CENTRIC 2024

The Seventeenth International Conference on Advances in Human oriented and Personalized Mechanisms, Technologies, and Services

ISBN: 978-1-68558-193-0

September 29 - October 03, 2024

Venice, Italy

CENTRIC 2024 Editors

Stephan Böhm, RheinMain University of Applied Sciences, Germany

Lasse Berntzen, University of South-Eastern Norway (USN), Norway

CENTRIC 2024

Forward

The Seventeenth International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services (CENTRIC 2024), held on September 29 – October 3, 2024 in Venice, Italy, addressed topics on human-oriented and personalized mechanisms, technologies, and services, commonly known as I-centric.

There is a cohort of technologies that favored the so called “user-centric” services and applications. While some of them reached some maturity, others are to prove their economics (WiMax, IPTV, RFID, etc). The human-oriented and personalized technologies and services rely on a key set of features, some to be deployed, others getting more mature (personal profiles, preferences, identity, proximity, personal devices, etc.). Following, advanced applications covering human related activities benefit from personalized and human-oriented networks and services, especially preventive and personalized medicine, body networks and devices, or anticipative systems.

The conference provided a forum where researchers were able to present recent research results and new research problems and directions related to them. The conference sought contributions presenting novel result and future research in all aspects of user-centric mechanisms, technologies, and services.

Similar to the previous editions, this event continued to be very competitive in its selection process and very well perceived by the international community. As such, it attracted excellent contributions and active participation from all over the world. We were very pleased to receive a large amount of top quality contributions.

We take here the opportunity to warmly thank all the members of the CENTRIC 2024 technical program committee as well as the numerous reviewers. The creation of such a broad and high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and efforts to contribute to the CENTRIC 2024. We truly believe that thanks to all these efforts, the final conference program consists of top quality contributions.

This event could also not have been a reality without the support of many individuals, organizations and sponsors. We also gratefully thank the members of the CENTRIC 2024 organizing committee for their help in handling the logistics and for their work that is making this professional meeting a success.

We hope the CENTRIC 2024 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in personalization research. We also hope that Venice provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city

CENTRIC 2024 Steering Committee

Stephan Böhm, RheinMain University of Applied Sciences - Wiesbaden, Germany
Tzung-Pei Hong, National University of Kaohsiung, Taiwan
Yasushi Kambayashi, Sanyo-Onoda City University, Japan

CENTRIC 2024 Publicity Chair

Lorena Parra Boronat, Universitat Politecnica de Valencia, Spain
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain
Jose Miguel Jimenez, Universitat Politecnica de Valencia, Spain
Francisco Javier Díaz Blasco, Universitat Politecnica de Valencia, Spain
Ali Ahmad, Universitat Politecnica de Valencia, Spain

CENTRIC 2024

Committee

CENTRIC 2024 Steering Committee

Stephan Böhm, RheinMain University of Applied Sciences - Wiesbaden, Germany
Tzung-Pei Hong, National University of Kaohsiung, Taiwan
Yasushi Kambayashi, Sanyo-Onoda City University, Japan

CENTRIC 2024 Publicity Chair

Lorena Parra Boronat, Universitat Politecnica de Valencia, Spain
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain
Jose Miguel Jimenez, Universitat Politecnica de Valencia, Spain
Francisco Javier Díaz Blasco, Universitat Politecnica de Valencia, Spain
Ali Ahmad, Universitat Politecnica de Valencia, Spain

CENTRIC 2024 Technical Program Committee

Khairuddin Ab Hamid, Waseda University, Japan
Youssef A. Attia, King Abdulaziz University, Saudi Arabia
Duygun Erol Barkana, Yeditepe University, Turkey
Ivon Bensason, Koç University, Turkey
Samir Brahim Belhaouari, Hamad Bin Khalifa University, Doha, Qatar
Lasse Berntzen, University of South-Eastern Norway, Norway
Stephan Böhm, RheinMain University of Applied Sciences - Wiesbaden, Germany
Daniel B.-W. Chen, Monash University, Australia
Sabine Coquillart, INRIA, France
António Correia, INESC TEC & University of Trás-os-Montes e Alto Douro, Vila Real, Portugal
Marco Costanzo, Università degli Studi della Campania "Luigi Vanvitelli", Italy
Carlos Cunha, Polytechnic Institute of Viseu, Portugal
Roberto Daza Garcia, Universidad Autónoma de Madrid, Spain
Rui Pedro Duarte, Polytechnic institute of Viseu, Portugal
Luciane Fadel, Federal University of Santa Catarina, Brazil
Rainer Falk, Siemens AG Corporate Technology, Germany
Filipe Fidalgo, Polytechnic Institute of Castelo Branco, Portugal
Alicia García-Holgado, GRIAL Research Group - University of Salamanca, Spain
Stefan Graser, CAEBUS Center for Advanced E-Business Studies | RheinMain University of Applied Sciences, Germany
Till Halbach, Norwegian Computing Center, Norway
Qiang He, Swinburne University of Technology, Australia
Koen Hindriks, Vrije Universiteit Amsterdam, Netherlands
Tzung-Pei Hong, National University of Kaohsiung, Taiwan
Takeshi Ikenaga, Kyushu Institute of Technology, Japan

Imène Jraidi, Advanced Technologies for Learning in Authentic Settings (ATLAS) Lab | McGill University, Montreal, Canada
Christos Kalloniatis, University of the Aegean, Greece
Yasushi Kambayashi, Sanyo-Onoda City University, Japan
Mazaher Kianpour, Norwegian University of Science and Technology (NTNU), Norway
Boris Kovalerchuk, Central Washington University, USA
Alexander Kröner, Technische Hochschule Nürnberg Georg Simon Ohm, Germany
Ravi Kuber, UMBC, USA
Stanislav Mamonov, Feliciano School of Business, USA
Ângela Cristina Marques de Oliveira, Instituto Politécnico de Castelo Branco, Portugal
Pierre-François Marteau, Université Bretagne Sud / Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA), France
Célia Martinie, Université Paul Sabatier Toulouse III, France
José Martins, LIAAD - INESC TEC / Polytechnic of Leiria, Portugal
Thomas Marx, TH Bingen, University of Applied Sciences, Germany
Erik Massarczyk, RheinMain University of Applied Sciences Wiesbaden Rüsselsheim, Germany
David Melhart, Institute of Digital Games | University of Malta, Malta
Pedro Merino, ITIS Software | University of Malaga, Spain
Toshiro Minami, Kyushu Institute of Information Sciences, Japan
Eduardo Miranda, University of Plymouth, UK
Fatma Najar, Concordia University, Montreal, Canada
Areolino Neto, Federal University of Maranhão, Brazil
Khoa Nguyen, Carleton University, Ottawa, Canada
Helder C. R. Oliveira, University of Calgary, Canada
Hugo Paredes, University of Trás-os-Montes e Alto Douro, Portugal
Monica Perusquía-Hernández, NTT Communication Science Laboratories, Japan
Stefan Pickl, Universität der Bundeswehr München, Germany
Melissa Ramos da Silva Oliveira, University of Vila Velha, Brazil
Jagat Jyoti Rath, Instituteof Infrastructure Technology Research and Management (IITRAM), Ahmedabad, India
Valentim Realinho, Instituto Politécnico de Portalegre, Portugal
Ann Reddipogu, RKode Ltd, UK
Michele Risi, University of Salerno, Italy
Armanda Rodrigues, NOVA LINCS | Universidade NOVA de Lisboa, Portugal
José Rouillard, University of Lille, France
Aurora Saibene, University of Milano - Bicocca, Italy
Sandra Sanchez-Gordon, Escuela Politécnica Nacional, Ecuador
Jungpil Shin, The University of Aizu, Japan
Alfredo Soeiro, University of Porto - FEUP, Portugal
Giandomenico Spezzano, National Research Council of Italy, Italy
Mu-Chun Su, National Central University, Taiwan
Patricia Torrijos Fincias, University of Salamanca, Spain
Carlos Travieso González, University of Las Palmas de Gran Canaria, Spain
Alberto Vergnano, University of Modena and Reggio Emilia, Italy
Christina Volioti, Aristotle University of Thessaloniki / University of Macedonia / International Hellenic University, Greece
Bo Yang, The University of Tokyo, Japan
Hao-Chun Yang, National Tsing Hua University, Taiwan

Christos Zaroliagis, University of Patras, Greece
Alejandro Zunino, ISISTAN-CONICET | Universidad Nacional del Centro (UNICEN), Argentina

Table of Contents

Enhancing UX Research Activities Using GenAI – Potential Applications and Challenges

Stefan Graser, Anastasia Snimshchikova, Stephan Böhm

{stefan.graser, stephan.boehm}@hs-rm.de
CAEBUS Center of Advanced E-Business Studies
RheinMain University of Applied Sciences
Wiesbaden, Germany

Martin Schrepp

martin.schrepp@sap.com
SAP SE
Walldorf, Germany

***Abstract*—User Experience (UX) Research covers various methods for gathering the users' subjective impressions of a product. For this, practitioners face different activities and tasks related to the research process. This includes processing a large amount of data based on qualitative and quantitative data. However, this can be very laborious in practice. Thus, the application of GenAI can support UX research activities. This paper provides a practical perspective on this topic. Based on previous studies, we present different use cases indicating the potential of GenAI in UX research. Moreover, we provide insights into an exploratory study using GenAI along an entire UX research process. Results show that Large Language Models (LLMs) are useful for various tasks. Thus, the research activities can be carried out more efficiently. However, the researcher should always review results to ensure quality. In summary, we want to express the potential of GenAI enhancing UX research.**

***Keywords*—User Experience (UX), UX Research, Generative Artificial Intelligence (GenAI), Usability Test, Surveys, Comment Analysis**

I. Introduction

A good UX fosters overall customer satisfaction and, in consequence, loyalty [1][2]. One of the main tasks of UX researchers is to evaluate the subjective impression of users towards products and to trigger improvements to ensure that they stay competitive over time. Users' subjective impressions can be broken down into three phases: (1) the anticipated use, (2) the actual use, and (3) after use. More precisely, it refers to the expectations before the use, the experience during, and the experience after using a product [3]. For this, different evaluation methods are applied in practical research.

A popular method to evaluate specific aspects of UX related to system or product quality is usability testing. Different testing procedures are available [4][5], but the general idea is similar. Test tasks are formulated, and participants are observed while working on them. Usually, participants are instructed to comment on what they do or want to achieve (thinking aloud). The researcher takes notes, or the comments are recorded. As a result of such a usability test, typically, an analysis of the comments of the participants together with some metrics (normally task completion rate and task completion time) is reported.

Usability tests provide detailed insights into problems and potential product improvements. But they require a lot of effort. For example, the preparation of good test scenarios and instructions for participants, the analysis of a large number of comments per participant, and the preparation of the insights for the test report. Thus, typically, only a small number of testers can be recruited. A survey with experts that conducted formative usability tests [6] found that 82% invited less than 15 testers and the overall median of test participants was 10 (similar results were reported in [7]).

Another quite popular method is the usage of surveys [8]. Compared to usability tests, it is very easy to recruit larger samples of participants in such studies. A survey can be sent by email to a large group of customers, posted on social media channels, or incorporated by a feedback button into the product's user interface.

UX and Usability surveys typically collect subjective impressions of users that are a mix of structured and unstructured data. Examples of structured data are demographic data (e.g., age, gender, profession), usage data (e.g., frequency of use, experience with the product), and scores from standardized UX and Usability questionnaires (e.g., SUS [9], UEQ [10], UMUX [11], or VISAWI [12]). Structured data are easy to analyze with standard statistical methods. Demographic or usage data often help to interpret the results of UX scores from questionnaires, for example, a higher frequency of use is often associated with higher UX ratings [13][14].

It is a common practice to allow participants to comment on the product by one or several text entry fields. Such comments create additional insights into why participants have answered a certain way and can often contain ideas for improvements. The collected comments are usually much shorter than transcripts from a usability test, and only a fraction of the participants fill in comments [15]. But their number can still be huge. If you get 10,000 responses and just 10% of them fill in some text, you still have 1,000 texts to analyze. This makes it clear that anything that helps to avoid a purely manual interpretation and analysis of comments is highly welcome.

In practice, it is common to conduct a mixed-method

 ISBN: 978-1-68558-193-0

approach to gather comprehensive insights into the users' subjective impressions. Therefore, different methods, such as usability testing and surveys, are applied together. Thus, research activities require the realization of different tasks, such as the creation of textual artifacts for the preparation and execution of the activity. Examples include instructions or task descriptions for usability tests, personas resulting from user research, or the interpretation of users' comments. Moreover, both qualitative and quantitative evaluation results are collected. This textual data must be processed. LLMs can naturally support tasks, such as creating and interpreting textual data. Thus, it is interesting to know how they can be used in typical user research activities to save effort.

This paper is structured as follows: Section II explains the specific objective of the research, including the research questions. Afterward, the paper structure is aligned with the four research questions. Section III, IV, and V provide insights into applying GenAI in UX research referring to RQ 1 - 3. Section VI presents a practical approach applying different methods supported by GenAI. We summarize answers to the research questions in the discussion in Section VII. This article ends with a conclusion and outlook in Section VIII.

II. RESEARCH OBJECTIVE AND RESEARCH QUESTIONS

The rapid development of GenAI techniques offers many possibilities to support UX research activities. Of course, it is not possible to oversee all the possibilities and risks associated with this new technology at the moment. However, some promising application areas can already be identified and are discussed in this paper. Against this background, we specifically address the following research questions:

- ***RQ1: How can GenAI assist UX researchers in categorizing and summarizing user comments?***
  We investigate the potential of ChatGPT to analyze user comments from surveys and transcripts from usability tests using existing survey data from a larger study.
- ***RQ2: How can GenAI be used to assist UX researchers in the preparation of research and design activities?***
  Several UX activities require the preparation of instructions for the participants (e.g., task descriptions for usability tests) or some artifacts used for design activities (e.g., personas). We investigate how GenAI can support UX researchers in those tasks.
- ***RQ3: Can GenAI be used to develop a deeper understanding of the concept of UX?***
  We investigate how LLMs can be used to analyze the semantic textual similarity of items and scales from UX questionnaires and how these results can be used to plan UX evaluations.
- ***RQ4: How can GenAI be integrated with existing methods in a typical UX evaluation scenario?***
  We describe the results of a UX evaluation based on a mixed-method approach that evaluates UX before (expectations concerning UX), during (analyzing Thinking-aloud comments from a product testing), and after (UEQ [10] results) use.

III. GENERATIVE AI-BASED COMMENT ANALYSIS

There are several established methods to analyze and communicate user comments. A semantic grouping of comments into categories, together with the number of comments per category, can be communicated. Alternatively, a short textual summary of the main points mentioned by the users can be created. Other common methods are sentiment analysis [16] or word clouds [17][18]. LLMs can be used to assist UX researchers in all these tasks.

We will discuss the first two methods (sentiment analysis with LLMs is an already established method, and the creation of word clouds does not really need GenAI capabilities) using an example of a survey that collected feedback concerning PayPal. For details of the study, see [19]. The comments are answers to the questions "What do you particularly like about PayPal?" (positive comments) and "Which aspects do you find particularly bad about PayPal?" (negative comments). Overall there were 113 negative comments and 105 positive comments.

A ChatGPT (version 4.0) prompt consisting of general instructions followed by a numbered list of all positive (negative) comments from the survey was used. The instruction had the following content:

The numbered list below contains user comments concerning PayPal. Ignore negative comments. Categorize the comments. One comment can belong to more than one category. Create a short name for each category that contains less than two words. Create a summary for each category in a very short sentence. List all categories. Show the short name of the category first. Then, in a new line, the summary of the category. Then the list of comments belonging to that category. Combine the short summaries from the 5 most frequent categories in a short text.

Participants entered negative comments in some cases to the question "What do you particularly like about PayPal?" These are filtered out by the "Ignore negative comments" statement. To analyze the negative comments, this part was replaced by "Ignore positive comments."

The prompt splits the task of creating a classification and a summary into a simple procedure. Firstly, categories are determined and described. Since the categories and the assigned statements are listed in the result, a manual quality check of the categorization is easy. Secondly, the short statements are combined into a summary. Of course, the number of categories is not known beforehand. However, this number can be determined by simply running the prompt several times with different values for this number, which was, in fact,

Copyright (c) IARIA, 2024. ISBN: 978-1-68558-193-0

used to come to our decision to use 5 categories. The output indicates if there are not enough comments per category or if the categories are too abstract or do not clearly separate between comments.

In our example, the following five categories and short texts were produced for the positive comments:

- Ease of Use: Users appreciate the simplicity and intuitiveness of PayPal. (53 comments)
- Speed: Users value the quickness of transactions with PayPal. (21 comments)
- Security: Users feel secure using PayPal for their transactions. (13 comments)
- Integration: Users like how PayPal integrates with other platforms and services. (11 comments)
- International/Currency Features: Users appreciate the ability to handle international transactions and currency conversions. (7 comments)

For the negative comments, the result is as follows:

- Security Concerns: Users have concerns about the security of PayPal. (17 comments)
- Difficulties with Transfers and Withdrawals: Users find it difficult and time-consuming to transfer or withdraw money. (14 comments)
- High Fees: Users are unhappy with the high fees charged by PayPal. (9 comments)
- Poor Customer Service: Users find PayPal's customer service to be lacking and unhelpful. (8 comments)
- User Interface Issues: Users find PayPal's user interface confusing and difficult to navigate. (4 comments)

As a positive summary, the following text was created: *Users appreciate PayPal for its ease of use, quick transaction speed, security features, integration with other platforms, and its ability to handle international transactions and currency conversions* and the negative summary was *Users have expressed dissatisfaction with PayPal due to high fees, poor customer service, difficulties with transfers and withdrawals, issues with the user interface, and concerns about security. These areas represent the most frequent categories of negative feedback about the platform.*

Since ChatGPT is not deterministic, it must be checked how reliable such a generated categorization or summary is. If the same prompt is used twice, the results will usually differ to some extent. However, since the effort to run the same prompt several times is minimal, this can be checked easily. In our example, repetitive runs produced only small deviations, for example, "Users feel secure using PayPal for their transactions" versus "Users appreciate the security features of PayPal, including protection of personal details" or "Users find the ability to handle and convert different currencies useful" versus "Users find PayPal's ability to handle, convert, and transfer different currencies beneficial".

Another question is how well the result describes the content of the comments. Of course, there is no "optimal" solution to such tasks. Different human experts will also produce slightly deviating results. The validity of the result can be checked by manually analyzing a small subset of comments and comparing them to the classification of ChatGPT. To get a deeper understanding, we did (which is unrealistic for real use cases with thousands of comments) a full analysis of all comments by a human expert. The classification was identical in 88% of the cases.

