transactions on Visualization and Computer Graphics**, v. 20, n. 12, p. 2191–2200, 2014.

MEDLER, B.; JOHN, M.; LANE, J. Data Cracker: Developing a Visual Game Analytic Tool for Analyzing Online Gameplay. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2011. (CHI '11), p. 2365–2374. ISBN 9781450302289.

MELEGATI, J.; GUERRA, E.; WANG, X. Understanding hypotheses engineering in software startups through a gray literature review. **Information and Software Technology**, v. 133, p. 106465, 2021. ISSN 0950-5849. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0950584920302111>.

MENEWEGER, T. et al. Characteristics of narrative textual data linked to user experiences. p. 2605–2610, 2014.

MILLET, B. Ux research methods for designing interactive media. In: _____. **Interaction in Digital News Media: From Principles to Practice**. Cham: Springer International Publishing, 2018. p. 85–113. ISBN 978-3-319-96253-5. Disponível em: <https://doi.org/10.1007/978-3-319-96253-5_5>.

MOELLENDORFF, M.; HASSENZAHL, M.; PLATZ, A. Dynamics of user experience: How the perceived quality of mobile phones changes over time. In User experience–towards a unified view. In: **Workshop at the 4th Nordic Conference on Human-Computer Interaction**. [S.l.: s.n.], 2006. p. 74–78.

MÓRO, R.; DARÁZ, J.; BIELIKOVÁ, M. Visualization of gaze tracking data for ux testing on the web. In: **HT (Doctoral Consortium/Late-breaking Results/Workshops)**. [S.l.: s.n.], 2014.

MÓRO, R.; DARÁŽ, J.; BIELIKOVÁ, M. Visualization of gaze tracking data for UX testing on the Web. In: **CEUR Workshop Proceedings**. [S.l.: s.n.], 2014. v. 1210.

MORRISON, C.; DOHERTY, G. Analyzing engagement in a web-based intervention platform through visualizing log-data. **Journal of Medical Internet Research**, v. 16, n. 11, p. e252, 2014.

MUNZNER, T. **Visualization analysis and design**. [S.l.]: A K Peters, 2014. (A.K. Peters visualization series).

NARANG, B.; TRIVEDI, P.; DUBEY, M. K. Towards an understanding of ux (user experience) and uxd (user experience design), an applicability based framework forecommerce, intranets, mobile & tablet & web usability. **International Journal of Advanced Research in Computer Science**, v. 8, p. 2764–2768, 2017.

NORMAN, D.; NIELSEN, J. **Nielsen Norman Group - The definition of user experience (UX)**. 2018. <https://www.nngroup.com/articles/definition-user-experience/>. Acessado: 20/09/2020.

OFFERMANN, P. et al. Artifact types in information systems design science – a literature review. In: WINTER, R.; ZHAO, J. L.; AIER, S. (Ed.). **Global Perspectives on Design Science Research"**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 77–92.

OHASHI, K. et al. Focusing requirements elicitation by using a ux measurement method. In: **2018 IEEE 26th International Requirements Engineering Conference (RE)**. [S.l.: s.n.], 2018. p. 347–357.

OLDENBURG, I.; HILLENBRAND, C. Using ux design in inclusive teacher education in higher education to bridge the research-to-practice gap. **10th International Conference on Higher Education Advances (HEAd'24)**, 2024.

ORMEÑO, Y. I.; PANACH, J. I. Mapping study about usability requirements elicitation. In: SALINESI, C.; NORRIE, M. C.; PASTOR, Ó. (Ed.). **Advanced Information Systems Engineering**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. p. 672–687. ISBN 978-3-642-38709-8.

PAUL, C. L. Analyzing Card-Sorting Data Using Graph Visualization. **Journal of Usability Studies**, Usability Professionals' Association, Bloomingdale, IL, v. 9, n. 3, p. 87–104, maio 2014.

PEPPING, J. et al. Motiis: Fostering Parents' Awareness of Their Adolescents Emotional Experiences during Gaming. In: **Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society**. New York, NY, USA: Association for Computing Machinery, 2020. (NordiCHI '20), p. 1–11. ISBN 9781450375795.

PETERSEN, K. et al. Systematic mapping studies in software engineering. In: **Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering**. Swindon, UK: BCS Learning & Development Ltd., 2008. (EASE '08), p. 68–77.

PETERSEN, K.; VAKKALANKA, S.; KUZNIARZ, L. Guidelines for conducting systematic mapping studies in software engineering: An update. **Information and Software Technology**, v. 64, p. 1–18, 2015. ISSN 0950-5849. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0950584915000646>.

PINHEIRO, E. G. et al. On the contributions of non-technical stakeholders to describing ux requirements by applying proto-persona. **Journal of Software Engineering Research and Development**, v. 7, p. 8:1 – 8:19, Dec. 2019. Disponível em: <https://sol.sbc.org.br/journals/index.php/jserd/article/view/155>.

RAHMAN, M. M. et al. Chatgpt and academic research: A review and recommendations based on practical examples. **Rahman, M., Terano, HJR, Rahman, N., Salamzadeh, A., Rahaman, S.(2023). ChatGPT and Academic Research: A Review and Recommendations Based on Practical Examples. Journal of Education, Management and Development Studies**, v. 3, n. 1, p. 1–12, 2023.

RAINER, A.; WILLIAMS, A. Using blog-like documents to investigate software practice: Benefits, challenges, and research directions. **Journal of Software: Evolution and Process**, v. 31, n. 11, p. e2197, 2019. E2197 smr.2197. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/smr.2197>.

RAMALHO, B.; JORGE, J.; GAMA, S. Representing uncertainty through sentiment and stance visualizations: A survey. **Graphical Models**, v. 129, p. 101191, 2023. ISSN 1524-0703.

RHIE, J. H. L. Y. L.; YUN, M. H. Evaluating representativeness of qualitative text data in identifying ux issues. **International Journal of Human–Computer Interaction**, Taylor & Francis, v. 33, n. 11, p. 868–881, 2017. Disponível em: <https://doi.org/10.1080/10447318.2017.1294337>.

RIJNDERS, F.; WALLNER, G.; BERNHAUPT, R. Live Feedback for Training Through Real-Time Data Visualizations: A Study with League of Legends. **Proceedings of the ACM on Human-Computer Interaction**, Association for Computing Machinery, New York, NY, USA, v. 6, n. CHI PLAY, p. 1–23, out. 2022.

RIVERO, L.; CONTE, T. A systematic mapping study on research contributions on ux evaluation technologies. In: **Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2017. (IHC 2017). ISBN 9781450363778.

RODDEN, K. Applying a sunburst visualization to summarize user navigation sequences. **IEEE computer graphics and applications**, IEEE, v. 34, n. 5, p. 36–40, 2014.