Comments differ concerning content and length between different data collection mechanisms. If, for example, a product survey is sent to users via an email campaign, then users answer the questions in retrospect. In such cases, comments are often relatively short and concentrate on the product's general strengths and weaknesses. If a survey is opened over a feedback link inside the product, users answer the questions in their actual working context. Many comments will refer to the screen on which the feedback button is clicked and will be quite concrete (refer to a UI element or feature on the screen). Since the users assume that it is clear what they are referring to, the screen itself is typically not mentioned in the comment. Such comments are hard to interpret if the context is not known. Thus, in such cases, a human UX expert who knows the application well obviously has a massive advantage over an LLM. In each case, a more detailed analysis of the comments is required, for example, a pre-processing by analyzing comments collected for separate screens or adding some context information to the prompt.

In our example, a common product (PayPal) was investigated. Therefore, it was not necessary to provide additional context information; the product name was sufficient. In practical applications, the LLM will typically not have any information about the product (it will not be in the training set). It is recommended to add several sentences to the prompt that explain the product's main use cases.

IV. Creation of Personas

Another area where LLMs can be quite helpful in UX research is the creation of texts used for research actions (we will show in Section VI how task descriptions for usability tests can be created) or for the communication of research results. In this section, we will show how the creation of personas can be supported by LLMs.

Personas [20][21] are fictional characters that represent typical target groups of users. They are created based on research results concerning the actual or potential users. Personas are merely a communication tool to help designers or developers empathize with users and consider their needs adequately. It is a matter of taste which information a persona description contains. There are several templates available that contain a pre-selection of categories and a way to visualize the persona description. Typical examples are demographics (name, age, gender, family status, etc.), goals and actions associated with the product (user journey, typical tasks, usage scenarios, etc.), skills (education, degrees, knowledge concerning product or work area, etc.), interests (for example, hobbies), personal

Copyright (c) IARIA, 2024. ISBN: 978-1-68558-193-0

goals or values (lifestyle, environment, etc.) and often a photo. Most of this information is not relevant to the product design decisions. They are added to make the persona ”real” [20]. The underlying assumption is that it is easier to discuss if a product fulfills the needs of a real person (described by the persona) than an abstract ”user”, where every member of a design team may have different ideas about ([20]).

The goal of this paper is not to discuss the pros and cons of personas in design. They are a frequently used tool, and we show how easily LLMs can generate realistic persona descriptions. As described above, the core of a persona is research about the user group. We take as an example a developer of cloud applications that are used to enhance an existing ERP system in a company.

The following prompt contains the core information about such a developer role: *In UX research, personas are often used in user interface design to communicate and discuss design decisions. Create a persona for a software developer. The person works for a medium-sized company and is a member of a small team of software developers. The company uses SAP as its major ERP system. The development team creates cloud applications that enhance this ERP system with additional functionalities. Some key challenges in this job are the technical complexity of the applications and the need to get a deep knowledge of the structure and existing interfaces of the ERP system. Can you create a persona description? Please provide, in addition, a quote that could realistically come from the persona, which sums up what matters most to them in relation to the product or service.*

The resulting persona was described as follows:

- Persona Name: John Smith
- Background: John is a 35-year-old software developer who has been working in a medium-sized company for the last 7 years. He holds a Bachelor’s degree in Computer Science and has a strong background in cloud computing and ERP systems.
- Role: John is a member of a small team of software developers. His main responsibility is to develop cloud applications that enhance the company’s major ERP system, SAP, by adding additional functionalities.
- Key Challenges: John often struggles with the technical complexity of the applications he is developing. He also finds it challenging to gain a deep understanding of the structure and existing interfaces of the ERP system.
- Goals: John’s primary goal is to develop robust, efficient, and user-friendly applications that can seamlessly integrate with the existing ERP system. He also aims to continuously improve his technical skills and understanding of the ERP system to overcome his challenges.
- Skills: John is proficient in several programming languages, including Java, Python, and C++. He also has a good understanding of cloud computing platforms, especially AWS and Google Cloud. He has a basic understanding of SAP but is still working on improving his knowledge in this area.
- Behaviors: John is a problem-solver who enjoys tackling complex technical challenges. He is a quick learner and is always eager to learn new technologies and tools that can help him in his job. He prefers to work in a collaborative environment and values clear communication within his team.
- Needs: John needs clear documentation and training materials to help him understand the ERP system better. He also needs a supportive work environment where he can freely discuss his challenges and get help from his team members.
- Quote: ”Understanding the ERP system is just as important as writing good code. We need to ensure that our applications not only work well but also integrate seamlessly with the existing system. It’s a challenge, but it’s also what makes our job interesting and rewarding.”

There are some remarkable points concerning this persona description. It contains several elements that are not explicitly mentioned in the prompt but inferred from this information. For example, ”programming skills in Java, Python, and C++”, or ”understanding of AWS and Google Cloud”. Also, the age (35), working experience (7 years), and degree (bachelor) fit together. The description under ”Behaviors” fits well for a young motivated developer. Overall, the generated description matches the small amount of information contained in the prompt quite well and is a realistic (a bit stereotypical) description of a person.

It was already noted that personas often contain stereotypes [22]. This is true for personas created by UX researchers and even more true for ChatGPT-created personas. If we run, for example, the prompt described above 10 times, then all 10 personas are male and between 32 and 35 years old. If this is really a problem, it is a matter of taste and assumptions on how personas should be used in a design process. Stereotypical personas are, on the one hand, more realistic [20]; on the other hand, [23], they can cause wrong conclusions and predictions in the design team. However, at least concerning demographics contained in a ChatGPT-created persona, this can easily be avoided by explicitly defining this information in the prompt.

V. Analyzing Semantic Similarity of Items and Scales

This section shows how LLMs can be used to get deeper insights into semantic similarities of UX items. We focus on the application aspects. More details concerning the methods can be found in [24][25].

The most common way of measuring UX is the usage of standardized questionnaires in surveys [26][27]. UX has many facets, thus a single questionnaire can not cover the whole concept [28]. Many different UX questionnaires are available, and each one focuses on different UX aspects. Thus, it depends on the evaluation scenario which questionnaire is suitable [29][30].

A UX questionnaire consists of different items and scales [29][30]. However, as the example items in Table I show, nearly identical items can be assigned to differently named scales. Conversely, scales with highly similar or even identical names can measure semantically different concepts. For example, AttrakDiff [31] and UEQ [10] both contain a scale named *Stimulation*. However, *Stimulation* in the sense of the UEQ refers to an interesting and stimulating experience. *Stimulation* in the sense of the AttrakDiff2 contains, in addition, the aspect that the design of the product is creative and innovative [8].

TABLE I. EXAMPLE OF SIMILAR ITEMS ASSIGNED TO DIFFERENTLY NAMED SCALES [30].

Item	Scale	Questionnaire	Source
The system is easy to use	Likeability	SASSI	[32]
I thought the system was easy to use	Usability	SUS	[9]
This system is easy to use	Overall	UMUX	[11]
It was simple to use this system	System Usefulness	PSSQU	[33]

Previous research concerning the dependency of UX scales mainly focused on an analysis of correlations between scales [29][30][34]. Other studies analyze the semantic textual similarity by applying NLP techniques [35][36], i.e., Sentence Transformers, to analyze the similarity of the encoded textual items in a vector space.

LLMs are good candidates for performing an analysis of semantic similarity between items and scales of UX questionnaires. They use word embeddings to represent the semantics of texts. Against this background, we applied ChatGPT-4 to (1) (re-) construct common UX concepts based on a set of UX items (see V-A), (2) detect and match suitable measurement items based on semantic textual similarity (see V-B), and (3) uncover the semantic textual similarity among the measurement items (see V-C).

A. Detecting a Semantic Structure on Items

We created an item set containing 408 items from 19 UX questionnaires (for construction details, see [24][25]). A series of six ChatGPT prompts are formulated and sequentially applied that ask ChatGPT to classify the items into categories and more fine granular sub-categories (see [24][25] for the detailed formulation of the prompts). As a result, the LLM generated six main topics and 16 subtopics:

- **System Usability and Performance**: Ease of Use — Efficiency and Speed — Functionality and Flexibility
- **User Engagement and Experience**: Engagement Level — Aesthetics and Design — Confusion and Difficulty
- **Information and Content**: Clarity and Understandability — Relevance and Utility — Consistency and Integration
- **Website-specific Feedback**: Navigation and Usability — Trust and Security — Aesthetics and Design
- **Learning and Adaptability**: Learning Curve — Adaptability
- **Overall Satisfaction and Recommendation**: Satisfaction — Recommendation

The full classification can be found in [24][25]. As an example, we show the top five items of the sub-category *Efficiency and Speed*:

1) The interaction with the system is fast.
2) The system responds too slowly.
3) This software responds too slowly to inputs.
4) The speed of this software is fast enough.
5) Has fast navigation to pages.

The generated topics refer to both pragmatic and hedonic properties, but they are rather broad and sometimes look strange from the perspective of a human UX expert. For example, the subtopic *Navigation and Usability* referring to a pragmatic value and *Aesthetics and Design* referring to a hedonic value are summarized under the main topic *Website-specific Feedback*. Most of the generated sub-topics can be easily related to established UX concepts, for example, a larger set of UX aspects used in several studies [30].

In another prompt, we inserted the UX quality aspects from [30] and asked ChatGPT to compare them with its own generated sub-topics. The comparison is shown in Table II:

TABLE II. COMPARISON OF EXISTING UX QUALITY ASPECTS BY [30] AND AI-GENERATED TOPICS.

(#)	UX Quality Aspects	AI-generated Sub-Topics
(1)	Perspicuity	Ease of Use — Learning Curve
(2)	Efficiency	Efficiency and Speed
(3)	Dependability	Consistency and Integration
(4)	Usefulness	Functionality and Flexibility—Relevance and Utility
(5)	Intuitive use	Ease of Use
(6)	Adaptability	Adaptability
(7)	Novelty	-
(8)	Stimulation	Engagement Level
(9)	Clarity	Clarity and Understandability
(10)	Quality of Content	Relevance and Utility
(11)	Immersion	Engagement Level
(12)	Aesthetics	Aesthetics and Design
(13)	Identity	-
(14)	Loyalty	Loyalty
(15)	Trust	Trust and Security
(16)	Value	Perceived value

Results show that most AI-generated sub-topics are named differently but can be allocated to established UX quality aspects. *Novelty* and *Identity* were not classified, but this was caused by the fact that most of the questionnaires in our list did not contain corresponding items.

To sum up, ChatGPT-4 generates a comprehensive overview of topics and subtopics based on UX measurement items. Moreover, both pragmatic and hedonic topics are contained. The different items almost completely match the respective topic. Thus, LLMs are useful in generating topics and thus can be used to investigate the semantic structure of sets of items.

B. Finding Suitable Items for Ad Hoc Surveys

Standardized UX questionnaires offer many advantages compared to ad hoc defined surveys [8]. However, it is also not uncommon to use surveys in an exploratory way to get insights into the opinions of users. Such ad hoc surveys often contain open-ended questions and some closed questions to get some rough scores concerning single UX aspects. Often single questions from standardized questionnaires are reused for this purpose. In this section, we show how LLMs can be used to determine suitable items.

ChatGPT-4 was applied to the data set of 408 items described above. The following exemplary prompt was used to find items that describe the aspect of *Usefulness* (can be adapted to other UX qualities by changing the bold part):

Below there is a list of statements and questions related to the UX of a software system. ***Select all statements or questions from this list that describe whether the software system is useful or not.*** *List these statements or questions. Start with those statements and questions that describe this best.*

As a result, ChatGPT identified 15 items. The top 5 are shown below.

1) The software helps me to complete my work tasks better than expected without extra effort.
2) With the software, I can sometimes even exceed my desired goals without any extra effort.
3) The software allows me to increase the quality of my work without any extra work.
4) The software offers me all the possibilities I need to work on my tasks.
5) The software is tailored to the tasks I need to work on.

The selected items fit pretty well with the intention formulated in the prompt. Thus, if an adequate list of items is available, it is easy to filter out candidates who match a given intention. For details and adjustments of the described prompt to other UX aspects, see [25]. This paper shows that, with a few exceptions, the selected items aligned well with the UX quality specified in the prompt. However, in this use case the intention is to detect one or a few items that describe a certain aspect of UX, thus researchers will pick items from the top of the list. Thus, ChatGPT can be used to select items from a candidate list that fit a certain research objective.

C. Investigating the Semantic Similarity between UX Concepts

In this section, we want to provide an insight into how effective GenAI is in measuring semantic similarity. Previous research has shown that innovative NLP techniques can be applied to compare the semantic similarity based on the encoded textual items in a vector space [35][36]. As GenAI is a sub-field of NLP, the semantic textual similarity analysis was conducted by applying ChatGPT-4.

A second item pool was used for this study. A list of 40 UX questionnaires [28] was analyzed, and artificial items with a highly standardized format were created. All positive adjectives from both existing semantic differential scales and statements in UX questionnaires were extracted. Based on the positive adjective, items were generated applying the same structure: *"I perceive the product as ¡adjective¿"*. This results in a data set of 135 artificially generated items. We refer to [8][37] for a similar technique.

We use ChatGPT to investigate how the 135 artificial items relate to existing UX concepts. Therefore, we applied a generic prompt with an instruction and an explanation regarding common UX concepts/quality aspects. In the following, an exemplary prompt concerning *Usefulness* is illustrated:

Below, there is a list of statements related to the user experience of a product. ***Select all statements from this list that describe that users perceive the product as useful.*** *List these statements or questions. Start with those statements and questions that describe this best.*

After the prompt, we inserted the artificially generated item pool. The part of the prompt marked in **bold** was adapted to the respective UX quality aspect. The remaining part of the prompt stayed stable. In summary, prompts for the UX quality aspects *Learnability*, *Efficiency*, *Dependability*, *Stimulation*, *Novelty*, *Aesthetics*, *Adaptability*, *Usefulness*, *Value*, *Trust*, and *Clarity* were applied. The respective prompt for each quality aspect runs three times. Only the items assigned to the respective UX quality aspect in all three runs were included. This results in an overview with semantically similar adjectives representing the respective UX quality aspects. For detailed results, we refer to [25].

Such an analysis can provide insights into the semantic overlap of common UX concepts. We see, for example, that *Novelty* and *Stimulation* share several assigned adjectives, thus there is some semantic overlap. This corresponds well to the realization of these concepts in UX questionnaires. In the AttrakDiff [31], these aspects are combined in one scale, while they form two separate scales in the UEQ [10]. There is also the expected close connection between the pragmatic qualities *Learnability*, *Efficiency* and *Dependability*. We can also see that *Clarity* and *Value* are somewhere in between the pragmatic qualities and *Aesthetics*, which is also known from empirical research [38][39]. However, many more dependencies become visible in the graphic. Thus, such an analysis can provide insights into otherwise hidden semantic overlaps of common UX concepts.

An analysis, as described above, can help UX researchers develop a deeper understanding of common UX scales and concepts and their semantic overlap. It can also help interpret the results of UX surveys that use corresponding UX scales for measurement.

VI. Exploratory Study for the Investigation of a Generative AI-supported UX Evaluation Scenario

The previous sections showed the potential of GenAI for different tasks in UX research. This section will provide a practical example of how GenAI can be applied in a research

 ISBN: 978-1-68558-193-0

process. We conducted an exploratory study based on a mixed-method approach, gathering qualitative and quantitative data. The study aimed to determine the extent to which GenAI is a useful enhancement in the practical UX research process. In the following, the study approach is described.

A. Study Approach

The study is designed to present an entire UX research process. All relevant aspects of UX are therefore intended to be examined. This results in three perspectives:

1) **Anticipated Use**: UX impressions and expectations towards a product before the use [40]
2) **Actual Use**: The momentary experience during the interaction with the product.
3) **After Use**: The retrospective experience after the usage of the product.

Thus, the methodological approach is broken down into these three parts. It is important to note that all three parts of this study are based on the theoretical foundation of the UEQ concerning the construct of UX [10]. Thus, the common ground is the six UX scales and their respective items. For details, we refer to [10].

We applied the UEEE method developed by [40] regarding part ***(1)*** **Anticipated Use**. Moreover, we conducted a usability test concerning part ***(2)*** **Experience during Use**. In particular, the Thinking-Aloud method based on AI-generated user tasks was performed. This results in a textual data pool of 115 pages and approximately 43.750 words. Afterward, the resulting user protocols were analyzed using ChatGPT-4. In relation to part ***(3)*** **Actual Use**, we applied the User Experience Questionnaire (UEQ) by [10]. The approach is visualized in Figure 1.

The social media platform **Instagram** was analyzed as an evaluation object. We conducted a convenience sample with a total of 30 study participants. The study was conducted in German. In the following, the three performed parts are further described.

1) User Expectations: User Experience Expectation and Evaluation (UEEE) [40][41] is a method to evaluate user expectations uniformly and efficiently. The method is based on the same idea as card sorting methods, such as *Product Reaction Cards* [42]. The researchers define a set of adjectives describing the respective product. These adjectives represent the expectations towards the product. The adjectives are displayed to the study participants in a tool described in [40]. The study participants are asked to evaluate these adjectives based on their expectations concerning the product. Therefore, the adjective can be classified into four categories (*unimportant, rather unimportant, rather important, important*). For this study, we used the positive adjectives of the 26 items of the UEQ.

Regarding the evaluation, study participants received a brief introduction and explanation of the tool and the evaluation object. Afterward, they had to classify the adjectives based on their expectations of using Instagram.

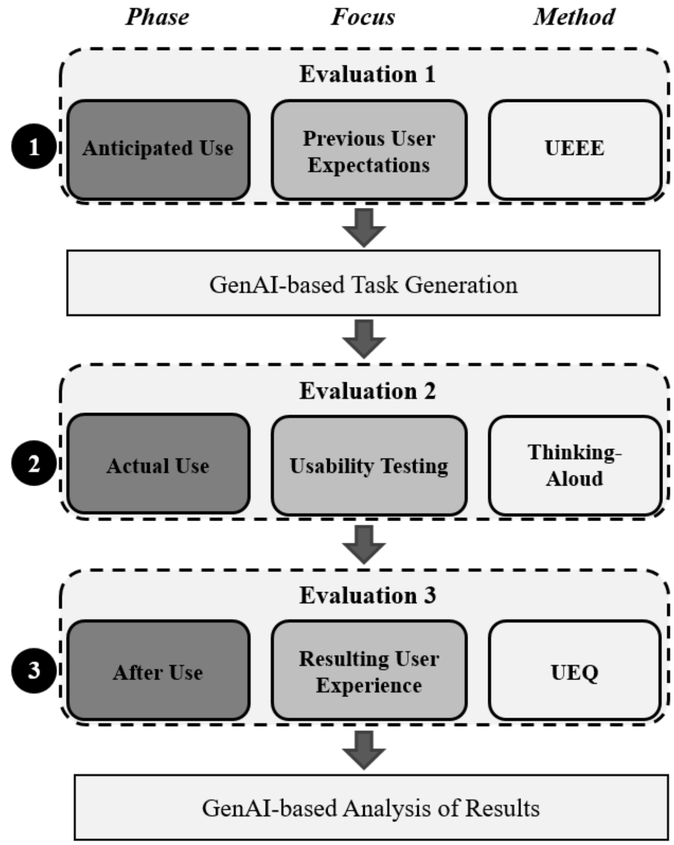

Figure 1. Methodological Approach of the Exploratory Study.

2) Generative AI-based Task Generation: We used ChatGPT-4 to generate tasks concerning usability testing. We applied four prompts to specify the tasks for usability testing. Each task refers to one UX scale regarding the UEQ [10]. Thus, this ensures that the user protocols regarding the Thinking-Aloud relate to the relevant UX factors. The four prompts and the final tasks are provided in Appendix A and B.

ChatGPT's ability to generate tasks results from the fact that Instagram is a well-known product. In the training phase of the LLM, there were obviously many texts concerning the usage of Instagram. In practice, usability tests are typically performed for new products, and such information is not available for the LLM. However, existing product documentation, specifications, or other sources of information can be added to the prompt in such cases to provide the required information.

3) Actual Use: To measure the experience during use, we conducted usability testing. More precisely, we deployed Thinking-Aloud. For this, the study participants were introduced to the UX test scenario. Moreover, they received the six AI-generated tasks to complete using Instagram.

4) User Experience: In the third part, we evaluated the actual UX. Therefore, various methods for measuring UX can be found in the scientific literature. The most established way is the use of standardized questionnaires.[26]. We implemented

Figure 2. Word Cloud Visualizing the Expectations of Participants.

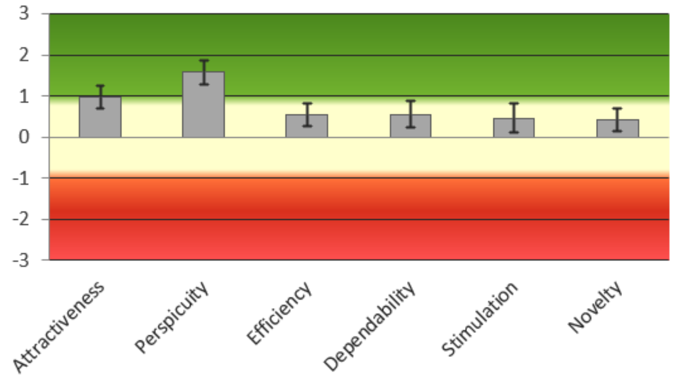

Figure 3. Scale Scores and Confidence Intervals for the UEQ.

the UEQ by [10] to examine the UX.

The results are shown below. Due to paper restrictions, we focus on the relevant results regarding GenAI.

B. Results

1) Anticipated Use – UEEE Results: Figure 2 shows a word cloud that visualizes how often an adjective was assigned to the category *important*. The bigger an item is displayed, the more participants consider it important for the product. The word cloud shows that users have high expectations concerning security, perspicuity (easy to learn, clear, understandable), and stimulation (interesting, exciting).

2) Actual Use: As shown in Section V-A based on [24][25], GenAI is useful in (re-)constructing UX factors. Thus, we further applied ChatGPT-4 to identify UX factors by analyzing the user protocols conducted with the Thinking-Aloud method. In particular, we consolidated the qualitative textual data based on the different tasks. This resulted in six consolidated protocols. We used ChatGPT-4 to analyze data for each task to identify key topics concerning the UX. The identified topics and descriptions are shown in the following:

- **Ease of Use and Intuitiveness:** Many users found Instagram's story creation features to be intuitive and easy to navigate. Features like adding location tags, filters, hashtags, and polls were generally considered user-friendly.
- **Feature Discovery:** Some users indicated challenges in discovering specific features or functions within the app, suggesting that while the app is generally easy to use, certain functionalities may not be immediately obvious to all users.
- **Personalization and Creativity:** The availability of various filters, stickers, and the ability to add music was positively received, offering users creative ways to personalize their stories.
- **Information Clarity:** There were mixed responses about the clarity of information presented in the app. Some users found it straightforward to understand and use different features, while others struggled with specific aspects like changing colors or understanding the purpose of certain icons.
- **Learning Curve and User Guidance:** New users or those less familiar with the app's functionalities experienced a learning curve. However, once familiarized, they found the app easier to navigate.
- **Functionality and Performance:** Overall, the app's functionalities were well-received. However, some users suggested improvements, like better filter variety or more options for customizing location stickers.
- **Accessibility and User Experience Variability:** Different users had varying experiences based on their familiarity with Instagram, indicating that user experience can vary significantly depending on the user's prior exposure to the app.

3) After Use – UEQ Results: Figure 3 shows the results of the UEQ. As we can see, there is a high rating for the scale *Perspicuity* and also a positive rating for the scale *Attractiveness*. The other scales show only moderate positive ratings. Of course, the confidence intervals are relatively big, based on the fact that we had only 30 participants. Thus, we should interpret the result with care. But if we compare this to the expectations evaluated before, we see that the expectations concerning *Perspicuity* are fulfilled, while this is not the case for *Stimulation*.

4) Matching Qualitative and Quantitative Evaluation Results: In the last step, we applied ChatGPT-4 to match qualitative and quantitative evaluation results. In particular, we aimed to match the textual data of the user protocols with the quantitative results of the UEQ and the UEEE. Thus, we provide a combination of quantitative scale values and qualitative text content. Two perspectives were provided for this.

1) **Quantitative Data Perspective**
2) **Qualitative Data Perspective**

Concerning (1) **Quantitative Data Perspective**, we aimed to identify the number of relevant statements in the textual data (user protocols) related to the respective items. In relation to (2) **Qualitative Data Perspective**, we aimed to classify relevant statements to the different items. Thus, each scale and its items consist of relevant textual data. We analyzed the textual data regarding the UX factors *Attractiveness* and *Stimulation*. We applied five prompts to ChatGPT-4. Each prompt was adjusted for the respective UX factor. In the following, we present the respective prompts and results. Due to paper restrictions, we conducted results referring to the UX factor *Attractiveness* and *Stimulation*.

The first prompt is as follows: ***Prompt1:*** *We performed usability testing regarding Instagram. For this, we generated a task. The task refers to the UX factor Attractiveness and its positive items.* ***Inserted definition and items****. The attached user protocol file contains 30 answers given to the task. As the tasks relate to the specific factor of Attractiveness and its items, it should be possible to match the textual data with the items. I want you to analyze all statements. Match all relevant statements to the respective items.*

We aimed to provide an initial classification of the statements for the items of Attractiveness. As a result, ChatGPT-4 generates a first classification. For this, the LLM generalized the analyzed textual data. Results are shown in the following. We only present an example of the first two item pairs:

- ***Attractiveness:*** **Annoying / Enjoyable**
  Annoying: Frequent complaints about too much advertising, unorganized feeds, and overwhelming amounts of content were common, making the experience less enjoyable for some participants (e.g., Participants 1, 7, 14, 22).
  Enjoyable: Others found the diversity in content and the presence of videos and stories to be enjoyable, citing the variety as an ”Eyecatcher” and appreciating the personalized feed (e.g., Participants 2, 5, 21).

- ***Stimulation:*** **Not Interesting/Interesting**
  Interesting: Participants who encountered content aligned with their interests found the Reels feature interesting. For example, Participant 6 appreciated the ease of scrolling and discovering content, though they wished for better control over video pausing.
  Not Interesting: Feedback from Participant 2 and Participant 9 suggested that not all content was appealing, with some users finding the Reels not tailored enough to their preferences or too similar to content from other platforms like TikTok.

Moreover, we want to provide both quantitative and qualitative perspectives as described before. Therefore, we applied **prompt2 and 3**.

Prompt2: *We want to show a quantitative perspective. Therefore, we want to illustrate the number of statements referring to each of the respective items. Can you analyze and classify all relevant statements to the six items?*

Prompt3: *Furthermore, we want to show a qualitative perspective. Therefore, we want to illustrate the relevant statements referring to each of the respective items. Can you analyze and extract all relevant statements from the file and sort them into the six items?*

As a result, ChatGPT analyzed the quantity *(n=x)* of suitable statements concerning the items. Moreover, original statements were extracted. Due to paper restrictions, we only illustrated the first original statement extracted by ChatGPT-4 for each item.

Attractiveness:

- **Annoying / Enjoyable**: *n=7*
  ”I think they are such eye-catchers.” (Participant 2)
- **Bad / Good**: *n=4*
  ”The design is clear and I can quickly see what I like.” (Participant 8)
- **Unlikable / Pleasing**: *n=3*
  ”Everything is very straightforward, clear and simple.” (Participant 3)
- **Unpleasant / Pleasant**: *n=3*
  ”I think that’s even better, because the dark design is set for me.” (Participant 3)
- **Unattractive / Attractive**: *n=2*
  ”Visually appealing in any case...I like round icons very much.” (Participant 22)
- **Unfriendly / Friendly**: *n=2*
  ”It is easy to scroll. It is understandable.” (Participant 5)

Stimulation:

- **Inferior / Valuable**: *n=22*
  ”I enjoy scrolling through the reels.” (Participant 10)
- **Boring / Exciting**: *n=15*
  ”But I find the reels very exciting. It motivates me to keep watching.” (Participant 4)
- **Not Interesting / Interesting**: *n=25*
  ”There are lots of interesting things coming up and that takes up a lot of my time in the evenings.” (Participant 3)
- **Demotivating / Motivating:**: *n=18*
  ”Sometimes you have to watch the reel twice. Maybe 3 or 4 until you finally stop at the right place.” (Participant 6)

In the last step, we inserted **prompt4 and 5** to match the quantitative and qualitative evaluation results. The prompts are as follows:

Prompt4: *In summary, we conducted a mixed-method approach. We examined the user expectations using UEEE, usability testing by applying Thinking Aloud, and the UX by applying the User Experience Questionnaire. Thus, we provide quantitative and qualitative results. In the following, quantitative results regarding the UX factor Attractiveness are shown.* ***Inserted quantitative evaluation results*** *The* *first number is the quantity of how often the adjective was evaluated as important as expectation. The second number is the mean value from the questionnaire using a 7-point Likert scale. Compare the selected statements of the qualitative results with the quantitative evaluation results.*

Prompt5: *Bring the relevant statements analyzed from the user protocols classified to the respective items together with the quantitative results provided.*

After **prompt4**, the results were generally summarized by ChatGPT, similar to **prompt1**. Therefore, **prompt5** has been introduced to provide a more specific representation referring to the respective user statements. As a result, ChatGPT-4 considered both evaluation results and generated a matching summary for each item. Again, we only show the results of the first two items.

- ***Attractiveness:*** **Annoying / Enjoyable**
  Despite no initial expectations, users found Instagram enjoyable, particularly highlighting dynamic content, such as videos. This suggests Instagram's ability to deliver an enjoyable experience beyond users' anticipations.
- ***Stimulation:*** **Not Interesting/Interesting**
  The statements underscore the importance of content relevance and variety in making the platform interesting, which is mirrored in the high mention count and positive mean value, indicating a strong correlation between user interest and the Stimulation factor.

We conducted this procedure with the textual data from all six tasks. Finally, we combined all results from the exploratory study. Therefore, evaluation results were brought together in Figure 4 regarding the expectation (*Anticipated use*), experience (*After use*), and the number of statements (*Actual use*) identified before. The items are scaled from -1 to 1 concerning expectation and experience (see [40]). Moreover, the dot size linearly represents the frequency of comments. The different colors summarize the items of the respective UX factor.

The results show that items, such as *easy* or *understandable*, have a high rating of both expectations and experience, but few comments. This indicates that it is taken for granted and that this is confirmed afterward. In contrast, the number of statements is higher for the items rated with high expectations but a lower experience. This is plausible as usability test results are usually problem-centered and, thus, rather focus on negative aspects more.

To underline this, we performed the correlation between the three dimensions (see Table III). Expectation and experience strongly correlate. This means that Instagram generally seems to fulfill expectations well. The correlation between experience and the quantity of statements is strongly negative. Thus, the number of statements is greater when expectations are not met. In contrast, less is said when expectations are met. This emphasizes the fact that usability testing focuses on problems.

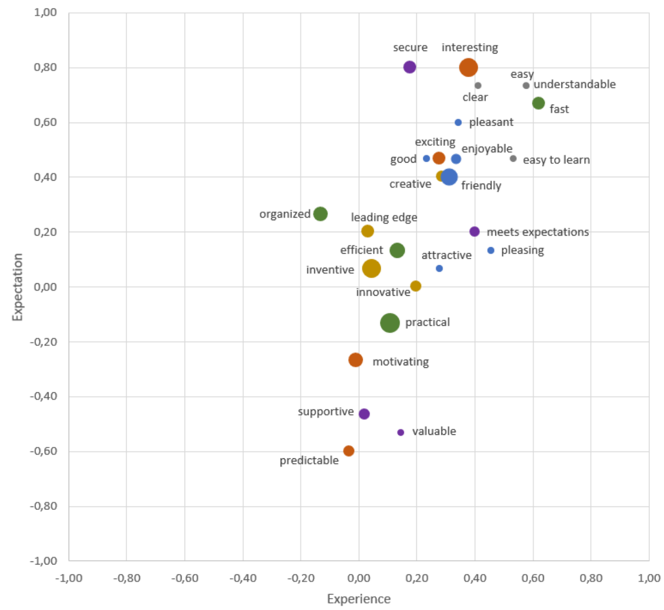

Figure 4. Comparison of Anticipated Use and Actual Use with Extended GenAI Results.

TABLE III. CORRELATION BETWEEN EXPECTATIONS, EXPERIENCE, AND STATEMENTS QUANTITY.

	Expectation	Experience	Statements Quantity
Expectation	1		
Experience	0.668819	1	
Statements Quantity	-0.29093	-0.55583	1

To sum up, the LLM was useful in processing large text data. In particular, ChatGPT-4 was able to (1) generate a generalized summary, (2) identify, count, and extract suitable statements, and (3) match quantitative with qualitative evaluation results.

C. Study Limitations

Concerning this approach, we want to draw several limitations from a practical perspective. As shown, various tasks can be enhanced by applying LLMs. However, it depends on the level of information provided for the LLMs. This concerns both information supplied by prompts and accessed by the LLM itself. Moreover, the non-deterministic nature of LLMs must be mentioned. Thus, results can differ by applying the same prompt twice. Moreover, we did not follow any strategy in prompt engineering. Furthermore, the small sample size must be mentioned. Thirty study participants are just sufficient for the quantitative evaluation part.

VII. Discussion

To sum up, we want to discuss the research questions based on the results. Therefore, we want to show the potential of applying LLMs for different tasks in the UX research process. Using LLMs makes it easy to adopt different perspectives in the research process. Thus, a large amount of data can be processed and analyzed quickly. This would not be possible with a manual analysis. Moreover, no specific strategy concerning prompt engineering was applied.

- ***RQ1: How can GenAI assist UX researchers in categorizing and summarizing user comments?***
  Concerning the first research question, we focus on processing short text data. Therefore, user comments from a study were analyzed and summarized. Results show that the LLM was useful in analyzing and summarizing text data. However, it must be mentioned that the text data from surveys usually differ concerning content and length. The data is extremely heterogeneous. Thus, it is necessary to review AI-generated results by an expert.
- ***RQ2: How can GenAI be used to assist UX researchers in the preparation of research and design activities?***
  Regarding the second research question, we provide insights into preparing a UX evaluation. In particular, we generated specific user tasks for the usability testing method Thinking-Aloud as well as personas. For this, the AI-generated user tasks were useful and specifically related to the research objective of the study. Moreover, ChatGPT-4 generated realistic personas. Remarkably, different aspects that were not described were generated. However, typical stereotypes were included.
- ***RQ3: Can GenAI be used to develop a deeper understanding of the concept of UX?***
  Moreover, we provide insights into identifying a common ground concerning the concept of UX [24][25]. Therefore, we applied ChatGPT-4 to analyze the semantic textual similarity of UX measurement items from established UX questionnaires. Results show that GenAI was useful in (1) (re-) constructing common UX factors, (2) detecting suitable items, and (3) covering semantic similarity. Furthermore, adjectives were assigned to semantic similar UX concepts.
- ***RQ4: How can GenAI be integrated with existing methods in a typical UX evaluation scenario?***
  Concerning the last research question, we conducted an exploratory study investigating users' subjective impressions comprehensively. We applied a mixed-method approach gathering both quantitative and qualitative data. Moreover, we used ChatGPT-4 for different tasks during the research process. In particular, we generated tasks for usability testing as well as analyzing results. Regarding the latter, we showed that quantitative and qualitative evaluation results could be matched. Moreover, improvement suggestions could be derived.

In summary, we state that using LLMs such as ChatGPT-4 is useful for various activities along the UX research process. Thus, the research process can be made more efficient. Above all, time is saved, and practical research processes can be enhanced. However, the researcher should always review results to ensure quality. As shown in this study, the following tasks concerning UX evaluation can be enhanced:

1) Analyzing both small and large text data
2) Creating personas
3) Analyzing semantic similarity
4) Generating user tasks
5) Matching quantitative and qualitative evaluation results

In the following section, we provide a conclusion and outlook.

VIII. Conclusion and Outlook

This article provides insight into the potential of the new technology, GenAI, for UX research. We present different research activities and use cases supported and conducted by using the LLM ChatGPT-4. Results offer concrete usage scenarios for practice.

GenAI is a new and fast-developing area. New potential use cases evolve in short intervals, and it is currently not possible to judge their potential and Challenges definitively. It is important to notice that developed prompts supporting GenAI use cases must be constantly monitored. Moreover, a human expert should critically evaluate their results. Firstly, the underlying GenAI, or especially LLM, also develops over time. Thus, a new version may produce a different output to a prompt than expected. Secondly, since LLMs are non-deterministic, it is, in each case, a good idea to check how stable the output of a prompt is during several runs.

In conclusion, we want to clearly emphasize the potential to enhance UX research and its activities by applying GenAI. The UX research process and its tasks can be improved. Especially in practice, labor-intensive activities along the research process can be accelerated and supported. Moreover, a benefit can be shown in the derivation of results. In practice, different evaluation methods are often combined, resulting in various data. By using GenAI, the results can be matched with each other. This allows the researcher to differentiate from pure scale values. Thus, specific aspects for improving the evaluation object can be derived due to data matching. Conclusions providing both quantitative and qualitative results can be drawn.