RUOTSALO, T. et al. Interactive Intent Modeling for Exploratory Search. **ACM Transactions on Information Systems**, Association for Computing Machinery, New York, NY, USA, v. 36, n. 4, p. 1–46, out. 2018. ISSN 1046-8188.

SAGA, R. et al. Visualization of customer expectations from Web text using co-occurrence graph and auto-labeling in the service market. In: **VISIGRAPP 2017 - Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications**. [S.l.: s.n.], 2017. v. 3, p. 317–321.

SALDAÑA, J. **The coding manual for qualitative researchers**. [S.l.]: Sage Publications, 2021.

SALMAN, I.; MISIRLI, A. T.; JURISTO, N. Are students representatives of professionals in software engineering experiments? In: **2015 IEEE/ACM 37th IEEE International Conference on Software Engineering**. [S.l.: s.n.], 2015. v. 1, p. 666–676. ISSN 1558-1225.

SALMINEN, J. et al. Personas and analytics: a comparative user study of efficiency and effectiveness for a user identification task. In: **Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems**. [S.l.: s.n.], 2020. p. 1–13.

______. Design issues in automatically generated persona profiles: A qualitative analysis from 38 think-aloud transcripts. In: **Proceedings of the 2019 Conference on Human Information Interaction and Retrieval**. New York, NY, USA: Association for Computing Machinery, 2019. (CHIIR ’19), p. 225–229. ISBN 9781450360258.

SCHMIDT, A. et al. Simulating the human in hcd with chatgpt: Redesigning interaction design with ai. **Interactions**, Association for Computing Machinery, New York, NY, USA, v. 31, n. 1, p. 24–31, jan. 2024. ISSN 1072-5520. Disponível em: <https://doi.org/10.1145/3637436>.

SCHMIDT, D. C. et al. Towards a catalog of prompt patterns to enhance the discipline of prompt engineering. **ACM SIGAda Ada Letters**, ACM New York, NY, USA, v. 43, n. 2, p. 43–51, 2024.

SEGURA, V.; FERREIRA, J. J.; BARBOSA, S. D. J. BONNIE: Building Online Narratives from Noteworthy Interaction Events. In: **2018 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)**. [S.l.: s.n.], 2018. p. 233–238.

SHAMIM, A.; BALAKRISHNAN, V.; TAHIR, M. Evaluation of opinion visualization techniques. **Information Visualization**, v. 14, n. 4, p. 339–358, 2014.

SHEN, Z. et al. Visual analysis of massive web session data. In: **IEEE Symposium on Large Data Analysis and Visualization (LDAV)**. [S.l.: s.n.], 2012. p. 65–72.

SHNEIDERMAN, B. The eyes have it: a task by data type taxonomy for information visualizations. In: **Proceedings 1996 IEEE Symposium on Visual Languages**. [S.l.: s.n.], 1996. p. 336–343.

SLONE, D. J. Visualizing qualitative information. **The Qualitative Report**, v. 14, p. 489–497, 2009.

SOURE, E. J. et al. CoUX: Collaborative Visual Analysis of Think-Aloud Usability Test Videos for Digital Interfaces. **IEEE Transactions on Visualization and Computer Graphics**, v. 28, n. 1, p. 643–653, 2022.

SOUSA, T. A. F. de; BARBOSA, S. D. J. Sistema de recomendação para apoiar a construção de gráficos com dados estatísticos. In: **Proceedings of the 12th Brazilian Symposium on Human Factors in Computing Systems**. Porto Alegre, BRA: SBC, 2013. (IHC ’13), p. 168–177.

STANDARD, I.; ISO, B. **ISO 9241-210: 2019-Ergonomics of human-system interaction–Part 210: Human-centred design for interactive systems**. [S.l.]: ISO, 2019.

STASKO, J. Value-Driven Evaluation of Visualizations. In: **Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization**. New York, NY, USA: Association for Computing Machinery, 2014. (BELIV ’14), p. 46–53. ISBN 9781450332095.

SUN, M.; CONVERTINO, G.; DETWEILER, M. Designing a unified cloud log analytics platform. In: **Proceedings - 2016 International Conference on Collaboration Technologies and Systems, CTS 2016**. [S.l.: s.n.], 2016. p. 257–266.

SYKOWNIK, P.; BORN, F.; MASUCH, M. Can You Hear the Player Experience? A Pipeline for Automated Sentiment Analysis of Player Speech. In: **2019 IEEE Conference on Games (CoG)**. [S.l.: s.n.], 2019. p. 1–4.

TEIXEIRA., G. V.; ZAINA., L. A. M. Using lean personas to the description of ux-related requirements: A study with software startup professionals. In: INSTICC. **Proceedings of the 24th International Conference on Enterprise Information Systems - Volume 2: ICEIS,**. [S.l.]: SciTePress, 2022. p. 211–222. ISBN 978-989-758-569-2. ISSN 2184-4992.

TONG, Y. et al. A data-driven approach for integrating hedonic quality and pragmatic quality in user experience modeling. **Journal of Computing and Information Science in Engineering**, 2022.

UNTERKALMSTEINER, M. et al. Software startups - a research agenda. **E-Informatica Software Engineering Journal**, v. 2016, p. 89–124, 10 2016.

UNTERKALMSTEINER, M.; FELDT, R.; GORSCHEK, T. A taxonomy for requirements engineering and software test alignment. **ACM Trans. Softw. Eng. Methodol.**, Association for Computing Machinery, New York, NY, USA, v. 23, n. 2, apr 2014. ISSN 1049-331X. Disponível em: <https://doi.org/10.1145/2523088>.

VASWANI, A. et al. Attention is all you need. **Advances in neural information processing systems**, v. 30, 2017.

VENABLE, J. A framework for design science research activities. **Proceedings of the 2006 Information Resource Management Association Conference**, p. 184–187, 01 2006.

VENKATESH, V.; BALA, H. Technology Acceptance Model 3 and a Research Agenda on Interventions. **Decision Sciences**, Wiley, v. 39, n. 2, p. 273–315, 5 2008.

VICTORELLI, E. Z.; REIS, J. C. D. Human-Data Interaction Design Guidelines for Visualization Systems. In: **Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2020. (IHC '20). ISBN 9781450381727.

WALL, E. et al. A Heuristic Approach to Value-Driven Evaluation of Visualizations. **IEEE Transactions on Visualization and Computer Graphics**, v. 25, n. 1, p. 491–500, 2019.

WALLNER, G.; HALABI, N.; MIRZA-BABAEI, P. Aggregated Visualization of Playtesting Data. In: **Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2019. (CHI '19), p. 1–12. ISBN 9781450359702.