This paper is a first step in providing a selected overview of how to apply LLMs in UX research. Future research should further consider the potential of enhancing UX research activities. Therefore, the different UX evaluation methods and their tasks should be investigated along the UX research process.

References

[1] A. Bilgihan, "Gen y customer loyalty in online shopping: An integrated model of trust, user experience and branding," *Computers in Human Behavior*, vol. 61, pp. 103–113, 2016.

[2] L. Deng, D. E. Turner, R. Gehling, and B. Prince, "User experience, satisfaction, and continual usage intention of it," *European Journal of Information Systems*, vol. 19, no. 1, pp. 60–75, 2010.

 ISBN: 978-1-68558-193-0

[3] I. O. for Standardization 9241-210:2019, *Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems*. ISO - International Organization for Standardization, 2019.

[4] J. F. Dumas and J. C. Redish, *A practical guide to usability testing*. Greenwood Publishing Group Inc., 1993.

[5] J. R. Lewis, "Usability testing," in *Handbook of Human Factors and Ergonomics*, Wiley Online Library, 2012, pp. 1267–1312.

[6] J. Sauro. "How many users do people actually test?" (2010), [Online]. Available: https://measuringu.com/actual-users/ [retrieved: 08/2024].

[7] J. L. Jeff Sauro, "Correlations among prototypical usability metrics: Evidence for the construct of usability," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2009, pp. 1609–1618.

[8] M. Schrepp, *User Experience Questionnaires: How to use questionnaires to measure the user experience of your products?* KDP, ISBN-13: 979-8736459766, 2021.

[9] J. Brooke, "Sus: A "quick and dirty'usability," *Usability Evaluation in Industry*, vol. 189, no. 3, pp. 189–194, 1996.

[10] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *HCI and Usability for Education and Work*, A. Holzinger, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 63–76, ISBN: 978-3-540-89350-9.

[11] K. Finstad, "The usability metric for user experience," *Interacting with Computers*, vol. 22, no. 5, pp. 323–327, 2010.

[12] M. Moshagen and M. Thielsch, "Facets of visual aesthetics.," *International Journal of Human-Computer Studies, 25 (13), 1717-1735.*, no. 68(10), pp. 689–709, 2010.

[13] S. Borsci, S. Federici, S. Bacci, M. Gnaldi, and F. Bartolucci, "Assessing user satisfaction in the era of user experience: Comparison of the sus, umux, and umux-lite as a function of product experience," *International Journal of Human-Computer Interaction*, vol. 31, no. 8, pp. 484–495, 2015.

[14] M. Schrepp, J. Kollmorgen, and J. Thomaschewski, "A comparison of sus, umux-lite, and ueq-s," *Journal of User Experience*, vol. 18, no. 2, pp. 86–104, 2023.

[15] D. Simov and M. Schrepp, "Wie man 6000 nutzer-kommentare in qualität umwandelt," *Mensch und Computer 2021-Usability Professionals*, 2021.

[16] B. Liu, *Sentiment analysis and opinion mining*. Springer Nature, 2022.

[17] M. A. Hearst and D. Rosner, "Tag clouds: Data analysis tool or social signaller?" In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*, IEEE, 2008, pp. 160–160.

[18] M. Wattenberg and F. Viegas, "Tag clouds and the case for vernacular visualization," *Interactions*, vol. 15, pp. 49–52, 2008.

[19] J. Kollmorgen, M. Schrepp, and J. Thomaschewski, "Influence of demographic variables and usage behaviour on the perceived user experience," in *Web Information Systems and Technologies*, M. Marchiori, F. J. Domínguez Mayo, and J. Filipe, Eds., Cham: Springer Nature Switzerland, 2023, pp. 186–208, ISBN: 978-3-031-43088-6.

[20] A. Cooper, *Why high-tech products drive us crazy and how to restore the sanity*. Sams Publishing, 2004.

[21] J. Pruitt and T. Adlin, *The persona lifecycle: Keeping people in mind throughout product design*. Elsevier, 2010.

[22] M. Pröbster, J. Hermann, and N. Marsden, "Personas and persons - an empirical study on stereotyping of personas," in *Proceedings of Mensch und Computer*, 2019, pp. 137–145.

[23] J. Grudin, "Why personas work: The psychological evidence," *The persona lifecycle*, vol. 12, pp. 642–664, 2006.

[24] S. Graser, S. Böhm, and M. Schrepp, "Using ChatGPT-4 for the identification of common UX factors within a pool of measurement items from established ux questionnaires," in *CENTRIC 2023: The Sixteenth International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services*, 2023, pp. 19–28.

[25] S. Graser, M. Schrepp, and S. Böhm, "Identifying semantic similarity for UX items from established questionnaires using ChatGPT-4," *International Journal On Advances in Systems and Measurements*, vol. 17, no. 1/2, 2024.

[26] W. B. Albert and T. T. Tullis, *Measuring the User Experience. Collecting, Analyzing, and Presenting UX Metrics*. Morgan Kaufmann, 2022.

[27] A. Assila, K. M. de Oliveira, and H. Ezzedine, "Standardized usability questionnaires: Features and quality focus," *Computer Science and Information Technology*, vol. 6, pp. 15–31, 2016.

[28] M. Schrepp, "A comparison of ux questionnaires - what is their underlying concept of user experience?" In *Mensch und Computer 2020 - Workshopband*, C. Hansen, A. Nürnberger, and B. Preim, Eds., Bonn: Gesellschaft für Informatik e.V., 2020. DOI: 10.18420/muc2020-ws105-236.

[29] A. Hinderks, D. Winter, M. Schrepp, and J. Thomaschewski, "Applicability of user experience and usability questionnaires," *Journal Universal Computer Science*, vol. 25, pp. 1717–1735, 2019.

[30] M. Schrepp *et al.*, "On the importance of UX quality aspects for different product categories," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. In Press, pp. 232–246, Jun. 2023. DOI: 10.9781/ijimai.2023.03.001.

[31] M. Hassenzahl, "The thing and I: Understanding the relationship between user and product," in *Funology: From Usability to Enjoyment*, M. A. Blythe, K. Overbeeke, A. F. Monk, and P. C. Wright, Eds. Dordrecht: Springer Netherlands, 2004, pp. 31–42, ISBN: 978-1-4020-2967-7. DOI: 10.1007/1-4020-2967-5_4.

[32] B. Weiss, I. Wechsung, A. Naumann, and S. Möller, "Subjective evaluation method for speech-based uni- and multimodal applications," in *Perception and Interactive Technologies*, Berlin, Heidelberg: Springer-Verlag, 2008, pp. 285–288, ISBN: 9783540693680. DOI: 10.1007/978-3-540-69369-7_33.

[33] J. Lewis, "Psychometric evaluation of the post-study system usability questionnaire: The pssuq," vol. 2, Jan. 1992, pp. 1259–1263.

[34] D. Winter, M. Schrepp, and J. Thomaschewski, "Faktoren der user experience: Systematische übersicht über produktrelevante UX-qualitätsaspekte," in *Workshop*, A. Endmann, H. Fischer, and M. Krökel, Eds. De Gruyter, 2015, pp. 33–41, ISBN: 9783110443882. DOI: doi:10.1515/9783110443882-005.

[35] S. Graser and S. Böhm, "Quantifying user experience through self-reporting questionnaires: A systematic analysis of the sentence similarity between the items of the measurement approaches," in *HCI International 2023 – Late Breaking Posters*, C. Stephanidis, M. Antona, S. Ntoa, and G. Salvendy, Eds., Cham: Springer Nature Switzerland, 2024, pp. 138–145, ISBN: 978-3-031-49212-9.

[36] S. Graser and S. Böhm, "Applying augmented sbert and bertopic in UX research: A sentence similarity and topic modeling approach to analyzing items from multiple questionnaires," in *Proceedings of the IWEMB 2023, Seventh International Workshop on Entrepreneurship, Electronic, and Mobile Business*, accepted for publication, to be published in 2024, 2023.

[37] B. Rummel and M. S. Martin, "UX fragebögen und wortwolken," *Mensch und Computer 2019-Workshopband*, 2019.

[38] W. Ilmberger, M. Schrepp, and T. Held, "Cognitive processes causing the relationship between aesthetics and usability," in *HCI and Usability for Education and Work: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008, Graz, Austria, November 20-21, 2008. Proceedings 4*, Springer, 2008, pp. 43–54.

[39] M. Schrepp, R. Otten, K. Blum, and J. Thomaschewski, "What causes the dependency between perceived aesthetics and perceived usability?" *International Journal of Interactive Multimedia and Artificial Intelligence*, pp. 78–85, 2021.

[40] D. Oelfin and M. Schrepp, *Evaluation of anticipated user experience*, Mensch und Computer 2021 - Workshopband, Bonn, 2021. DOI: 10.18420/muc2021-mci-ws01-363.

[41] M. Schrepp, "Measuring user experience with modular questionnaires," in *2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, IEEE, 2021, pp. 1–6.

[42] J. Benedek and T. Miner, "Measuring desirability: New methods for evaluating desirability in a usability lab setting," 2002.

APPENDIX

A: Prompts for User Tasks in relation to the Thinking-Aloud

prompt1: Imagine that you want to analyze the user experience of Instagram. To do this, you will prepare a 12-minute user experience test with participants who already have an Instagram account. It is essential that the 6 UX factors and the items associated with them are taken into account and that they are explicitly addressed. For each factor, provide a singular and very detailed UX test task along with very specific instructions for the tester to accomplish that task. Write which goal is to be achieved with it. Dedicate around two minutes per task. In the end, there should be six very specific tasks with a clear aim. Testers should experience and evaluate all factors. The overall UX must be evaluated, not the content! In the end, the testers will be answering the UEQ. Here are the six factors contributing to the user experience (UX) and their evaluation scores: –inserted the six UX scales and their respective items.

prompt2: Make the tasks more specific, e.g., give a certain topic to search for.

prompt3: The user test will be conducted using the think-aloud method. Rewrite the instructions for the testers so that they report on their experiences as much as possible during the test. The test should still last 12 minutes and analyze the UX of Instagram, not the content of the app.

prompt4: The test should have a common thread (e.g., traveling, fitness, healthy eating).

B: User Tasks in relation to the Thinking-Aloud

1. **Attractiveness (2 minutes)**

- Task: Discover and interact with new travel content.

- Instructions: Go to the 'Explore' tab and type "travel" in the search bar. Choose and interact with three different posts (like, comment, or share) that you find most appealing. Look for posts with diverse content like landscapes, cityscapes, and cultural highlights.

- Goal: To evaluate how attractive and engaging the content discovery process is on Instagram.

2. **Efficiency (2 minutes)**

- Task: Post a travel-themed photo efficiently.

- Instructions: Select a personal travel photo from your gallery. Apply the 'Clarendon' filter, write a caption describing the location and your experience, add hashtags, such as TravelMemories and AdventureAwaits, and post it. Time how long it takes to complete this process.

- Goal: To assess the speed and practicality of posting content on Instagram.

3. **Perspicuity (2 minutes)**

- Task: Create and publish a multi-feature Instagram Story.

- Instructions: Create a Story featuring a travel photo. Add a geotag of the location, use the 'Poll' sticker to ask followers about their travel preferences, and add a relevant GIF. Ensure the Story is coherent and visually appealing.

- Goal: To evaluate how easily new users can understand and use various Story features.

4. **Dependability (2 minutes)**

- Task: Review and adjust account security settings.

- Instructions: Navigate to your profile, tap on the three-line menu, select 'Settings', then 'Privacy'. Set the account to private, then go to 'Security', enable two-factor authentication, and review active sessions. Revert any changes if desired after completing the task.

- Goal: To determine how dependable and user-friendly Instagram's privacy and security settings are.

5. **Stimulation (2 minutes)**

- Task: Engage with travel community content.

- Instructions: Follow the official Instagram account of a popular travel destination (e.g., @VisitParis or @ExploreNewZealand). Comment on their latest post with a question or a meaningful remark about the destination.

- Goal: To assess whether interaction with thematic content is stimulating and enjoyable.

6. **Novelty (2 minutes)**

- Task: Explore innovative content formats.

- Instructions: Go to the 'Reels' tab and watch three travel-related Reels. Look for content that showcases unique destinations, creative editing, or unusual travel experiences. Observe the layout and presentation of these Reels.

- Goal: To evaluate the creative and innovative aspects of Instagram's newer features.

Construction of the UXAR-CT – a User eXperience Questionnaire for Augmented Reality in Corporate Training

Stefan Graser
stefan.graser@hs-rm.de
Center for Advanced E-Business Studies
RheinMain University of Applied Sciences
Wiesbaden, Germany
ORCID: 0000-0002-5221-2959

Martin Schrepp
martin.schrepp@sap.com
SAP SE
Walldorf, Germany
ORCID: 0000-0001-7855-2524

Stephan Böhm
stephan.boehm@hs-rm.de
Center for Advanced E-Business Studies
RheinMain University of Applied Sciences
Wiesbaden, Germany
ORCID: 0000-0003-3580-1038

***Abstract*—Measuring User Experience (UX) with questionnaires is essential for developing and improving products. However, no domain-specific standardized UX questionnaire exists for Augmented Reality (AR) in Corporate Training (CT). Thus, this study introduces the UXAR-CT questionnaire - an AR-specific UX questionnaire for CT environments. We describe the construction procedure and the evaluation process of the questionnaire. A set of candidate items was constructed, and a larger sample of participants evaluated several AR-based learning scenarios with these items. Based on the results, we performed a Principal Component Analysis (PCA) to identify relevant measurement items for each scale. The three best-fitting items were selected based on the results to form the final questionnaire. The first results regarding scale quality indicate a high level of internal consistency. The final version of the UXAR-CT questionnaire is provided and will be evaluated in further research.**

Keywords–UXAR-CT; User Experience (UX); UX Measurement; UX Quality Aspects; Questionnaire Construction and Evaluation; Augmented Reality (AR); Corporate Training (CT).

I. Introduction

Augmented Reality (AR) enhances the real world with digital content. Therefore, AR refers to three characteristics: (1) a combination of reality and virtuality, (2) real-time interaction, and (3) registration in 3-D [1]. AR can be applied in various domains [2]. Among these, the field of training and education indicates a potential for improving both teaching and learning [3][4]. The field can be divided into academic teaching and Corporate Training (CT). The latter refers to training scenarios in a professional environment. Only a little research was conducted on the latter [2].

Innovative technologies create new interaction paradigms and, thus, a new experience for the user [5]. User Experience (UX) refers to the subjective impression of users towards a product [6]. Measuring the UX is essential for the design and improvement of products. Different methods for measuring the UX can be found in the literature [7]. Applying standardized UX questionnaires to measure subjective impressions is the most established method in UX research.

Various standardized questionnaires can be found in scientific literature [5]. A questionnaire is based on different dimensions, items, and measurement scales, breaking down the construct of UX. However, there is a lack of common ground based on the level of factors and scales [8][9]. Moreover, questionnaires follow different approaches regarding their structure. Not every UX questionnaire can be applied equally in every evaluation scenario. Thus, it is important to use AR-specific UX questionnaires to capture users' perceptions of the technology successfully.

Only three AR-specific UX questionnaires could be found in the literature. However, none of the existing questionnaires refers to CT. This indicates a lack of research concerning UX measurement approaches for AR in CT. Thus, this study introduces the **UXAR-CT**, a domain-specific standardized UX questionnaire for AR in CT. The UXAR-CT was developed on a common concept concerning UX. The UXAR-CT evaluated different AR-based CT scenarios at the Chamber of Crafts for Lower Franconia in Schweinfurt, Bavaria (Germany). This article is based on our previous paper providing initial insights into the questionnaire development and design [10]. In this study, we describe (1) the construction process in more detail, (2) the research design and procedure of the empirical study, and (3) the first evaluation results of the UXAR-CT. In conclusion, the final version of the UXAR-CT questionnaire is proposed.

Section II provides insights into related work regarding the concept of UX and UX questionnaires for AR. Section III describes the construction procedure of the UXAR-CT. Section IV describes the evaluation scenario and the procedure of the empirical study. Evaluation results are shown in Section V. Finally, Section VI gives a conclusion and outlook.

II. Related Work

This Section II introduces our understanding of the construct UX. We discuss the relevant explanatory approach on which the questionnaire is based. Furthermore, the existing AR-specific UX questionnaires are presented.

A. The Concept of User Experience

UX is proposed as a multidimensional construct with different dimensions regarding the subjective impression of users. Various definitions can be found in the literature. The most common definition is given by ISO 9241-210, which defines UX as a "person's perceptions and responses that result from the use or anticipated use of a product, system or service" [6]. This implies that UX is a subjective construct depending on the user's perception.

Moreover, the definition is rather broad and abstract. Thus, this does not help quantify and measure UX. Different approaches were conducted to break down the construct of UX and achieve a better understanding. One common distinction was made by [11] dividing into *pragmatic* and *hedonic* qualities. Furthermore, some research aimed to consolidate UX factors based on the empirical as well as the semantic similarity on the level of measurement items [8][9][12]–[14].

Schrepp et al. [14] broke down UX into a set of semantically clearly described quality aspects. In this regard, the terms UX factor and UX quality aspect are considered the same. In summary, [14] proposed 16 UX quality aspects shown in the following table I:

TABLE I. CONSOLIDATED UX QUALITY ASPECTS BASED ON [14]

	UX Quality Aspect	Definition
(1)	Perspicuity (PE)	Is it easy to get familiar with the product and to learn how to use it?
(2)	Efficiency (EF)	Can users solve their tasks without unnecessary effort? Does the product react fast?
(3)	Dependability (DE)	Does the user feel in control of the interaction? Does the product react predictably and consistently to user commands?
(4)	Usefulness (US)	Does using the product bring advantages to the user? Does using the product save time and effort?
(5)	Intuitive Use (IU)	Can the product be used immediately without any training or help?
(6)	Adaptability (AD)	Can the product be adapted to personal preferences or personal working styles?
(7)	Novelty (NO)	Is the design of the product creative? Does it catch the interest of users?
(8)	Stimulation (ST)	Is it exciting and motivating to use the product? Is it fun to use?
(9)	Clarity (CL)	Does the user interface of the product look ordered, tidy, and clear?
(10)	Quality of Content (QC)	Is the information provided by the product always actual and of good quality?
(11)	Immersion (IM)	Does the user forget time and sink completely into the interaction with the product?
(12)	Aesthetics (AE)	Does the product look beautiful and appealing?
(13)	Identity (ID)	Does the product help the user to socialize and to present themselves positively to other people?
(14)	Loyalty (LO)	Do people stick with the product even if there are alternative products for the same task?
(15)	Trust (TR)	Do users think that their data is in safe hands and not misused to harm them?
(16)	Value (VA)	Does the product design look professional and of high quality?