WALLNER, G.; KRIGLSTEIN, S. A spatiotemporal visualization approach for the analysis of gameplay data. In: **Conference on Human Factors in Computing Systems - Proceedings**. [S.l.: s.n.], 2012. p. 1115–1124.

______. Visualization-based analysis of gameplay data - a review of literature. **Entertainment Computing**, v. 4, n. 3, p. 143–155, 2013.

______. PLATO: A visual analytics system for gameplay data. **Computers & Graphics**, v. 38, p. 341–356, 2014. ISSN 0097-8493.

______. Multivariate Visualization of Game Metrics: An Evaluation of Hexbin Maps. In: **Proceedings of the Annual Symposium on Computer-Human Interaction in Play**. New York, NY, USA: Association for Computing Machinery, 2020. (CHI PLAY '20), p. 572–584. ISBN 9781450380744.

WANG, Y. et al. Animated Narrative Visualization for Video Clickstream Data. In: **SIGGRAPH ASIA 2016 Symposium on Visualization**. New York, NY, USA: Association for Computing Machinery, 2016. (SA '16), p. 1–8. ISBN 9781450345477.

WARE, C. **Information Visualization: perception for design**. 3. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2012. ISBN 9780123814647, 9780123814654.

WATSON, R. B.; SPYRIDAKIS, J. H. A Tool to Remotely Collect and Visualize Users' Interactions with Web-Based Content. In: **Proceedings of the 34th ACM International Conference on the Design of Communication**. New York, NY, USA: Association for Computing Machinery, 2016. (SIGDOC '16), p. 1–6. ISBN 9781450344951.

WIęCKOWSKA, M.; RUDNICKA, P. Visualising user–website interaction: description and evaluation of a teaching method. **Universal Access in the Information Society**, Springer-Verlag, Berlin, Heidelberg, v. 22, n. 2, p. 297–313, out. 2021. ISSN 1615-5289.

WOHLIN, C. et al. **Experimentation in software engineering**. [S.l.]: Springer Science & Business Media, 2012.

YANG, S.-Y.; CHOU, W. H.; CHEN, H.-X. Development of the automated tool to perform analysis and visualisation of user reviews. **International Journal of Arts and Technology**, v. 12, n. 3, p. 218–237, 2020. Disponível em: <https://www.inderscienceonline.com/doi/abs/10.1504/IJART.2020.109814>.

YEN, Y.-C. G.; KIM, J. O.; BAILEY, B. P. Decipher: An Interactive Visualization Tool for Interpreting Unstructured Design Feedback from Multiple Providers. In: **Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2020. (CHI '20), p. 1–13. ISBN 9781450367080.

YOO, S. et al. Saliency-based gaze visualization for eye movement analysis. **Sensors**, v. 21, n. 15, p. 5178, 2021.

ZAINA, L. A.; SHARP, H.; BARROCA, L. Ux information in the daily work of an agile team: A distributed cognition analysis. **International Journal of Human-Computer Studies**, v. 147, p. 102574, 2021. ISSN 1071-5819. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1071581920301762>.

ZHANG, Y. et al. Controlled vocabularies for scientific data: Users and desired functionalities. **Proceedings of the Association for Information Science and Technology**, v. 52, n. 1, p. 1–8, 2015.

Apenddix A

See Table 52.

ID	Title
P01	A 3D visualization framework to social network monitoring and analysis (Campos Filho; NOVAES; GOMES, 2015)
P02	A Generic Visualization Approach Supporting Task-Based Evaluation of Usability and User Experience (BERNHAUPT et al., 2020a)
P03	A spatiotemporal visualization approach for the analysis of gameplay data (WALLNER; KRIGLSTEIN, 2012)
P04	A Tool to Remotely Collect and Visualize Users' Interactions with Web-Based Content (WATSON; SPYRIDAKIS, 2016)
P05	A Visual Analytics Approach for Understanding Reasons behind Snowballing and Comeback in MOBA Games (LI et al., 2017)
P06	ADQDA: A Cross-device Affinity Diagramming Tool for Fluid and Holistic Qualitative Data Analysis (LIU; EAGAN, 2021)
P07	Aggregated Visualization of Playtesting Data (WALLNER; HALABI; MIRZA-BABAEI, 2019)
P08	An interactive visualization for tabbed browsing behavior analysis (CERNEA et al., 2014)
P09	Analyzing Card-Sorting Data Using Graph Visualization (PAUL, 2014)
P10	Analyzing engagement in a web-based intervention platform through visualizing log-data (MORRISON; DOHERTY, 2014)
P11	Analyzing spatial user behavior in computer games using geographic information systems (DRACHEN; CANOSSA, 2009a)
P12	Animated Narrative Visualization for Video Clickstream Data (WANG et al., 2016)
P13	Assessing the impact of visual design on the interpretation of aggregated playtesting data visualization (HALABI; WALLNER; MIRZA-BABAEI, 2019)
P14	BONNIE: Building Online Narratives from Noteworthy Interaction Events (SEGURA; FERREIRA; BARBOSA, 2018)
P15	Can you hear the player experience? A pipeline for automated sentiment analysis of player speech (SYKOWNIK; BORN; MASUCH, 2019)
P16	Collecting and harnessing rich session histories (GOODELL et al., 2006)
P17	CoUX: Collaborative Visual Analysis of Think-Aloud Usability Test Videos for Digital Interfaces (SOURE et al., 2022)
P18	Data cracker: Developing a visual game analytic tool for analyzing online gameplay (MEDLER; JOHN; LANE, 2011)
P19	Decipher: An Interactive Visualization Tool for Interpreting Unstructured Design Feedback from Multiple Providers (YEN; KIM; BAILEY, 2020)
P20	Designing a unified cloud log analytics platform (SUN; CONVERTINO; DETWEILER, 2016)
P21	E-customized product: User-centered co-design experiences (LI; LIU, 2020)
P22	Echo: Analyzing Gameplay Sessions by Reconstructing Them from Recorded Data (MACCORMICK; ZAMAN, 2020)
P23	Echoing the Gameplay: Analyzing Gameplay Sessions across Genres by Reconstructing Them from Recorded Data (MACCORMICK; ZAMAN, 2023)
P24	EmojiText: An Information Visualization Technique for Analyzing Phrases and Sentiments (COSTA et al., 2021)
P25	Emotion-prints: Interaction-driven emotion visualization on multi-touch interfaces (CERNEA et al., 2015)
P26	Enriching task models with usability and user experience evaluation data (BERNHAUPT et al., 2019)
P27	Extracting relationship between browser history items for improved client-side analytics and recommendations (KOTAPALLE; KANDALA; GADE, 2018)
P28	FeaturEyeTrack: automatic matching of eye tracking data with map features on interactive maps (GÖBEL; KIEFER; RAUBAL, 2019)
P29	Gameplay analysis through state projection (ANDERSEN et al., 2010)
P30	GameVis: Game data visualization for the web (FEITOSA et al., 2015)
P31	InteracDiff: Visualizing and Interacting with UX-Data (DITTRICH; HOF; WIETHOFF, 2019b)
P32	Interactive Exploration of Music Listening Histories (DIAS; FONSECA; GONÇALVES, 2012)
P33	Interactive Intent Modeling for Exploratory Search (RUOTSALO et al., 2018)
P34	Live Feedback for Training Through Real-Time Data Visualizations: A Study with League of Legends (RIJNDERS; WALLNER; BERNHAUPT, 2022)
P35	Mapping User Trajectories to Examine Behavior and Outcomes in Digital Health Intervention Data (CHEN et al., 2019)

Table 52 – Selected papers identified (Continued on the next page).