This distinction based on UX quality aspects is a common perspective in UX research [14]–[16]. For instance, common UX questionnaires are based on this approach [17]–[22]. Identifying the relevant UX quality aspects that differ depending on the application field and evaluation object is essential. Not every quality aspect suits every evaluation scenario. Thus, the importance of the UX quality aspects regarding the objective must be considered [10][14]. In the following, the existing UX AR questionnaires are described.

B. UX Questionnaires for Augmented Reality

As described, many UX questionnaires are available in the literature [5]. However, most of them are general and unrelated to AR. Only three AR-specific UX questionnaires were identified, as shown in the following.

1) **Handheld Augmented Reality Usability Scale (HARUS)** [23][24]
2) **Augmented Reality Immersion (ARI) Questionnaire** [25]
3) **Customizable Interactions Questionnaire (CIQ)** [26]

The Handheld Augmented Reality Usability Scale (HARUS) developed by [23][24] specifically focuses on the usability of handheld AR devices. The questionnaire consists of the factors *Manipulability* referring to the ease of handling, and *Comprehensibility* referring to the ease of understanding. Each factor contains eight items. The evaluation is based on a seven-point Likert scale resulting in a computed score from 0 to 100, similar to the System Usability Scale (SUS) [23][24][27].

The Augmented Reality Immersion (ARI) Questionnaire developed by [25] consists of three main scales *Engagement*, *Engrossment*, and *Total Immersion*, and further six subscales *Interest*, *Usability*, *Emotional Attachment*, *Focus of Attention*, *Presence*, and *Flow* with a total of 21 items on a seven-point Likert scale. The focus is on measuring the immersion which concerns the cognitive and emotional absorption. The questionnaire is intended to be applied in location-aware AR settings.

The Customizable Interactions Questionnaire (CIQ) developed by [26] aims to gather the interaction quality between the user and virtual objects in AR scenarios. The questionnaire is based on the five scales *Quality of Interactions*, *Comfort*, *Assessment of Task Performance*, *Consistency with Expectation*, and *Quality of the Sensory Enhancements* with a total of 17 items on a five-point Likert scale [26].

To sum up, the HARUS specializes in AR devices with a specific Usability focus. In contrast, the ARI questionnaire focuses on immersion in location-based settings. The CIQ, in turn, focuses on the quality of interaction. This shows the still great heterogeneity of the questionnaires even in a specific domain.

In summary, the AR-specific questionnaires consist of different UX factors, indicating different measurement focuses. Furthermore, the questionnaires did not refer to CT. It remains unclear which UX aspects are relevant for AR in CT. This indicates a lack of applicability for UX evaluation in the field of CT.

III. Construction of the UXAR-CT Questionnaire

In the following, the construction of the **UXAR-CT** questionnaire is described. The developement process consists of three steps illustrated in the following:

1) Determination of relevant UX quality aspects (Section III-A, [10])
2) Construction of UX measurement items (Section III-B, [10])
3) Evaluation and derivation of the final questionnaire (Sections IV and V)

Steps (1) and (2) are shortly introduced in Sections III-A and III-B. Detailed information for the initial development phase can be found in the previous paper [10]. Step (3) was conducted within this study. We conducted an empirical study described in Section IV. Based on this, the evaluation results and construction of the final questionnaire are illustrated in Section V.

Finally, we would like to add a statement regarding the previous study. Within this, we declared the UXAR-CT as a measurement approach concerning **Mobile** Augmented Reality. However, several reasons have emerged that cause us to depart from this terminology. On the one hand, the distinction between the terms is still not clearly outlined in the literature. On the other hand, handheld and head-mounted devices were used

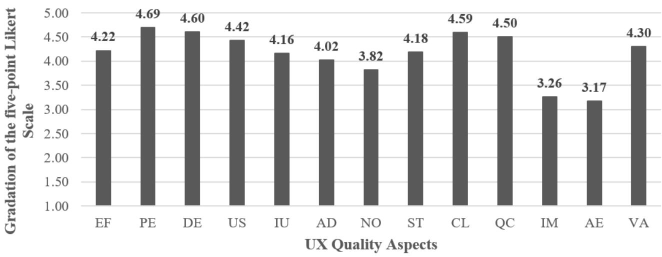

Figure 1. Importance Rating of UX Quality Aspects Concerning AR in CT.

at the Chamber of Crafts in Schweinfurt. To keep it simple, we will only refer to AR now. This also does not affect the first study on the importance of UX quality aspects. Thus, this change is therefore not considered critical. In the following, the three steps of construction are illustrated.

A. *Determining Relevant UX Quality Aspects*

The questionnaire is based on the perspective concerning UX quality aspects according to [14]. As this is a common approach of several questionnaires, we also rely on this understanding (See Section II-A, Table I). Therefore, the relevant UX quality aspects for AR in CT were considered. Data concerning the importance of different UX quality aspects were collected over an online survey. In the following, we only refer to the main results of this survey. Details are described in our previous paper [10].

Participants could start the survey by clicking a link in an invitation mail. In an introductory part, the technology AR was explained in detail. To give participants more context, a video showing a CT scenario using AR was displayed. The 13 of the 16 UX aspects described in Table I are displayed. The aspects *Identity*, *Loyalty*, and *Trust* were excluded since they obviously play no role in the CT scenario). Participants were asked to rate the importance of each presented UX quality aspect in relation to the shown CT scenario. For the rating, we applied a five-point Likert scale (from *"not important at all"* scored as 1 to *"very important"* scored as 5). We collected 121 complete responses using this survey.

Figure 1 illustrates the overall importance ratings for the UX aspects. The Y-axis presents the gradation of the applied five-point Likert scale (from *"not important at all"–1* to *"very important"–5*). The X-axis shows the mean values for the evaluated UX quality aspects. The highest rated and thus most important aspect for AR in CT is *Perspicuity* with an average score of 4.69, whereas the lowest rated one is *Aesthetics* with an average score of 3.17. Nevertheless, none of the quality aspects were rated as unimportant (or less than 3) on average.

B. *Construction of Measurement Items*

We decided to consider only the five most important UX quality aspects for the questionnaire to keep the length manageable. For the construction of the item pool, a large set of items from 60 established UX questionnaires (with overall around 1500 single items) was analyzed. Ten suitable items for each UX quality aspect were extracted and reformulated to match our research context of mobile augmented reality in learning. This initial list of statements was then again reviewed concerning their formulations and potential duplicates or items that were, after the reformulation, too similar to other items and were removed. In the next step, the study's three authors independently selected the most representative items based on their expertise.

This resulted in the following list of candidate items (the term in brackets is later used to refer to the item):

Efficiency

- Using the application for learning is practical (EF1)
- The application reduces the learning effort (EF2)
- The application helps me to learn faster (EF3)
- The application saves me time while learning (EF4)
- The application improves my learning and work performance (EF5)

Perspicuity

- It was clear from the start how I had to use the application for learning (PE1)
- It is easy/simple to learn how to use the application (PE2)
- The information in the application is easy to understand (PE3)
- The operation of the application is logical (PE4)
- It is easy to navigate between individual parts of the application (PE5)

Dependability

- The behavior of the application always meets my expectations (DE1)
- I am confident in using the application at all times (DE2)
- The application is easy to control (DE3)
- I always have control over the application at every step (DE4)
- It is easy to find your way around the application (DE5)
- The application always responds comprehensible (DE6)

Usefulness

- The application helps me to learn (US1)
- It is a great advantage to use the application when learning (US2)
- The application is useful for learning (US3)
- I find the application useful for learning (US4)
- The application fully meets my expectations (US5)

Clarity

- The information on the display is clearly laid out (CL1)
- The information on the display is clear (CL2)
- The display of the application looks tidy (CL3)
- It's easy to find the information I need (CL4)

We collected data in a study with German participants. Thus, the items were carefully translated into German. The German items can be found in the appendix. A seven-point Likert scale was applied (See Figure 2). We chose an emoji-based scale to reduce mental effort and increase both attention and clarity for the trainees.

The following describes the evaluation scenario and procedure of the empirical study.

Figure 2. Emoji-based seven-point Likert Scale.

IV. Empirical Study Design for the Questionnaire Evaluation

We conducted an empirical study to evaluate the constructed UXAR-CT questionnaire within AR-based CT learning scenarios. In the following, we present the respective AR-based CT scenarios and the study procedure.

A. Evaluation Scenario: Augmented Reality-based CT Applications

We collaborated with the Chamber of Crafts for Lower Franconia in Schweinfurt, Bavaria (Germany) for the evaluation and data collection. In summary, 53 Chambers of Crafts can be found in Germany. The main concept is representing the entire crafts sector in Germany. Moreover, they are responsible for the education and training of apprentices in the craft sector, including over 130 apprenticeships in the fields of construction, wood, metal/electrical, clothing, food, health, glass, and paper. Therefore, various courses take place at the chambers, which are relevant and mandatory for the apprenticeships [28].

The institution in Schweinfurt has carried out the project **ARihA – Augmented Reality in Corporate Training**. The idea of the project was to develop and implement innovative digital learning and teaching methods using AR in CT, resulting in an immersive and action-oriented learning experience. The aim was to enhance the learning effectiveness and motivation of the trainees. The ARihA project was funded by the German Federal Ministry of Education and Research (BMBF). In the context of the project, five AR-based CT scenarios among three craft sectors were developed. The scenarios are part of the apprenticeship in the electrical engineering, metal construction, and automotive engineering sectors. Head-mounted displays (Holo-Lens 2) and handheld devices (tablets) were applied for the applications. An overview is given in the following Table II [29]. Moreover, some examples from the AR-based CT scenarios are shown in Figures 3 and 4.

TABLE II. AR APPLICATIONS FOR CORPORATE TRAINING [14]

Training Process	Craft Sector	AR Device
Testing of electronic devices	electrical engineering	head-mounted
Processing of high-grade steel and aluminum	metal construction	head-mounted
Installation of locking and security systems	metal construction	handheld
Troubleshooting and use of measurement devices on a car lighting wall	automotive engineering	both
Changing the timing belt on a car engine	automotive engineering	handheld

Further information can be found online [29]. The described AR applications of the CT scenarios are the basis for evaluating the UXAR-CT. The study approach is described below.

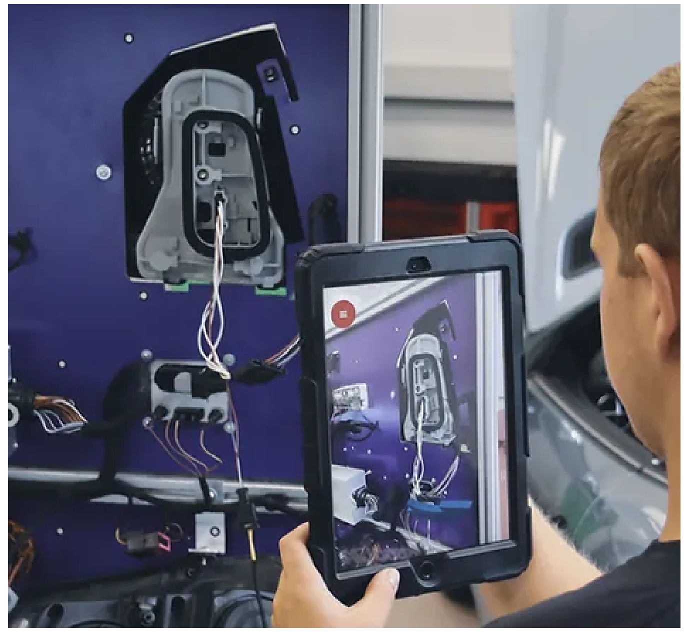

Figure 3. Troubleshooting and use of measurement devices on a car lighting wall (1) [10][29].

Figure 4. Troubleshooting and use of measurement devices on a car lighting wall (2) [10][29].

B. Empirical Survey Structure and Procedure

We applied the UXAR-CT questionnaire to the five AR-based CT scenarios in the Chamber of Crafts. The AR applications of the different CT scenarios are regularly used in daily teaching activities. After the participants finished their learning activities, they filled out a survey (paper-pencil) containing a short motivation and instruction.

The instruction was followed by questions concerning age, gender, the learning scenario just completed, and the participants' apprenticeship. Moreover, we added two open questions:

- *What did you like about the application?*
- *What should be improved?*

After these initial questions, the candidate items described above (See Section III-B) were presented. In addition to the candidate items, one additional question was added:

Overall, I am satisfied with the support provided by the application for my learning tasks.

This is a classical item to measure overall satisfaction, and the responses should help to select the best-fitting items for a scale (see description below). To sum up, the applied UXAR-CT questionnaire has the following structure:

1) Motivation and Instruction
2) Demographics (4 Items)
3) Open Questions (2 Items)
4) Overall Satisfaction (1 Item)
5) Efficiency (5 Items)
6) Perspicuity (5 Items)
7) Dependability (6 Items)
8) Usefulness (5 Items)
9) Clarity (4 Items)

The items referring to the UX scales (See Section III-B) were presented in random order. The questionnaire contains 32 items, 26 of which relate to the UX (Overall Satisfaction included).

We collected responses to our survey from December 2023 to May 2024 in the Chamber of Crafts for Lower Franconia in Schweinfurt, Bavaria (Germany). The course instructors were briefed on the questionnaire and were available to answer the trainees' questions during completion. In this regard, we refer to the concept of the Chamber of Handicrafts. Courses usually last one week. In addition, courses, including AR-based CT scenarios, are not held every week. This emphasizes the difficulty of data collection in the CT application field. The evaluation results of the UXAR-CT are presented below.

V. Evaluation Results of the UXAR-CT Questionnaire

This Section V provides the evaluation results of the empirical study. Both descriptive results and results of the Principal Component Analysis (PCA) are shown. The best-fitting items were selected. Moreover, the first results regarding scale quality are presented. Finally, further development suggestions are provided.

A. Descriptive Evaluation Results

We first want to present the descriptive evaluation results of the empirical study. During the study period of six months, a total of ***106*** completed questionnaires were collected. There were 12 female and 93 male participants. One person did not provide any information about the gender. The average age is 19.

Moreover, we reviewed the qualitative results regarding the open questions. Thus, some qualitative insights can be provided. The trainees were generally positive about the AR-based CT scenarios. In the following, we list the qualitative answers mentioned at least five times or more. The trainees particularly liked the visual presentation of the learning and teaching content and 3D objects. Furthermore, it was perceived as simple and supportive to understand the explanations regarding the learning content. Moreover, the interaction with the learning content was perceived as useful. Additionally, the use and handling of the applications was described as simple. In summary, using AR was perceived as novel, fun, and varied in contrast to other learning methods.

However, some aspects for improvement were declared. In this context, the reaction and function of the system has been mentioned. In some cases, this did not always work correctly. Furthermore, the quality of learning and teaching content presented was criticized. Lastly, it was noted that extended use can cause both headaches and eye pain. Moreover, the comfort of wearing HoloLens2 decreases. The following Table III summarizes the qualitative results.

TABLE III. QUALITATIVE INSIGHTS FROM THE EVALUATION.

Positive Aspects	Improvement Potential
visual representation and 3D objects	content quality
simplicity and support of understanding	functionality and system reaction
interaction	headaches and eye pain
simple handling	wearing comfort
novelty, fun, and variety	

B. Semantic Homogenity

A Principal Component Analysis (PCA) [30][31] with varimax rotation was performed for each of the proposed scales. The goal of this analysis was first to see if the candidate items are uni-dimensional or if they split into separate components that represent different semantic meanings. Secondly, the loading of the single items on the corresponding component demonstrates how well the items represent this component.

We show in the following the results of the PCA per suggested scale. The figures show the eigenvalues of the components. To determine if a scale is semantically homogeneous, we use the scree-test [32] (determine the point after which eigenvalues differ only slightly) and the Kaiser-Gutmann [33] criterion (remove components with eigenvalues less than 1).

Efficiency: The scree test and the Kaiser-Gutmann criterion (See Figure 5) both indicate a solution with one single component. This component explains 76% of the variability in the data.

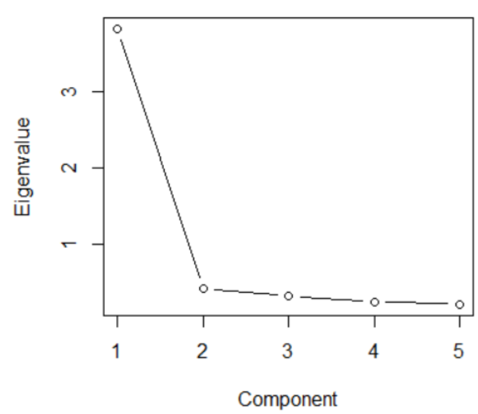

Figure 5. Screen plot of the eigenvalues for the efficiency scale.

Perspicuity: Again, both criteria (See Figure 6) indicate a solution with one single component that explains 64% of the variability in the data.

Dependability: Also, for this semantic group of items, both criteria (See Figure 7) indicate a single component. The component explains 66% of the variability in the data.

Usefulness: The scree test and the Kaiser-Gutmann criterion (See Figure 8) both indicate a solution with one single component. This component explains 79% of the variability in the data.

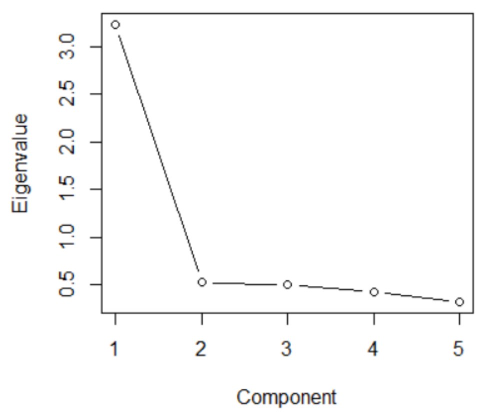

Figure 6. Screen plot of the eigenvalues for the perspicuity scale.

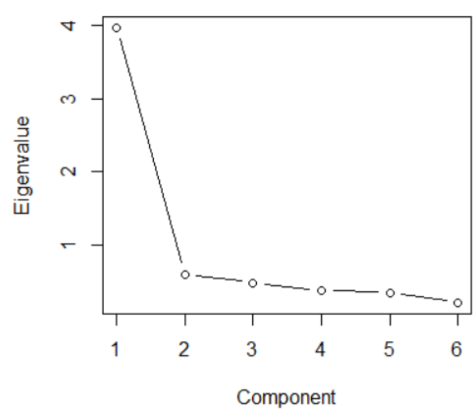

Figure 7. Screen plot of the eigenvalues for the dependability scale.