ID	Title
P36	Miria: A mixed reality toolkit for the in-situ visualization and analysis of spatio-temporal interaction data (BÜSCHEL; LEHMANN; DACHSELT, 2021)
P37	Motiis: Fostering Parents' Awareness of Their Adolescents Emotional Experiences during Gaming (PEPPING et al., 2020)
P38	Multivariate Visualization of Game Metrics: An Evaluation of Hexbin Maps (WALLNER; KRIGLSTEIN, 2020)
P39	Personal Web Library: organizing and visualizing Web browsing history (DU et al., 2018)
P40	PLATO: A visual analytics system for gameplay data (WALLNER; KRIGLSTEIN, 2014)
P41	PopHistory: Animated Visualization of Personal Web Browsing History (CARRASCO; KOH; MALIK, 2017)
P42	Quantified UX: Towards a common organizational understanding of user experience (LACHNER et al., 2016b)
P43	Reconstructing experiences with iScale (KARAPANOS; MARTENS; HASSENZAHL, 2012b)
P44	Saliency-based gaze visualization for eye movement analysis (YOO et al., 2021)
P45	See, Feel, Move: Player Behaviour Analysis through Combined Visualization of Gaze, Emotions, and Movement (KEPPLINGER et al., 2020)
P46	Supporting Serendipitous Discovery and Balanced Analysis of Online Product Reviews with Interaction-Driven Metrics and Bias-Mitigating Suggestions (JASIM et al., 2022)
P47	Towards gameplay analysis via gameplay metrics (DRACHEN; CANOSSA, 2009b)
P48	Towards the detection of UX Smells: The support of visualizations (BUONO et al., 2020b)
P49	Understanding User Behavior in Digital Libraries Using the MAGUS Session Visualization Tool (BOGAARD et al., 2020)
P50	UXmood - A Tool to Investigate the User Experience (UX) Based on Multimodal Sentiment Analysis and Information Visualization (InfoVis) (Da Silva Franco et al., 2019)
P51	VETA: Visual eye-tracking analytics for the exploration of gaze patterns and behaviours (GOODWIN et al., 2022)
P52	Visual analysis of massive web session data (SHEN et al., 2012)
P53	Visualising user–website interaction: description and evaluation of a teaching method (WIęCKOWSKA; RUDNICKA, 2021)
P54	Visualization of customer expectations from Web text using co-occurrence graph and auto-labeling in the service market (SAGA et al., 2017)
P55	Visualization of gaze tracking data for UX testing on the Web (MÓRO; DARÁŽ; BIELIKOVÁ, 2014)
P56	Visualizing event sequence data for user behavior evaluation of in-vehicle information systems (EBEL; LINGENFELDER; VOGELSANG, 2021b)
P57	Visualizing group user behaviors for social network interaction design iteration (GU et al., 2015)

Table 52 – Selected papers identified (Continued).

Apenddix B

See Table 53.

Term	Description	Source
Andrienko and Andrienko (2011) Algorithm	Algorithm to partition a spatial region based on the clustering of points in a movement trajectory.	Andrienko, N., Andrienko. G. (2011). Spatial Generalization and Aggregation of Massive Movement Data.
ArcGIS	Geographic information system for spatial data visualization.	https://www.esri.com/en-us/arcgis/about-arcgis/overview
Artificial Neural Networks	Machine learning model based on the interconnection of units with iteratively adjusted weights to achieve an optimal non-linear function.	Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986). Learning representations by back-propagating errors.
Auditok	Python library to detect sound events in audio.	https://pypi.org/project/auditok/0.1.8/
Box2DWeb	Game engine that allows animation of symbols based on physical principles.	https://box2d.org/
Bubble Sets	Visualization technique to portray spatial and membership relationships simultaneously.	Collins, C., Penn, G., Carpendale, S. (2009). Bubble Sets: Revealing Set Relations with Isocontours over Existing Visualizations.
Classical Multidimensional Scaling (CMDS)	Algorithm to visualize items spatially based on a dissimilarity matrix.	Kruskal, J. B., Wish, M. (1978). Multidimensional Scaling.
D3.js	JavaScript library for creating data visualizations.	http://d3js.org/
Density Based Spatial Clustering of Applications with Noise (DBSCAN)	Density-based clustering algorithm.	Ester, M., Kriegel, H.-P., Sander, J., Xu, X. (1996) A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.
Gensim	Python library with natural language processing algorithms.	Řehůřek, R., Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora.
Graphviz5	Software for visualizing graphs and networks.	https://graphviz.org/
HAMSTERS-XLE	Software environment for modeling tasks that compose an interaction.	Martinie, C., Palanque, P., Bouzekri, E., Cockburn, A., Canny, A., Barboni, E. (2019) Analysing and Demonstrating Tool-Supported Customizable Task Notations.
Hadoop	Java framework for dealing with the processing of large databases.	White, T. (2010). Hadoop: The definitive guide.
INTUI	Questionnaire to measure different dimensions of an intuitive interaction.	Ullrich, D., Diefenbach, S. (2010). INTUI. Exploring the Facets of Intuitive Interaction.
K-means	Clustering algorithm based on the minimization of intra-cluster distance.	Lloyd, S. (1982). Least squares quantization in PCM.
Linear Associative Reinforcement Learning (LinRel)	Reinforcement learning model to deal with the exploitation-exploration trade-off in decision making.	Auer, P. (2000). Using confidence bounds for exploitation-exploration trade-offs.
Multidimensional Scaling for Java (MDSJ)	Java library to apply the multidimensional scaling algorithm.	Algorithmics Group. (2009). MDSJ: Java Library for Multidimensional Scaling (Version 0.2).
NASA-Task Load Index (NASA-TLX)	Questionnaire to measure the perceived workload during a task.	Hart, S. G., Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research.
Newman (2004) Algorithm	Clustering algorithm for community detection in graphs.	Newman, M. E. J., 2004. Fast algorithm for detecting community structure in networks.