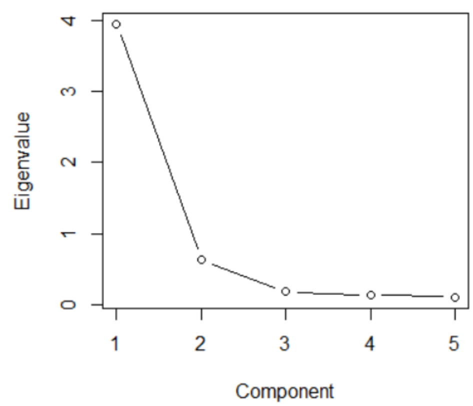

Figure 8. Screen plot of the eigenvalues for the usefulness scale.

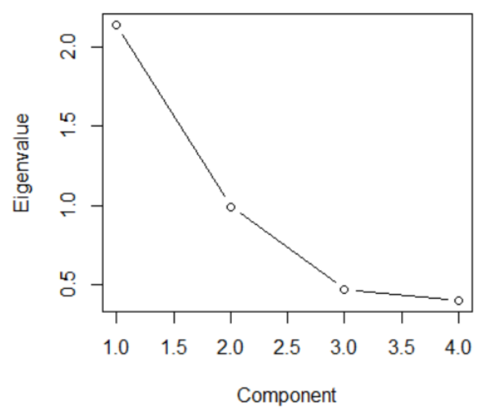

Figure 9. Screen plot of the eigenvalues for the clarity scale.

Clarity: Again, the scree test and the Kaiser-Gutmann criterion (See Figure 9) indicate considering only a single component. This component explains 53% of the variability in the data.

Thus, the analysis indicates semantic homogeneity of the candidate items for all proposed scales. Therefore, we can safely assume that all candidate items in a proposed scale measure the same semantic concept and can now, in the next step, identify those items that form the best representation of this concept.

C. *Selection of the Items*

We use established criteria to select the best-fitting items per scale. Firstly, guidelines suggested by [34] state that loadings greater than 0.4 are generally considered acceptable. Thus, items with a lower loading are removed from the selection. Secondly, the higher the loading, the better the component represented by the item. However, in cases where the loadings of the remaining items are very similar, we used the item's correlation to the overall satisfaction rating as a basis for decision-making.

For Efficiency, all items show very similar loadings on the first component. The items EF1 (0.72), EF3 (0.62), and EF5 (0.62) show clearly higher correlations to the overall satisfaction than EF2 (0.41) and EF4 (0.47) and are thus selected to represent this scale (values in parenthesis are the correlations). For Perspicuity, all items show similar loadings on the component. Based on the correlations to overall satisfaction, the items PE2 (0.63), PE3 (0.63), and PE4 (0.55) were selected to represent the scale since they show a slightly higher correlation to overall satisfaction than PE1 (0.53) and PE5 (0.51), but the difference is not massive in this case. For Dependability, items DE3 (0.63), DE4 (0.55), and DE5 (0.51) showed higher loadings on the component than DE1, DE2, and DE6 and are thus selected. For Usefulness, items US1 (0.67), US2 (0.67), US3 (0.65), and US4 (0.68) showed similar loadings on the component and similar correlations with the overall satisfaction, while the loading of US5 was clearly lower. We select US1, US2, and US4 to represent this scale. For Clarity, item CL3 shows a much lower loading than the other three items, CL1 (0.54), CL2 (0.48), and CL4 (0.55), which are therefore selected.

The final version of the UXAR-CT in English can be found in Appendix B.

In the following, preliminary results concerning scale quality are illustrated.

D. *First Results concerning Scale Quality*

From our data set, we can calculate the Cronbach Alpha coefficient, which is a basic score for internal scale consistency. The corresponding values are 0.90 for Efficiency, 0.81 for Perspicuity, 0.85 for Dependability, 0.95 for Usefulness, and 0.79 for Clarity. These scores indicate a high level of internal consistency. Further quality criteria must be determined in the practical application of the questionnaire.

E. Further Development Suggestions

Based on the results, we can present a reliable questionnaire with the UXAR-CT. However, the questionnaire indicates potential for further development. On the one hand, there is no weighting along the five UX scales regarding importance and relevance. On the other hand, the UXAR-CT only provides a purely UX perspective. Both aspects are discussed in the following.

When analyzing questionnaires with numerical evaluation results, the question of how these can be interpreted always arises. It is common practice to compare the results with others. However, there is a lack of common ground within the UX questionnaires differing in structure and focus [5][8][9][15]. In addition, the UXAR-CT is a domain-specific questionnaire. This results in the difficulty of comparing evaluation results. Establishing a benchmark, such as with the UEQ, is common practice. However, this requires a large amount of data [19][35]. Creating a benchmark for a domain-specific questionnaire like the UXAR-CT is almost impossible. Another way to differentiate between the scales is by including and using an ***external criterion***. Thus, the relationship between the external criterion and the scales based on their correlation and evaluation results can be considered, and weightings can be derived. Concerning the UXAR-CT, we applied the *Overall satisfaction* as an external criterion. Thus, we can use this for further development to determine the relevance of the different UX scales.

Another aspect are the the questionnaire components. Up to now, the UXAR-CT only contains UX quality aspects. While these quality aspects were identified as relevant for AR in CT (See Section III-A, [10]), there is no relation to specific system properties concerning the applications. In other words, a developer can only draw limited conclusions about the improvement potential of the system properties on the basis of the evaluation results. Relevant measurement items referring to specific AR system properties should be integrated. In this way, practical suggestions for improving the AR-based CT applications could be implemented, and thus, the learning experience could be further improved. In this regard, we refer to an exemplary questionnaire with scales and items addressing relevant system properties: the *ARcis questionnaire*. The ARcis conceptualizes the three AR characteristics by [1] and relates them to the learning context. The questionnaire measures learners' perceptions of these characteristics [36].

VI. Conclusion

This article is based on our previous research [10] and describes the final construction of the UXAR-CT, a domain-specific standardized UX questionnaire referring to AR-based CT scenarios. We conducted an empirical survey applying the first version of the UXAR-CT to evaluate different AR-based CT applications at the Chamber of Handicrafts for Lower Franconia in Schweinfurt (Germany). We determined the best-fitting items per scale by performing a PCA based on the evaluation results. As a result, we provide a reliable questionnaire based on established UX quality aspects from a user perspective.

A. Implications

No standardized UX questionnaire for AR exists in the CT application field. Thus, there is a lack of research. By constructing the UXAR-CT questionnaire, we provide a valuable contribution to this research field. The UXAR-CT, therefore, is based on an established common ground referring to UX quality aspects by [14]. Based on this, the importance of the respective quality aspects in relation to AR in CT was evaluated. Therefore, the questionnaire consists of UX quality aspects that are relevant to AR in CT. Thus, we can provide a reliable questionnaire based on the relevant theoretical foundation from a research perspective. The questionnaire can be applied in AR-based CT scenarios to evaluate the subjective impression of the users in practical settings. This provides valuable insights into the AR applications to improve them further.

B. Outlook and Future Research

Finally, we want to give an outlook for future research activities. The UXAR-CT should be regularly applied in different AR-based CT scenarios in different organizations to gather further data for validation, which is the last step of a questionnaire creation process. Moreover, the questionnaire was constructed in German. Thus, an English version of the UXAR-CT would be useful for further expansion of the questionnaire and the data collection process. Additionally, we declared some suggestions for further development (See Section V-E. Up to now, the UXAR-CT only covers UX quality aspects as components. In future research activities, we want to expand the questionnaire to include relevant measurement items concerning AR system properties, enabling practical derivations regarding AR application improvement.

Acknowledgement

We want to express our extraordinary thanks to our research partner, the Chamber of Crafts for Lower Franconia in Schweinfurt, Bavaria (Germany), for supporting this research.

References

[1] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators & Virtual Environments*, vol. 6, pp. 355–385, 1997.

[2] S. Graser and S. Böhm, "A systematic literature review on technology acceptance research on augmented reality in the field of training and education," in *CENTRIC 2022, The Fifteenth International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services*, 2022, pp. 20–22, ISBN: 978-1-61208-999-7.

[3] Subandi, Joniriadi, A. A. Syahidi, and A. Mohamed, "Mobile augmented reality application with multi-interaction for learning solutions on the topic of computer network devices (effectiveness, interface, and experience design)," in *2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE)*, 2020, pp. 1–6. DOI: 10.1109/ICVEE50212.2020.9243292.

[4] S. Criollo-C *et al.*, "Towards a new learning experience through a mobile application with augmented reality in engineering education," *Applied Sciences*, vol. 11, no. 11, 2021, ISSN: 2076-3417. [Online]. Available: https://www.mdpi.com/2076-3417/11/11/4921 [retrieved: 08/2024].

[5] M. Schrepp, "A comparison of ux questionnaires - what is their underlying concept of user experience?" In *Mensch und Computer 2020 - Workshopband*, C. Hansen, A. Nürnberger, and B. Preim, Eds., Bonn: Gesellschaft für Informatik e.V., 2020. DOI: 10.18420/muc2020-ws105-236.

[6] I. O. for Standardization 9241-210:2019, *Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems*. ISO - International Organization for Standardization, 2019.

[7] W. B. Albert and T. T. Tullis, *Measuring the User Experience. Collecting, Analyzing, and Presenting UX Metrics*. Morgan Kaufmann, 2022.

[8] S. Graser and S. Böhm, "Quantifying user experience through self-reporting questionnaires: A systematic analysis of sentence similarity between the items of the measurement approaches," in *Lecture Notes in Computer Science, LNCS, volume 14014*, Springer Nature, 2023, pp. 138–145.

[9] S. Graser, S. Böhm, and M. Schrepp, "Using ChatGPT-4 for the identification of common ux factors within a pool of measurement items from established UX questionnaires," in *CENTRIC 2023, The Sixteenth International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services*, 2023, pp. 19–28.

[10] S. Graser, M. Schrepp, and S. Böhm, "UXAR-CT – an approach for measuring UX for mobile augmented reality applications in corporate training," in *Human-Centered Design, Operation and Evaluation of Mobile Communications*, J. Wei and G. Margetis, Eds., Cham: Springer Nature Switzerland, 2024, pp. 211–234, ISBN: 978-3-031-60458-4.

[11] M. Hassenzahl, "The thing and i: Understanding the relationship between user and product," in *Funology: From Usability to Enjoyment*, M. A. Blythe, K. Overbeeke, A. F. Monk, and P. C. Wright, Eds. Dordrecht: Springer Netherlands, 2004, pp. 31–42, ISBN: 978-1-4020-2967-7. DOI: 10.1007/1-4020-2967-5_4.

[12] D. Winter, M. Schrepp, and J. Thomaschewski, "Faktoren der user experience: Systematische übersicht über produktrelevante ux-qualitätsaspekte," in *Workshop*, A. Endmann, H. Fischer, and M. Krökel, Eds. Berlin, München, Boston: De Gruyter, 2015, pp. 33–41, ISBN: 9783110443882. DOI: doi:10.1515/9783110443882-005.

[13] A. Hinderks, D. Winter, M. Schrepp, and J. Thomaschewski, "Applicability of user experience and usability questionnaires," *J. Univers. Comput. Sci.*, vol. 25, pp. 1717–1735, 2019.

[14] M. Schrepp *et al.*, "On the importance of UX quality aspects for different product categories," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. In Press, pp. 232–246, Jun. 2023. DOI: 10.9781/ijimai.2023.03.001.

[15] M. Schrepp, *User Experience Questionnaires: How to use questionnaires to measure the user experience of your products?* KDP, ISBN-13: 979-8736459766, 2021.

[16] Y. Rogers, H. Sharp, and J. Preece, "Interaction design-beyond human-computer interaction wiley," *Google Scholar Digital Library Digital Library*, 2002, ISSN: 9781119547259. [retrieved: 08/2024].

[17] J. P. Chin, V. A. Diehl, and K. L. Norman, "Development of an instrument measuring user satisfaction of the human-computer interface," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1988, pp. 213–218.

[18] J. Kirakowski and M. Corbett, "Sumi: The software usability measurement inventory," *British Journal of Educational Technology*, vol. 24, no. 3, pp. 210–2, 1993.

[19] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *HCI and Usability for Education and Work*, A. Holzinger, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 63–76, ISBN: 978-3-540-89350-9.

[20] H. X. Lin, Y.-Y. Choong, and G. Salvendy, "A proposed index of usability: A method for comparing the relative usability of different software systems," *Behaviour & Information Technology*, vol. 16, no. 4-5, pp. 267–277, 1997.

[21] M. Schrepp and J. Thomaschewski, "Design and validation of a framework for the creation of user experience questionnaires," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. InPress, pp. 88–95, Dec. 2019. DOI: 10.9781/ijimai.2019.06.006.

[22] H. Willumeit, G. Gediga, and K. C. Hamborg, "Isometricsl: Ein verfahren zur formativen evaluation von software nach iso 9241/10," *Ergonomie und Informatik*, vol. 27, pp. 5–12, 1996.

[23] M. E. C. Santos *et al.*, "Toward standard usability questionnaires for handheld augmented reality," *IEEE Computer Graphics and Applications*, vol. 35, no. 5, pp. 66–75, 2015. DOI: 10.1109/MCG.2015.94.

[24] M. E. C. Santos *et al.*, "A usability scale for handheld augmented reality," ser. VRST '14, Edinburgh, Scotland: Association for Computing Machinery, 2014, ISBN: 9781450332538. [Online]. Available: https://doi.org/10.1145/2671015.2671019 [retrieved: 08/2024].

[25] Y. Georgiou and E. A. Kyza, "The development and validation of the ari questionnaire: An instrument for measuring immersion in location-based augmented reality settings," *International Journal of Human-Computer Studies*, vol. 98, pp. 24–37, 2017, ISSN: 1071-5819. DOI: https://doi.org/10.1016/j.ijhcs.2016.09.014.

[26] M. Gao and D. Boehm-Davis, "Development of a customizable interactions questionnaire (ciq) for evaluating interactions with objects in augmented/virtual reality," *Virtual Reality*, vol. 27, pp. 1–18, Aug. 2022. DOI: 10.1007/s10055-022-00678-8.

[27] J. Brooke, "Sus: A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, Nov. 1995.

[28] *Crafts in germany*, 2023. [Online]. Available: https://www.handwerkskammer.de/ [retrieved: 08/2024].

[29] *Augmented reality in vocational training (ariha)*. [Online]. Available: https://www.projekt-ariha.de/ [retrieved: 08/2024].

[30] K. Pearson, "LIII. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[31] H. Hotelling, "Analysis of a complex of statistical variables into principal components.," *Journal of Educational Psychology*, vol. 24, no. 6, p. 417, 1933.

[32] R. B. Cattell, "The scree test for the number of factors," *Multivariate Behavioral Research*, vol. 1, no. 2, pp. 245–276, 1966.

[33] L. Guttman, "Some necessary conditions for common-factor analysis," *Psychometrika*, vol. 19, no. 2, pp. 149–161, 1954.

[34] A. L. Comrey and H. B. Lee, *A first course in factor analysis*. Psychology press, 2013.

[35] M. Schrepp, J. Thomaschewski, and A. Hinderks, "Construction of a benchmark for the user experience questionnaire (ueq)," 2017.

[36] J. M. Krüger and D. Bodemer, "Work-in-progress—measuring learners' subjective experience in augmented reality: First evaluation of the arcis questionnaire," in *2022 8th International Conference of the Immersive Learning Research Network (iLRN)*, 2022, pp. 1–3. DOI: 10.23919/iLRN55037.2022.9815900.

Appendix

A) German Translations of the Measurement Items

Overall satisfaction (Gesamtzufriedenheit)

- Insgesamt bin ich mit der Unterstützung der Anwendung fürs Lernen zufrieden.

Efficiency (Effizienz)

- Die Nutzung der Anwendung fürs Lernen ist praktisch. (EF1)
- Die Anwendung reduziert den Lernaufwand. (EF2)
- Die Anwendung hilft mir schneller zu lernen. (EF3)
- Die Anwendung spart mir Zeit beim Lernen. (EF4)
- Die Anwendung verbessert meine Lern- und Arbeitsleistung. (EF5)

Perspicuity (Durchschaubarkeit)

- Es war von Beginn an klar, wie ich die Anwendung zum Lernen nutzen muss. (PE1)
- Es ist leicht/einfach zu lernen, wie man die Anwendung benutzt. (PE2)
- Die Informationen der Anwendung sind einfach zu verstehen. (PE3)
- Die Bedienung der Anwendung ist logisch. (PE4)
- Es ist einfach, zwischen einzelnen Teilen der Anwendung zu navigieren. (PE5)

Dependability (Steuerbarkeit)

- Das Verhalten der Anwendung entspricht stets meinen Erwartungen. (DE1)
- Ich bin zu jeder Zeit sicher im Umgang mit der Anwendung. (DE2)
- Die Anwendung ist einfach zu kontrollieren. (DE3)
- Ich habe bei allen Schritten immer die Kontrolle über die Anwendung. (DE4)
- Es ist einfach sich in der Anwendung zurechtzufinden. (DE5)
- Die Anwendung reagiert immer verständlich. (DE6)

Usefulness (Nützlichkeit)

- Die Anwendung hilft mir beim Lernen. (US1)
- Es ist von großem Vorteil, die Anwendung beim Lernen zu nutzen. (US2)
- Die Anwendung ist nützlich beim Lernen. (US3)
- Ich finde die Anwendung fürs Lernen nützlich. (US4)
- Die Anwendung erfüllt meine Erwartungen vollständig. (US5)

Clarity (Übersichtlichkeit)

- Die Darstellung der Informationen im Display ist übersichtlich. (CL1)
- Die Darstellung der Informationen im Display ist klar. (CL2)
- Das Display der Anwendung wirkt aufgeräumt. (CL3)
- Es ist einfach, die Informationen zu finden, die ich benötige. (CL4)

B) Final UXAR-CT Questionnaire

UXAR-CT Questionnaire

Overall, I am satisfied with the support provided by the application for my learning tasks.	☹ ○ ○ ○ ○ ○ ○ ○ ☺
Using the application for learning is practical.	☹ ○ ○ ○ ○ ○ ○ ○ ☺
The application helps me to learn faster.	☹ ○ ○ ○ ○ ○ ○ ○ ☺
The application improves my learning and work performance.	☹ ○ ○ ○ ○ ○ ○ ○ ☺
It is easy/simple to learn how to use the application.	☹ ○ ○ ○ ○ ○ ○ ○ ☺
The information in the application is easy to understand.	☹ ○ ○ ○ ○ ○ ○ ○ ☺
The operation of the application is logical.	☹ ○ ○ ○ ○ ○ ○ ○ ☺
The application is easy to control.	☹ ○ ○ ○ ○ ○ ○ ○ ☺
I always have control over the application at every step.	☹ ○ ○ ○ ○ ○ ○ ○ ☺
It is easy to find your way around the application.	☹ ○ ○ ○ ○ ○ ○ ○ ☺
The application helps me to learn.	☹ ○ ○ ○ ○ ○ ○ ○ ☺
It is a great advantage to use the application when learning.	☹ ○ ○ ○ ○ ○ ○ ○ ☺
I find the application useful for learning.	☹ ○ ○ ○ ○ ○ ○ ○ ☺
The information on the display is clearly laid out.	☹ ○ ○ ○ ○ ○ ○ ○ ☺
The information on the display is clear.	☹ ○ ○ ○ ○ ○ ○ ○ ☺
It's easy to find the information I need	☹ ○ ○ ○ ○ ○ ○ ○ ☺

User Experience Evaluation of Augmented Reality: A Systematic Literature Review

Stefan Graser
stefan.graser@hs-rm.de
Center for Advanced E-Business Studies
RheinMain University of Applied Sciences
Wiesbaden, Germany
ORCID: 0000-0002-5221-2959

Felix Kirschenlohr
Schwarz IT KG
Neckarsulm, Germany

Stephan Böhm
stephan.boehm@hs-rm.de
Center for Advanced E-Business Studies
RheinMain University of Applied Sciences
Wiesbaden, Germany
ORCID: 0000-0003-3580-1038

***Abstract*—Due to technological development, Augmented Reality (AR) can be applied in different domains. However, innovative technologies refer to new interaction paradigms, thus creating a new experience for the user. This so-called User Experience (UX) is essential for developing and designing interactive products. Moreover, UX must be measured to get insights into the user's perception and, thus, to improve innovative technologies. We conducted a Systematic Literature Review (SLR) to provide an overview of the current research concerning UX evaluation of AR. In particular, we aim to identify (1) research referring to UX evaluation of AR and (2) articles containing AR-specific UX models or frameworks concerning the theoretical foundation. The SLR is a five-step approach including five scopes. From a total of 498 records based on eight search terms referring to two databases, 30 relevant articles were identified and further analyzed. Results show that most approaches concerning UX evaluation of AR are quantitative. In summary, five UX models/frameworks were identified. Concerning the UX evaluation results of AR in Training and Education, the UX was consistently positive. Negative aspects refer to errors and deficiencies concerning the AR system and its functionality. No specific metric for UX evaluation of AR in the field of Training and Education exists. Only three AR-specific standardized UX questionnaires could be found. However, the questionnaires do not refer to the field of Training and Education. Thus, there is a lack of research in the field of UX evaluation of AR in Training and Education.**

Keywords–User Experience (UX); UX Evaluation; (Mobile) Augmented Reality (M)AR; Systematic Literature Review (SLR).