Table 53 – Description of emerging methods, techniques, and tools from the SML (Continued on the next page).

Term	Description	Source
NVivo	Software for qualitative data analysis.	<https://help-nv.qsrinternational.com/20/win/Content/about-nvivo/about-nvivo.htm>
OGAMA	Software for analyzing eye tracking and mouse tracking data.	Voßkühler, A., Nordmeier, V., Kuchinke, L., Jacobs, A. M. (2008) OGAMA (Open Gaze and Mouse Analyzer): Open-source software designed to analyze eye and mouse movements in slideshow study designs.
OpenCV	Computer vision and machine learning library for multiple programming languages.	https://opencv.org/about/
OpenGL	Computer graphics API.	https://www.opengl.org/
Praat-Parselmouth	Python library for phonetic analysis.	Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat.
Protovis	JavaScript library for creating data visualizations.	Bostock, M., Heer, J. (2009). Protovis: A Graphical Toolkit for Visualization.
Quality Threshold (QT)	Clustering algorithm based on the diameter of the clusters.	Heyer, L. J., Kruglyak, S., and Yooseph, S. (1999). Exploring expression data: Identification and analysis of coexpressed genes.
Recommender systems' Quality of user experience (ResQue)	User-centered evaluation framework for recommender systems.	Pu, P., Chen, L., Hu, R. (2011). A user-centric evaluation framework for recommender systems.
Repertory Grid technique	Structured interview technique based on the terms used by the participant while describing their experience.	Fransella, F., Bell, R., Bannister, D., (2003). A Manual for Repertory Grid Technique.
Reverse Cuthill–McKee (RCM)	Sorting algorithm to reduce the envelope of sparse matrices.	George, A., Liu, J. W. H. (1981). Computer solution of large sparse positive definite systems.
Sun et al. (2012) Algorithm	Algorithm to verify isomorphism between graphs.	Sun, Z., Wang, H., Wang, H., Shao, B., Li, J. (2012). Efficient subgraph matching on billion node graphs.
Support Vector Machine (SVM)	Machine learning model based on the selection of support vectors that maximize the separation margin between distinct classes.	Cortes, C., Vapnik, V. (1995) Support-vector networks.
SWISH DataLab	Software environment for data analysis that supports the Prolog and R languages.	Bogaard, T., Wielemaker, J., Hollink, L., Van Ossenbruggen, J. (2017). SWISH DataLab: A web interface for data exploration and analysis.
Term frequency-inverse document frequency (TF-IDF)	Algorithm to measure the importance of a term in a document.	Salton, G., Buckley, C. (1988). Term-weighting approaches in automatic text retrieval.
User Experience Questionnaire (UEQ)	Questionnaire to measure the user experience of a product.	Laugwitz, B., Held, T., Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire.
Usefulness, Satisfaction, and Ease of use (USE)	Questionnaire to measure three dimensions of a product: usefulness, satisfaction, and ease of use.	Lund, A. (2001). Measuring Usability with the USE Questionnaire.
Unity	Game engine for creating 2D and 3D games and other visual content.	https://unity.com/
Universal Visualization Platform	System for creating analysis and visualization applications.	Gee, A.G., Li, H., Yu, M., Smrtic, M.B., Cvek, U., Goodell, H., Gupta, V., Lawrence, C., Zhou, J., Chiang, C.-H. and Grinstein, G.G. (2005) Universal Visualization Platform.
Valence Aware Dictionary and Sentiment Reasoner	Python library for classifying texts based on lexical analysis.	Hutto, C., Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text.
Velocity-Threshold Identification (I-VT)	Algorithm to differentiate types of gaze based on movement speed.	Salvucci, D., Goldberg, J. (2000). Identifying fixations and saccades in eye-tracking protocols.

Table 53 – Description of emerging methods, techniques, and tools from the SML (Continued).

Apenddix C

Approved grey articles see Table 54, Table 55, Table 56, and Table 57.

Article Citation	Short description	PDF link
AC1	Essential UX questions you need to ask about your digital product	9-120.PDF
AC2	Strategies to improve how your team communicates about data	2-17.PDF
AC3	Ultimate Guide to Collecting and Analyzing UX Research Data	4-54.PDF
AC4	How do you introduce data-driven design into your framework? What are the essential components for you and your colleagues to be able to effectively execute on a data-based approach?	7-137.PDF
AC5	What is Data-driven UX?	2-33.PDF
AC6	Designing Data-Driven Interfaces - I wanted to share some thoughts on how to arrive at a distinct and meaningful product.	7-94.PDF
AC7	Informed by data, driven by empathy: an introduction to user experience research at Booking.com	7-153.PDF
AC8	How to get started with quantitative UX research	1-9.PDF
AC9	How to choose the right UX metrics for your product	3-65.PDF
AC10	when applied data-driven	7-93.PDF
AC11	case study design data-driven	5-42.PDF
AC12	Role of Data in Measuring UX	2-121.PDF
AC13	4 Steps for Founders to Start Thinking Like a Data Scientist	14-86.PDF
AC14	The slipperiness of UX data	3-92.PDF
AC15	SEO & CRO-Informed UX | Data-Driven Ecommerce	7-129.PDF
AC16	In this article we will dig in what and why data-driven design is a better approach in delivering a great user experience	2-2.PDF
AC17	How Can Data Science Improve UX Design?	15-23.PDF
AC18	How to Make the Most of Quantitative Research in UX Design	1-42.PDF
AC19	Q&A - How do I analyse both quantitative and qualitative UX data?	8-70.PDF
AC20	Q&A - Is data analytics useful for UX research or design?	2-62.PDF
AC21	Q&A - What are some of the best data-driven eCommerce UI/UX optimization techniques?	7-53.PDF
AC22	3 Data Storytelling Tips for Software Companies to Design Analytics Features Instead of Hiring a UI/UX Designer	4-69.PDF
AC23	6 Steps to Optimize Your Website With Data-Driven Design	7-73.PDF
AC24	how Gitlab team explore ux data	4-62.PDF
AC25	THE COMPLETE GUIDE TO USER EXPERIENCE METRICS	14-77.PDF
AC26	project on github that uses google analytics to explore quantitative UX data. The main publication is an article on Uxdesign.	1-86.PDF
AC27	project in python to visualize UX data base on examples of Chapter 4 of book "Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics"	3-86.PDF
AC28	UX Checklist techniques with explanations about each topic.	4-108.PDF
AC29	A Beginner's Guide to Quantitative UX Research	1-3.PDF
AC30	Three Uses for Analytics in User-Experience Practice	3-7.PDF
AC31	Quantitative UX Research in Practice - To investigate how digital product teams use quantitative research to get "hard" data (or why they don't) we surveyed 429 UX professionals	1-2.PDF
AC32	Quantitative User-Research Methodologies: An Overview	1-1.PDF
AC33	Optimize your Website with UX Behavioral Data Analysis and Usability Testing	3-25.PDF
AC34	4 Ways to Implement Data For A Better User Experience Design	7-143.PDF
AC35	Manage Data Density, High-Level to Low-Level	2-114.PDF
AC36	UX Design (of visualizations) of data-intensive applications	2-71.PDF
AC37	Extracting Liquidware Stratusphere UX Data with Microsoft PowerBI	9-17.PDF
AC38	DATA DRIVEN WEB DESIGN: HOW TO BUILD BETTER-PERFORMING B2B WEBSITES	3-98.PDF
AC39	Using Data to Make Your Users Even Happier	6-88.PDF
AC40	9 Types of Web Analytics Tools — And How to Know Which Ones You Really Need	12-85.PDF
AC41	Digital Analytics: the great ally of User Experience (UX)	4-71.PDF
AC42	Customer Experience vs. User Experience, and How Customer Data Improves Both	2-55.PDF
AC43	3 Tools to Measure User Experience Data for Enterprise Applications	13-33.PDF
AC44	How to avoid being intimidated by quantitative UX metrics	6-53.PDF
AC45	User Experience Optimization: Process and Techniques for Continuous App Improvement	3-116.PDF

Table 54 – List of GLR approved articles (Part 1)

Article Citation	Short description	PDF link
AC46	The UX of data	12-77.PDF
AC47	4 Places to Find UX Data When User Research Isn't an Option	2-88.PDF
AC48	Tools for data-driven design of digital services	7-156.PDF
AC49	Interpreting analytics data in UX design	2-109.PDF
AC50	Data Driven Design: When Data Helps Build Great Products	7-57.PDF
AC51	How to make a UX data dashboard for your design team	2-54.PDF
AC52	Data Driven UX: It's More Than An Art	7-69.PDF
AC53	Data-informed design: Getting started with UX analytics	2-19.PDF
AC54	Methodology for Creating a Winning UX & Persuasive Design	4-127.PDF
AC55	Top 5 Web Analytics Tools Great for UX	12-47.PDF
AC56	Data-driven UX – how it can improve creating data visualizations	7-55.PDF
AC57	Common UX Design Methods and Techniques	8-23.PDF
AC58	What is UX Analytics?	1-74.PDF
AC59	UX Analysis Based on Usage Data – How Big Data helps to optimize the usability of applications	2-34.PDF
AC60	Data Driven UI/UX for a successful game	7-141.PDF
AC61	How To Build A Tracking Strategy For Your Customer Data	6-84.PDF
AC62	A Guide To UX Data Analysis For Actionable Insights	1-39.PDF
AC63	How to Create a More Data-Driven UX on Your E-commerce Website?	7-151.PDF
AC64	Quantitative research in Google Analytics : UX Metrics That Matter	1-33.PDF
AC65	What is analytics for UX designers? What is data driven design?	2-52.PDF
AC66	Why Embedded Analytics is a Game Changer for UX	4-98.PDF
AC67	UX Analytics 101: What is UX Analytics, Why It's Needed, & How to Use It	5-24.PDF
AC68	What UX Data Should Every Company Collect?	2-39.PDF
AC69	Data-Driven Design: What It Is and Why It Matters	2-95.PDF
AC70	How to Improve UX With AI and Machine Learning	4-122.PDF
AC71	How to Leverage UX Analytics for Data-Driven Product Design	2-83.PDF
AC72	The 7 most important UX KPIs and how to measure them	13-46.PDF
AC73	Top 3 Data Analytics to Track for Better UX	4-124.PDF
AC74	How to use the data to improve UX?	2-10.PDF
AC75	Data-Driven Design: Tap into Your Users' Micro-Moments for Better Engagement	7-155.PDF
AC76	text about Data Analysis and Interpretation	4-112.PDF
AC77	Formative UX Study Data Analysis:2 Simple Recipes	2-74.PDF
AC78	Process: Analyze UX interview data and communicate results	3-41.PDF
AC79	Selecting a UX Design Method: Making Small UX Data Big	2-69.PDF
AC80	4 Ways to Implement Data For A Better User Experience Design	2-6.PDF
AC81	Data Science + UX Design = More Conversions	15-7.PDF
AC82	Towards boosting the efficiency of UX researchers with intelligent visualization	18-7.PDF
AC83	Juicebox It's a multi-layered challenge in practice, as there are many ways data can be gathered, stored, and presented. In my early days at Juice, I designed an exercise primer called 30 Days to Data Storytelling(2), which spans a broad set of examples from principles.	5-43.PDF
AC84	With emphasis often placed on data analytics, you know by now it's important to track what's going on with your users. It's often hard to understand what to track, what to look for, and why any of it matters.	4-50.PDF
AC85	UX, Data & Analytics: the perfect match! Data gathering methods are totally accessible to anyone who wants to use them. Any company that wants to improve and move forward should think about implementing them. Statistics not only can guide you but give amazing results in your journey!	3-47.PDF
AC86	Colette Kolenda and Kristie Savage discuss how a three-step process for mixing qualitative and quantitative methods can avoid data discrepancies and fuel product decisions.	15-26.PDF
AC87	10 Top Tools for User Experience Testing	13-112.PDF
AC88	Audience data is quantitative data about the users of your website. The best-known and most easily-accessible repository of this can be found in Google Analytics (GA).	2-73.PDF
AC89	How to become a data-driven UX writer (and how to A/B test copy)	2-107.PDF
AC90	We use behavioral data to determine what amounts and types of interaction are happening on the site. This quantitative data, or WHAT data, coupled with the WHY data coming from UX and usability testing gives us a comprehensive view (or what I refer to as a 360 degree view) into website activity.	2-113.PDF

Table 55 – List of GLR approved articles (Part 2)