I. Introduction

Over the last decades, AR[1] as an innovative technology has emerged in different domains. AR, therefore, enhances the real environment with digital information and 3D data using different devices [1]. To separate AR from other technologies, [1] defined three characteristics of the technology: Therefore, AR (1) combines reality with virtuality, (2) creates an interaction of both in real-time, and (3) registers digital content in 3-D [1]. Due to technological development, AR can be easily deployed into application fields of daily life, such as Education, Entertainment, or Medicine [2]. Concerning the field of education, AR provides the potential for improving teaching as well as learning [3][4]. The field can be further divided into academic teaching and Corporate Training, which refers to training in a corporate environment.

Applying AR enhances both learning and teaching and thus, a benefit in education is created. Educational content can be experienced in a new way due to multimodality, interactivity, and engagement. In this context, different learning effects occur. As a learning effect, we understand the change in knowledge, skills, or abilities resulting from learning activities by applying AR. For instance, memory ability, learning motivation, or learning effectiveness can be enhanced [4][5].

However, innovative technologies always relate to new interaction paradigms and, thus, a new experience for the user [6][7]. UX refers to the subjective impression of the user towards a product, system, or service [8]. A positive UX is an essential success factor for interactive products [9]. For this, it is crucial to consider the user's perception of the respective product. Thus, the UX must be measured to provide insights into improving AR and creating a positive UX [10].

In this regard, this article focuses on the UX evaluation of AR. In particular, we analyzed (1) research articles containing an empirical UX evaluation of AR to provide insights into the status quo of UX evaluation. Moreover, we (2) analyzed articles containing a UX model or framework[2] in relation to AR to provide the theoretical foundation. Therefore, a Systematic Literature Review (SLR) was conducted to provide these insights into the current state of research in this field. The conducted SLR follows a five-step approach, including five defined scopes based on the Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) guidelines [5][11].

This article is structured as follows: In Section II, we introduce the related work referring to UX and AR. Moreover, the research objective and research questions are explained. In Section III, the methodological approach of the SLR is shown. Results are illustrated in Section IV. Lastly, a conclusion is given in Section V.

II. Related Work

A. The Concept of User Experience

UX is defined by the ISO as "a person's perceptions and responses that result from the use and/or anticipated use of a product, system or service" [8]. This rather broad description indicates that UX is a multidimensional construct of different dimensions [6][7].

In this context, Usability is strongly related to UX. Usability is defined as the "extent to which a product, system or service can be used by specific users to achieve specific goals with effectiveness, efficiency, and satisfaction in a specific context of use" [8]. Moreover, Usability is declared as a subdimension and, thus, part of the concept of UX.

[1]**Please note**: This article contains both AR and MAR. Both are considered identical and refer to the technology in general. No distinction is made in the article.

[2]In this article, the authors use *model* and *framework* synonymously.

 ISBN: 978-1-68558-193-0

Different approaches were conducted to break down the construct of UX and get a better understanding. [12] differentiated UX into pragmatic and hedonic properties. Pragmatic aspects refer to task-related, functional factors, whereas hedonic aspects describe emotional factors. This differentiation is a common perspective and many UX researchers rely on this. However, some problems concerning the quantification occur. Pragmatic qualities, such as efficiency, are task-related and, thus, can be measured. In contrast, hedonic factors refer to the emotional perception of the user. In this regard, it is quite difficult to quantify this as there is no specific underlying concept. Moreover, it depends on the specific context whether some quality is pragmatic or hedonic. Hence, it could be difficult to determine and quantify the UX in some cases [7].

Against this background, [7] developed a further distinction. For this, UX was broken down into a set of quality aspects defined as follows: "A UX quality aspect describes the subjective impression of users towards a semantically clearly described aspect of product usage or product design" [7][13].

These aspects can be further applied to quantify and measure UX and, thus, provide insights into the user's perception. This concept is the foundation of several UX metrics. Table I shows the UX quality aspects.

TABLE I. UX QUALITY ASPECTS [7].

(#)	Factor	Descriptive Question
(1)	Perspicuity	Is it easy to get familiar with the product and to learn how to use it?
(2)	Efficiency	Can users solve their tasks without unnecessary effort? Does the product react fast?
(3)	Dependability	Does the user feel in control of the interaction? Does the product react predictably and consistently to user commands?
(4)	Usefulness	Does using the product bring advantages to the user? Does using the product save time and effort?
(5)	Intuitive use	Can the product be used immediately without any training or help?
(6)	Adaptability	Can the product be adapted to personal preferences or personal working styles?
(7)	Novelty	Is the design of the product creative? Does it catch the interest of users?
(8)	Stimulation	Is it exciting and motivating to use the product? Is it fun to use?
(9)	Clarity	Does the user interface of the product look ordered, tidy, and clear?
(10)	Quality of Content	Is the information provided by the product always actual and of good quality?
(11)	Immersion	Does the user forget time and sink completely into the interaction with the product?
(12)	Aesthetics	Does the product look beautiful and appealing?
(13)	Identity	Does the product help the user to socialize and to present themselves positively to other people?
(14)	Loyalty	Do people stick with the product even if there are alternative products for the same task ?
(15)	Trust	Do users think that their data is in safe hands and not misused to harm them?
(16)	Value	Does the product design look professional and of high quality?

In the following section, we will introduce common UX evaluation and measurement approaches in UX research.

B. UX Evaluation and Measurement

Various methods measuring the UX can be found in scientific literature [14][15]. Therefore, the methods cover a wide range of different research objectives and questions. However, the methods differ in terms of the research objective and the application scenario. The following Figure 1 illustrates the most common methods referring to [15]:

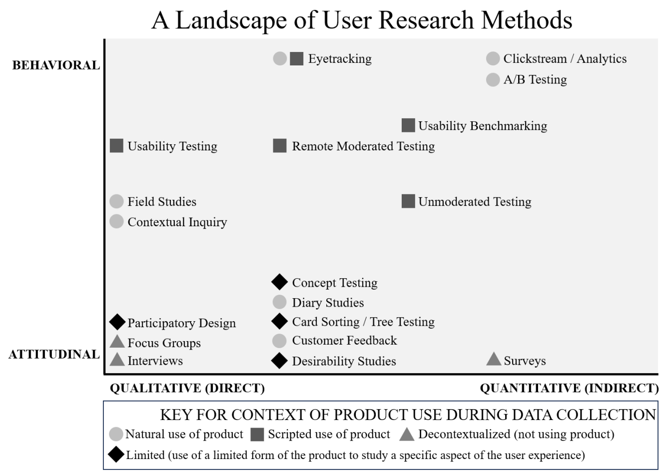

Figure 1. The most common UX Research Methods based on [15].

In general, methods are differentiated into subjective and objective evaluation. Subjective methods relate to self-reported data as direct feedback from the user, e.g., questionnaires. Objective methods refer to analytical data, e.g., eye-tracking data or time measurement. It is common to gather self-reported data as it provides direct user feedback referring to the subjective impression. Applying questionnaires is a simple, fast, and cost-efficient way of collecting self-reported data. Standardized questionnaires are, therefore, the most established method in quantitative UX research.

C. User Experience Questionnaires

As standardized UX questionnaires are the most common way of collecting self-reported data, we want to introduce them in this section. In general, UX questionnaires aim to gather data on the subjective impression of users. The structure is based on different factors, measurement items, and scales concerning the respective focus. Thus, the construct of UX is broken down by different factors. However, due to the lack of common ground concerning the construct of UX, the UX questionnaires indicate a high heterogeneity. On the one hand, factors can measure the same but are named differently. On the other hand, factors have the same designation but measure something different. Thus, existing UX questionnaires differ on the level of the measurement items and their respective factors [13][16][17].

In the literature, approximately 40 established UX questionnaires can be found [6]. Among them, the User Experience Questionnaire (UEQ) developed by [18] is the most widely used questionnaire [19]. The UEQ was based on the UX foundation in relation to [12] and consists of the following six scales divided into pragmatic and hedonic [18]:

- **Attractiveness**: Overall impression of the product. Do users like or dislike it?
- **Perspicuity**: Is it easy to get familiar with the product and to learn how to use it?
- **Efficiency**: Can users solve their tasks without unnecessary effort? Does it react fast?
- **Dependability**: Does the user feel in control of the interaction? Is it secure and predictable?

- **Stimulation**: Is it exciting and motivating to use the product? Is it fun to use?
- **Novelty**: Is the design of the product creative? Does it catch the interest of users?

The questionnaire was designed to evaluate the holistic impression of interactive products. The different factors consist of semantic differential scales and a 7-point Likert scale [18]. Further information can be found online [20].

However, other questionnaires follow another concept of quantifying the UX. There is a huge variety among the formulation of items. Moreover, no questionnaire can measure all UX factors. Thus, it is important to identify and evaluate the relevant UX factors concerning the respective evaluation object. The User Experience Questionnaire Plus (UEQ+) developed by [21] represents a modular framework that can be individualized regarding the specific evaluation context. Therefore, the UEQ+ consists of 16 UX quality aspects that can be combined to create an individual questionnaire. The UEQ+ is a modular extension of the UEQ and follows the common foundation of UX quality aspects [7] (See Section II-A). Further information can be found online [22].

In the following, we want to specify the research objective of this study.

D. Research Objective and Research Questions

This article focuses on the UX of AR. The overall research goal is to provide the current state of research concerning UX of AR. More precisely, this SLR follows two directions: (1) We aim to conduct the current state of research regarding UX evaluation. Therefore, we did not specify an application field. Besides this, we aim to collect the respective results of UX evaluation in the field of training and education as this is part of the researcher's doctoral project. (2) we analyzed research articles, including models, frameworks, or reviews in relation to the UX of AR to provide the theoretical foundation of this research topic. Against this background, we address the following research questions:

RQ1: *Which methods were applied for measuring UX in the context of AR?*

RQ2: *What theoretical models and frameworks exist concerning UX and AR?*

RQ3: *What results were conducted in UX research regarding AR in the domain of training and education?*

Based on this, the SLR was conducted. The detailed approach is declared in the following Section III. While speaking from the technology in general within this paper, the term AR also includes the different types such as AR.

III. METHODOLOGICAL APPROACH

A. Procedure

In the following, we illustrate the methodological approach used in this article. The three authors conducted the research to identify the current state of research concerning the UX of AR. Therefore, a Systematic Literature Review was conducted. The SLR follows a five-step approach (See Figure 2) including five defined scopes (See Figure 3). The processed steps include basic and advanced screening and filtering. Furthermore, a qualitative assessment based on two metrics for selecting articles to ensure the quality of records was applied. Moreover, specific articles in the field of Corporate Training were conducted. The procedure was generally based on the Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) guidelines [11]. We chose this procedure because other approaches for conducting SLR show some limitations, e.g., a lack of explicit guidelines for the quality assessment as well as the insufficient transparency of reporting intermediate results [5]. The detailed approach is shown in the following Figure 2.

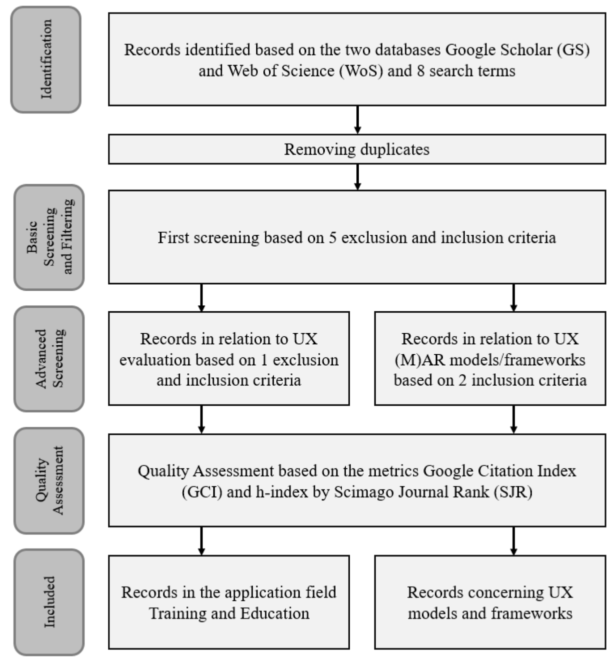

Figure 2. Methodological Approach.

In the following, the different scopes of this study are explained.

B. Scopes

For this SLR, we have defined five scopes for filtering and screening relevant records regarding the research questions. The five scopes are illustrated in the following Figure 3.

The first three stages aim to identify relevant articles referring to a UX evaluation of AR, including empirically collected data. Stage 4 introduces a specific quality assessment to ensure reliable and valid results within the respective records. Therefore, quality parameters were applied. In the last Stage 5, we particularly analyzed research conducted in the field of Training and Education. The different stages are further described below. The authors processed all the steps.

C. Stages of Literature Search

1) Stage 1: Identification: We used **Google Scholar (GS)** and **Web of Science (WoS)** for the literature search and record identification as both are two of the most common and largest databases for scientific research. Thus, both cover mostly all published research articles. We applied eight search terms to both databases. The search terms are composed of the keywords *"User Experience"* and *"Augmented Reality"* and their abbreviations as listed in the following:

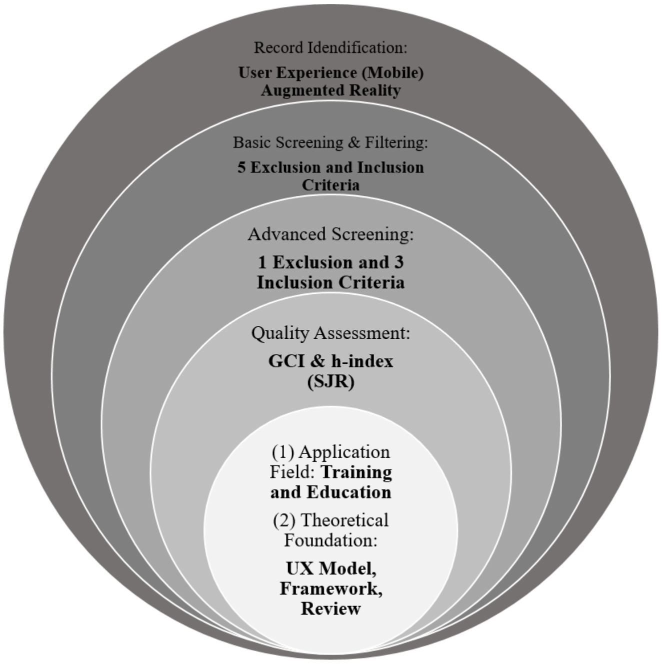

Figure 3. Scopes.

- ”User Experience Augmented Reality”
- ”User Experience Mobile Augmented Reality”
- ”UX Augmented Reality”
- ”UX Mobile Augmented Reality”
- ”UX AR”
- ”UX MAR”
- ”User Experience AR”
- ”User Experience MAR”

We have used the term *”Mobile Augmented Reality”* and its abbreviation to identify all relevant articles. We also included the abbreviations to ensure that all relevant articles were found. In relation to GS, the search terms were applied in the Advanced search performing *”allintitle: first keyword ”second keyword”*. Referring to WoS, the keywords were searched using Basic search (Title and Topic). As the focus is on UX evaluation, we assumed UX as the first keyword for the search. Thus, 16 datasets (8 search terms for both databases) were considered. The last searches were conducted on *July 31, 2023*. Based on the 16 datasets, a total number of **498** records was found.

2) Stage 2: Basic Screening and Filtering: All duplicates were deleted before basic screening and filtering as the second stage. Afterward, all records were screened based on different inclusion as well as exclusion criteria (see Table II):

TABLE II. INCLUSION AND EXCLUSION CRITERIA

Inclusion criteria	Exclusion criteria for Basic Screening
(in1) Focus on UX of AR	(ex1) Focus on VR instead of AR
(in2) Accessibility of full-text	(ex2) No accessibility of full-text
(in3) Research language English	(ex3) Written in non-English
(in4) Peer-reviewed	(ex4) Grey literature
(in5) Empirical data collection or theoretical model/framework (also SLR)	(ex5) insufficient information

For basic screening and filtering, the titles and abstracts of each record were analyzed based on the criteria *in1 - in5* and *ex1 - ex5* described in Table II. We specifically included all papers focusing on the UX of AR. We analyzed whether the full text was available or not. We only considered English language literature as the official research language. Lastly, we only included peer-reviewed records. Grey literature, such as white papers, theses, etc., were excluded. This results in a total of **223** records.