Article Citation	Short description	PDF link
AC91	Combining the behavioral UX and usability testing data to analyze results and make recommendations for website optimization is the final step of the 4 Big UX Optimization steps.	8-84.PDF
AC92	This approach of data-driven design helps to create a user-centric design and a better user experience. It enables you to make better design choices based on real evidence about the user's behavior, attitude, needs, etc.	7-5.PDF
AC93	the data soon adds up, and just as quickly, it becomes a valuable asset to those wanting to create personal digital experiences.	15-27.PDF
AC94	Rather than just guessing what your customers want or trying out new strategies to see what happens, you should be using a data-driven approach to the UX. This will help to eliminate time-consuming and expensive changes to allow your team to focus more on the solutions that work for your audience.	2-104.PDF
AC95	Big data provides essential insight into the buyer's persona and human behavior. Based on the analysis of big data, providers can stop guessing and start taking calculated measures to increase their product's success. Big data helps to make a design of a product data-driven rather than intuitive. And it results in better UX as customers get exactly what they wanted and even more.	7-99.PDF
AC96	Incorporating analytics into UX processes and research methods can add value in many ways. Project teams should use a combination of different data types to consistently measure data (in the wild) about usability and user behavior.	2-116.PDF
AC97	The Chrome User Experience Report provides user experience metrics for how real-world Chrome users experience popular destinations on the web.	2-20.PDF
AC98	It is no easy task to see through a whole bundle of user research data (quantitative and qualitative) and convert them into design ideas. Though we used 'n' number of methodologies to gather those data that are required for our design process, the idea of consolidating them is mind-blogging.	2-93.PDF
AC99	DATA-DRIVEN DESIGN: NEW "TREND" OR AN ETERNAL UX SOLUTION?	7-140.PDF
AC100	How Can Data Science Improve UX Design?	2-103.PDF
AC101	We, as product designers, need to have a great understanding of data in order to make informed decisions that will impact both on the business and the user experience (UX). Visualizing User Experience Data (Part I)	3-119.PDF
AC102	In our previous article 'Visualizing User Experience Data', we defined a framework to measure the User Experience of a product. In this article, we want to share how we can use Google Data Studio to visualise those metrics and facilitate decision making during the design process.	5-28.PDF
AC103	Data Analysis in UX Design - A prerequisite for a cross-functional product.	4-110.PDF
AC104	Five Critical Quantitative UX Concepts	1-61.PDF
AC105	Applying Machine Learning To User Research: 6 Machine Learning Methods To Yield User Experience Insights	9-51.PDF
AC106	What is Machine Learning + UX?	12-17.PDF
AC107	How to Capture UX Data for Experience-Driven Apps on Salesforce	13-6.PDF
AC108	UX research methods (part III): When to use quantitative data to justify product improvement decisions	1-15.PDF
AC109	User Research to Improve Your App Design	12-61.PDF
AC110	How to Interpret Data to Improve Your Website Design and Performance	13-71.PDF
AC111	Leveraging User Data for Improved UX	2-63.PDF
AC112	How to Use Data Analytics to Build Good UX	4-84.PDF
AC113	Partnering user experience designers with data scientists	15-8.PDF
AC114	A New Formula for Quantitative UX Decision Making	1-53.PDF
AC115	Just Add Water: Lessons Learned from Mixing Data Science and Design Research Methods to Improve Customer Service	15-39.PDF
AC116	In this guide, learn how to create your own custom CrUX Dashboard to track an origin's user experience trends.	2-75.PDF
AC117	How UX Designers Use Data In Design	2-101.PDF
AC118	Using Data to Drive UX & Design	7-154.PDF
AC119	In this chapter, we explain how to extract the data from your results, analyze it, and turn it into an action plan for improving your site.	4-63.PDF
AC120	UX analysis: best methods and key tools	1-50.PDF
AC121	In this article, we'll learn about data-driven design best practices and suggest a few tools to help you get started.	7-52.PDF
AC122	How designers can use data to create amazing work	15-116.PDF
AC123	How qualitative data fuels prototyping at NASA	7-103.PDF
AC124	Turn User Research Data Analysis into Actionable UX Insights in 5 Steps	4-32.PDF
AC125	6 Types of User Behavior to Track on Your Website & the Tools to Do It	6-61.PDF

Table 56 – List of GLR approved articles (Part 3)

Article Citation	Short description	PDF link
AC126	Making Data-Driven Decisions for Better Website UX	7-71.PDF
AC127	Data-Driven Design In The Real World	7-87.PDF
AC128	Analyzing User Behavior to Create the Best User Experience	4-86.PDF
AC129	If You're Not Using UX Data, It's Not UX Design	2-1.PDF
AC130	Strength in Numbers – An Overview of Data-driven Design	7-7.PDF
AC131	In this module, we'll go over everything you need to know about analyzing your data, using it to tell a meaningful story, crafting a deliverable that is most appropriate for your stakeholders, and best practices for creating the deliverables themselves.	1-58.PDF
AC132	Great UX leaders know their own UX data using scorecards	13-68.PDF
AC133	Why big data and UX need each other	2-8.PDF
AC134	This data collection and analysis process can be categorized into five main stages: discover, capture, analyze, plan, and execute. Find software tools for each of these five main stages and ensure each tool can be integrated into the same analysis stream. Here are the tools we use in these five stages.	14-101.PDF
AC135	Data-Driven Design for User Experience	7-3.PDF
AC136	Google's data-driven approach to superior user experience, revisited	7-83.PDF
AC137	Why Data Science and UX Design Need to Become Best Buds	15-2.PDF
AC138	Card sorting data analysis - Qualitative & quantitative ways	4-18.PDF
AC139	Data-Driven Design: An Integral Part of UX Design	2-3.PDF
AC140	Data-Informed Design: Three Data Stories	3-77.PDF
AC141	The Repertory Grid: Eliciting User Experience Comparisons in the Customer's Voice	9-108.PDF
AC142	UX101: Complete Guide to Building Data-Driven User Journeys	7-32.PDF
AC143	Three Ways UX Researchers And Analytics Specialists Can Collaborate And Deliver Great Insights	4-15.PDF
AC144	How to get UX data from non-UX research	14-2.PDF

Table 57 – List of GLR approved articles (Part 4)

Apenddix D

Figure 63 shows de homepage of the UX Data Wiki application and the options that users can explore. Figure 64 shows the page when the user already know what objective they want to achieve. Figure 65 shows the information (i.e., purposes, UX data and people that can help) about the selected objective.

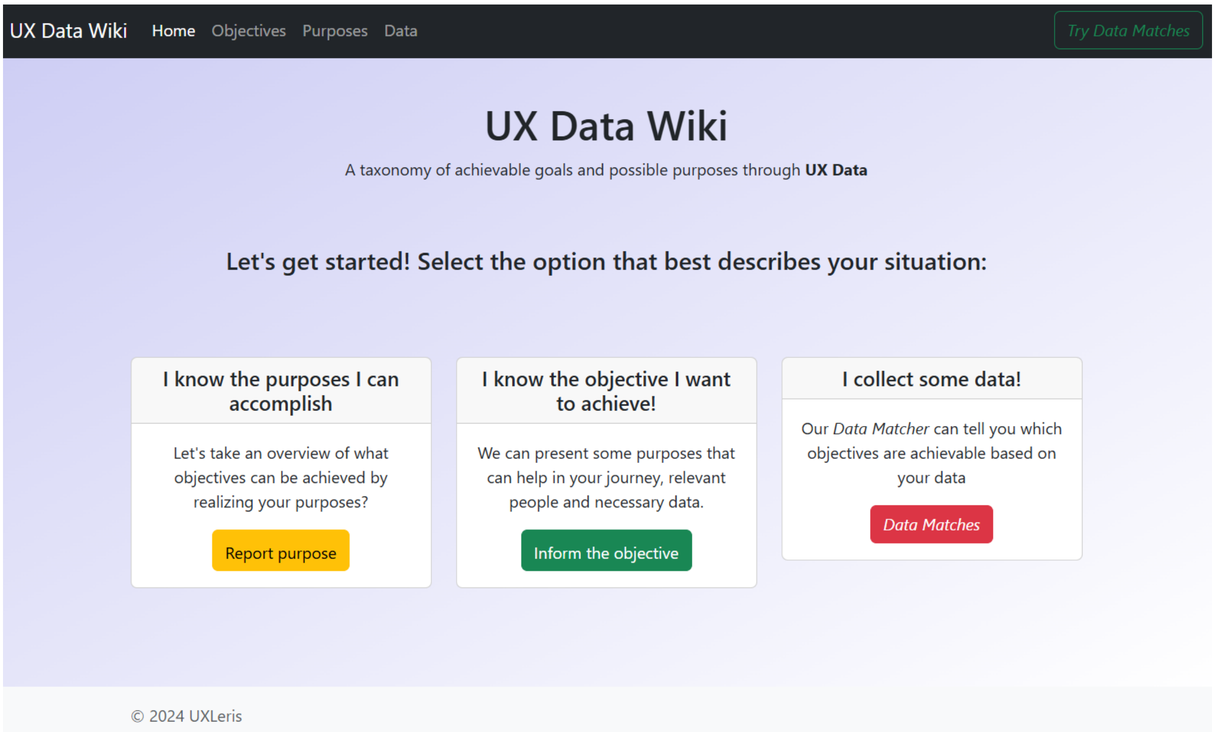


Figure 63 – Homepage of UX Data Wiki.

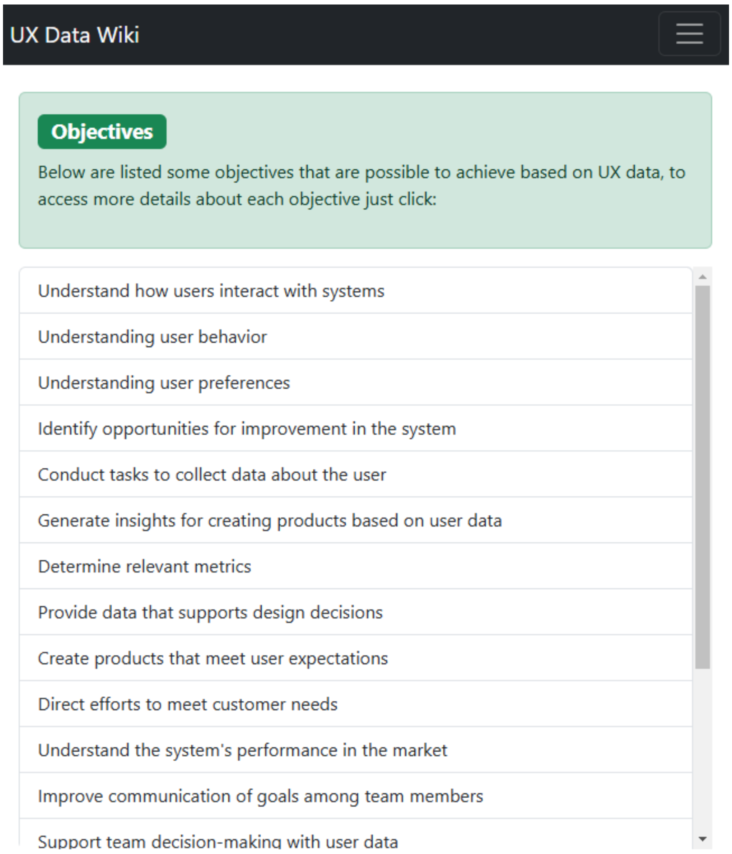


Figure 64 – Example of objectives available in UX Data Wiki.

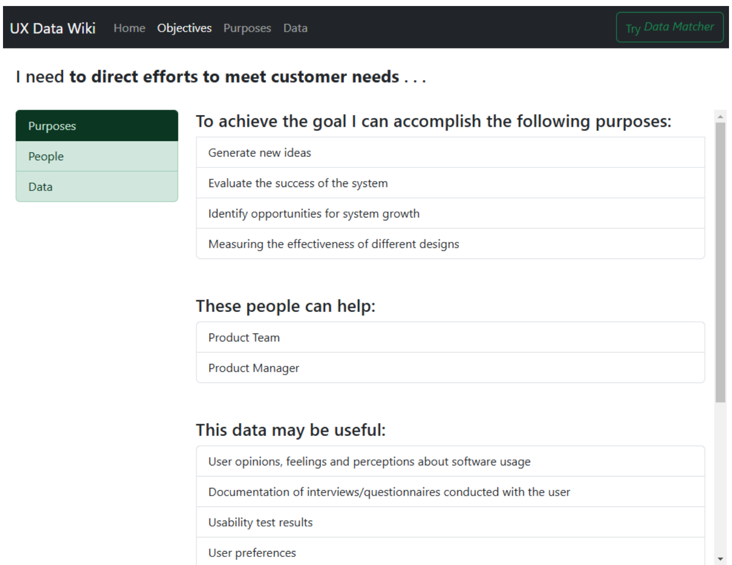


Figure 65 – Information showed when the user selected one objective.

Apenddix E

For grey articles quality assessment form see Figure 66 and Figure 67.

Article ID *

Sua resposta

Is an individual author? *

- ○ anonimous
- ○ 1 author
- ○ 2 or more authors

Has the author published other works? *

- ○ No
- ○ Yes
- ○ Yes, in the UX Data field

Does the work refers to a real situation *

- ○ No
- ○ Academic setup
- ○ Industry setup

Does the work have a bibliography? *

- ○ No
- ○ Mentioned in text
- ○ Have a list

Figure 66 – Questions used in the quality assessment of the grey articles (part 1).

Are the conclusions presented? *

- ◯ No
- ◯ Theoretical content
- ◯ Yes, lessons learned

Is there vested interest? (advertisiment or paid tests) *

- ◯ Yes
- ◯ No

What are the contributions?

- ☐ how to explore/analyze UX Data
- ☐ how to visualize UX Data
- ☐ how to collect/store UX Data
- ☐ how to communicate UX Data findings
- ☐ when we should explore/handle UX Data
- ☐ what we can extract of the UX Data
- ☐ what we consider UX Data (definitions)
- ☐ what UX Data can tell for us
- ☐ who manage UX Data
- ☐ informations about UX Data tools/approaches

Figure 67 – Questions used in the quality assessment of the grey articles (part 2).