3) Stage 3: Advanced Screening: In the third stage, an advanced screening was conducted applying the criteria *in6, ex6, in7, in8* (See Table III). Therefore, the criteria *in6* and *ex6* refer to records concerning the empirical evaluation, whereas *in7* and *in8* relate to the records for the theoretical foundation.

TABLE III. INCLUSION AND EXCLUSION CRITERIA FOR ADVANCED SCREENING

Inclusion criteria	Exclusion criteria
(in6) UX/Usability evaluation goal	(ex6) Lack of focus in UX/Usability evaluation goal
(in7) UX model/framework included	
(in8) Systematic Literature Review	

In particular, the abstracts and full texts of the records were analyzed to determine whether the primary goal of the respective study in relation to the research questions was addressing UX/Usability. Moreover, we considered whether an empirical study collecting empirical data was conducted. A number of **121** can be provided.

Furthermore, we analyzed whether a UX model or framework was proposed or contained regarding the identification of the theoretical foundation. As a result, **12** articles were identified.

4) Stage 4: Quality Assessment: In stage four, we employed a qualitative assessment to identify articles with high impact in the research field. Therefore, we applied two measures - the Google Citation Index (GCI) and the h-index provided by Scimago Journal Rankings (SJR). Even though both metrics show certain limitations, they are increasingly applied for the indication of paper impact [5][23][24]. Based on these metrics, we aimed to provide qualitative results.

Firstly, all records regarding UX evaluation were processed. We made a record classification according to their type *(B = book chapter, J = Journal article, C = conference proceedings, A = none of the three)*. In the following, the types and the respective record number is shown:

- **B** = Book chapter: **6** records
- **J** = Journal article: **54** records
- **C** = Conference proceedings: **61** records

In the second step, we conducted both metrics for all records. For the h-index, we looked up the specific scores of the respective publisher concerning each article. If the h-index of the latest published issue was missing, we used the closed available score of the previous issues. Afterward, we calculated the median h-index of each type. We used the median value rather than the mean due to large ranges of the respective scores [5]. Thus, the median h-index represents the threshold for quality classification.

Concerning the GCI, we computed the *Average Citation Count (ACC)* of each record by dividing the overall citation count by the number of years [5][25]. We also calculated the median score of overall citations as the relevant threshold. Thus, all records with an ACC above the median were considered qualitative. As all records with the type **A** do not have any citations and the h-index does not apply to them, these articles were excluded. The different median scores are shown in the following Table IV:

TABLE IV. MEDIAN VALUES OF RECORDS REGARDING UX EVALUATION

Median	B	J	C
GCI	1.62	2.5	1.5
h-index	0	46	7

Both metrics with the respective thresholds were applied, resulting in a final batch of **71** records. The resulting distribution by type is divided as follows:

- **B** = Book chapter: **2** records
- **J** = Journal article: **31** records
- **C** = Conference proceedings: **38** records

All records are listed in the appendix VI.

In relation to theoretical models and frameworks, 12 records were conducted with the following distribution:

- **B** = Book chapter: **1** records
- **J** = Journal article: **3** records
- **C** = Conference proceedings: **7** records
- **A** = other type: **1** records

Due to the large discrepancy in the metrics values, the quality assessment for the eleven articles was rather difficult. However, all articles have at least five citations. Thus, the eleven articles were further analyzed.

5) Stage 5: Included records: In the final stage 5, we made a further record selection and examined all articles in relation to the application field of *Training and Education*. This was done due to the specific research focus of the authors. This results in a number of **18** records. For this, we considered the respective UX evaluation results of these articles.

Furthermore, all records concerning a theoretical foundation were analyzed. This results in a final record batch of **30**. Figure 4 shows the approach with the respective number of records in each stage.

In the following, the results are illustrated.

IV. RESULTS

This section presents the results of this SLR. The section IV-A gives an overview of general information concerning research. We examined the records based on the pattern by (1) publication year, (2) origin, (3) application field, (4) application device/hardware, (5) software, and (6) methodological approach. For (1) and (2), both records concerning UX evaluation (***n = 71***) and UX models/frameworks (***n = 12***) identified in Stage III-C4 (Stage 4) Quality Assessment were examined. For (3) - (6), only the 71 UX evaluation articles were considered.

To provide a deeper insight into the UX of AR in the field of Training and Education, as well as the theoretical foundation, the articles resulting from Stage III-C5 (Stage 5) were further considered. Section IV-B illustrates the respective UX evaluation results in the field of *Training and Education*. Therefore, the ***18*** full articles were analyzed. Lastly, Section IV-C shows all records, including a UX model, framework, or review, to provide an overview of the theoretical foundation concerning UX of AR. The respective ***12*** were analyzed for this.

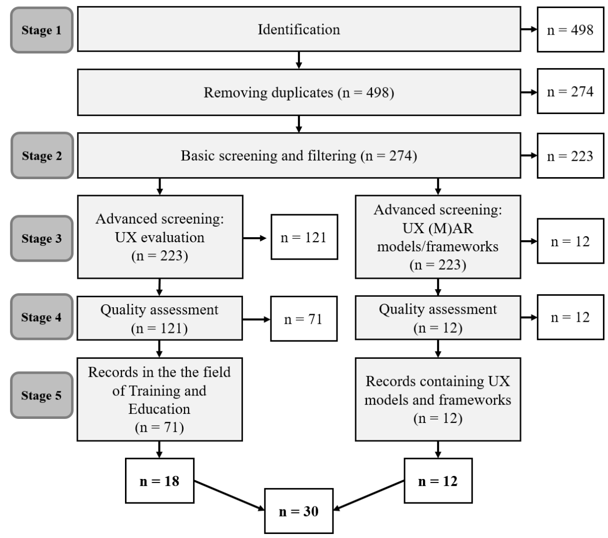

Figure 4. Methodological Approach.

A. General Information

1) Pattern by Year: The earliest publication year records (n = 5) were identified is 2011. Until 2017, record numbers were rather low. Since 2018, there has been a strong increase in articles. The highest number of publications reported in 2019 was 12. Figure 5 illustrates the number of records over the years.

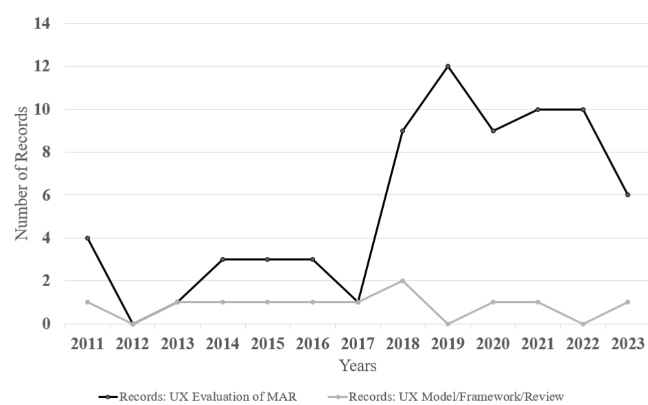

Figure 5. Number of Records Regarding the Year.

2) Pattern by Origin: Regarding the origin, the identified records are spread across 4 continents. Most studies (n = 40) were conducted by researchers from Europe, followed by Asia with a total of 24. One article can be classified as Asia and Europe. Two records are located in North America, whereas four articles are assigned to South America. No articles could be identified in relation to Africa or Australia. Figure 6 presents the paper distribution by origin.

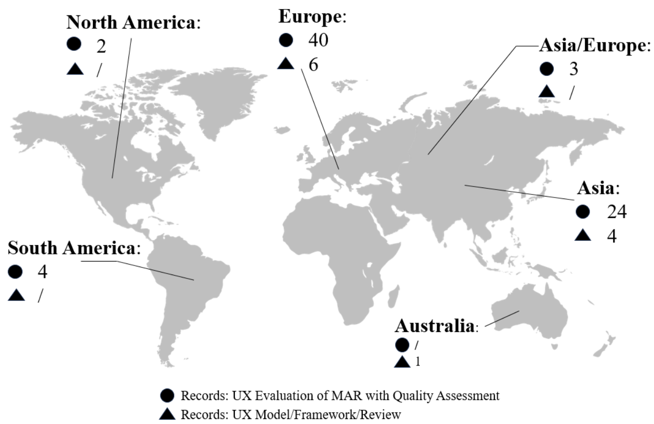

Figure 6. Records by Origin.

The 71 identified records regarding the UX evaluation of AR are considered in the following. It must be noted that some articles conducted several studies within one research article. Therefore, it is possible that the numbers of the respective results do not correspond to the numbers of the identified papers.

3) Pattern by Application Field: We categorized the records according to the specific application domain. Therefore, the researchers examined the papers and clustered the different domains into six main classes:

- Training and Education (n = 18)
- Marketing / Commercial Applications (n = 16)
- Culture Heritage / Museum (n = 18)
- Entertainment (n = 5)
- Medicine (n = 3)
- Navigation (n = 11)

Most records were collected in the field of training and education. As already described, the field can be further divided into academic teaching and Corporate Training. Therefore, ten records can be identified regarding academic teaching, whereas eight articles refer to Corporate Training scenarios.

The further major domains are Culture Heritage / Museum, with a total 18, and Marketing / Commerical Applications, with 16 identified articles. Additionally, applying AR for Navigation could be recorded eleven times. Lastly, five records in the field of entertainment and three in medicine were conducted.

4) Pattern by Application Device: Concerning AR, the technology can be applied to different hardware settings. By observing the literature, Mobile Augmented Reality often refers to Mobile Devices [5][26]. However, this is not always clear. For instance, AR glasses are also mobile in the sense of their property as they are not firmly linked to a place/position. Thus, we made a distinction between *handheld* and *head-mounted* in this study.

Results indicate that most studies have used handheld devices. This was determined for 48 articles. Android was used 28 times, and iOS 10 times as an operation system.

In comparison, 14 records applied head-mounted displays. Against this, the Microsoft HoloLens was used ten times, whereas the MagicLeap was implemented four times.

One study used a Web AR application. Eight studies did not specify the application device.

In the end, we took the respective underlying development software into account. This resulted in a very opaque overview, which was not specified for 43 articles. However, it can be noted that Unity is the most commonly used platform for prototype development, extended with different common AR development PlugIns, e.g., Vuforia, or its own programming.

5) Pattern by Methodological Approach: As described in Section II-B, various methods for UX evaluation can be found. The respective methods have different characteristics concerning their approach. Lastly, we examined the records' methodological approach to provide insights into the evaluation method. Results show that 38 articles applied purely quantitative methods. In contrast, ten records used a qualitative method. A total of 23 conducted a mixed-method approach applying both quantitative and qualitative methods. Questionnaires (n = 69), interviews (n = 19), (Usability) performance analysis (n = 2), observation logs (n = 2), NLP approaches (n = 2), Thinkin'-Aloud (n = 1), and eye-tracking (n = 1) can be listed as applied methods. For this, questionnaires are the most commonly used method.

As quantitative UX evaluation in questionnaires is the most established, we further examined the existing applied questionnaires. In summary, 40 records include individualized developed questionnaires. In contrast, 29 studies applied standardized, existing questionnaires. Regarding existing ones, eleven metrics could be identified. These include the UEQ [18], SUS [27], QUIS [28], AttrakDiff [29], SSQ [30], NASA TLX [31], TPI [32], HARUS [33][34], PSSUQ [35], TAM [36], and UTAUT [37], although the latter two originally belong to the field of technology acceptance research. Most of them were only used one to three times. Only the UEQ and the SUS have been applied more frequently. The SUS was applied eight times, whereas the UEQ was applied ten times. Therefore, it can be stated that the UEQ is the most widely used UX questionnaire.

Up to here, we considered relevant records resulting from stage four (See Section III-C4) to provide a comprehensive overview into general insights concerning the UX evaluation of AR. In the next Section IV-B, we took papers resulting from Stage 5 (See Section III-C5 into account to present details about UX evaluation in the field of Training and Education as well as the theoretical foundation.

B. User Experience Evaluation Results in Training and Education

In this section, we present results regarding the UX evaluation of AR in the field of Training and Education. In particular, ***18*** records were identified referring to Training and Education. As shown above (See Section IV-A3), the articles can be further classified into academic teaching (n = 10) and Corporate Training (n = 8). The relevant articles are shown in the Appendix (See Table V). In the following, we provide detailed results in relation to these papers.

1) Records in the Field of Corporate Training: For the field of Corporate Training, the records [38]–[45] were identified. [38], [42], and [39] conducted a quantitative method whereas [40][41][43]–[45] applied a mixed-method approach. In particular, Questionnaires (n =10), qualitative Interviews (n = 2), and Observation (n = 1) were applied as research methods. Individualized and standardized questionnaires were applied to the questionnaires. For standardized UX questionnaires, the (1) SUS [27], (2) QUIS [28], (3) NASA-TLX [31], (4) AttrakDiff [12][29], and (5) UEQ [18] were identified. Besides these, the Smart Glasses User Satisfaction (SGUS) questionnaire [46] was implemented.

Okimoto et al. conducted a study about welding simulation training with a group of twelve students. Participants perceived the applications as motivating for training due to the novelty of the technology. Introducing and using the application was reported as satisfactory. However, difficulties regarding the visual accommodation were mentioned. In particular, the initial sense of spatial depth and perception of the manual movement can be stated [38].

Helin et al. evaluated an AR system for astronauts' manual work. Results indicate that the use of AR in daily operations is usable. The system was working properly without delays. Based on the quantitative evaluation, the application was considered nearly acceptable. Furthermore, participants could accomplish their tasks quite well. It was perceived as easy to learn new tasks. The attention was captivated positively. However, the experience was perceived as not natural. Moreover, it was not easy to follow the instructions. Information seemed to be hard to read, confusing, and inconsistent. The overall rating was quite positive. Suggestions for improving the UX mainly refer to detailed information and feedback regarding the system control. Besides this, the participants indicated concern in relation to the quality and accuracy of the displayed content [38].

Alenljung and Lindblom evaluated a prototype for assembly instructions with a sample of five participants. The authors set up nine UX goals. The procedure was conducted in two rounds. Results indicate that the majority of the UX goals were not met. The problems mentioned here are not properly working AR function, the clarity of instructions, the sharpness in content projection, and the object detection [45].

Aromaa et al. applied AR in maintenance processes. A positive UX was evaluated and the system was well accepted. It was perceived as useful and supportive. Moreover, it was rated as flexible, effective, wonderful, and satisfying. The system's visual appearance was liked and easy to use [40] .

Both studies by Thomaschewski et al. evaluated an AR system supporting the temporal coordination of spatially dispersed teams. 22 participants took part in this study. In this context, Usability and UX evaluation were applied to identify the best interface. Thus, no further derivations can be reported [43][44].

Heo et al. provides insights into the UX evaluation of an AR-based automobile maintenance content application using a mobile device and the HoloLens. The study was conducted on 44 participants. In particular, the Awareness, Comfort, Functionality, and Space Perception were conducted and compared. Results show that the UX was not significantly different between the mobile device and the AR glass. Comfort was experienced more positively on the mobile device. In contrast, space perception of the AR glass showed better results [42].

Scavo et al. explored AR for telementoring. Twelve participants took part in the study. The application was perceived as greatly stimulating and playful, improving engagement. Results indicate that the interaction was intuitive, natural, and fast to learn [39].

To sum up, the results of the study report rather positive UX evaluation results concerning pragmatic as well as hedonic aspects. Thus, AR can be a benefit in a Corporate Training environment. However, it can be shown that problems often occur with the system functions of the AR applications.

2) Records in the Field of Academic Teaching: For the field of Academic Teaching, the records [47]–[56] were conducted. [50][51][53][55][57] and [52] conducted a quantitative approach. In comparison, [47]–[49] and [56] applied a mixed-method approach. Only [54] followed a qualitative approach by applying NLP techniques to analyze app reviews and ratings. Similar to the field of Corporate Training, Questionnaires (n = 10), qualitative Interviews (n = 3), and Observations (n = 1) were conducted. Applied questionnaires are individualized as well as standardized. Among the standardized UX questionnaires, the (1) UEQ [18], (2) PSSUQ [35], and (3) NASA-TLX [31] were used. Besides these, the Technology Acceptance Model (TAM) [36], the Emotions Questionnaire, the Temple Presence Inventory [32], and the Simulator Sickness Questionnaire [30] were used for evaluation.

Redondo et al. conducted a case study applying AR for spatial analysis in urban design processes. A student satisfaction survey and the Students ' Evaluation of Educational (SEEQ) questionnaire were applied for evaluation. Results indicate that AR can improve performance, shorten project development time, and promote creativity. However, hardware restrictions in object registration and high implementation costs are concluded [48].

Pribeanu and Iordache evaluated the usability of an AR-based learning scenario focusing on motivational value in a chemistry learning environment. Results show that AR application is perceived as supportive, exciting, motivating, and easy to use. The visualization and user guidance were perceived as positive. As negative aspects, the representation and augmentation of the educational content were difficult to distinguish [47].

Sarkar and Pillai considered user expectations toward AR. The authors developed expectations based on the three dimensions (1) content, (2) incentive, and (3) interaction in relation to learning with AR. Relevant characteristics concerning the expectations were defined for each of the three dimensions. These include Visual Cues, Informative, Situational Regeneration, and Dynamic for (1), Developing Interest, Cognitive Sustenance, Creative Instances, and Playfulness for (2), and Immersive, Tangible, Familiarity, and Exploratory for (3) [49].

Kazanidis and Pellas conducted a study applying AR in Mathematics. In summary, participants perceived the learning procedure using AR as positive. The AR application was perceived as visually appealing, helpful, and easy to use. Moreover, participants considered it to be exciting and useful. As negative aspects, a longer period of familiarization concerning AR, less effectiveness, and marker recognition, which did not work, was reported [50].

Kim-Berman et al. developed and tested a virtual tooth identification test. The evaluation shows problems in viewing and manipulating the AR application. Moreover, a high loading time and battery consumption could be determined. Nevertheless, the application was evaluated as effective [51].

Smaragdina et al. studied the UX of computer graphics educational comics applying markerless augmented reality. Evaluation results show a positive UX among all six scales of the UEQ. Pragmatic qualities were rated higher than hedonic properties [52].

Vrellis et al. applied UX and technology acceptance measures to evaluate an AR application for science literacy. Results indicate moderate spatial presence, low simulator sickness, and high acceptance as well as satisfaction [53].

Dominguez Alfaro and Puyvelde investigated the UX of AR apps by analyzing app reviews. Results show that technical issues, features, and user instructions must be improved [54] .

Liu et al. evaluated a web-based AR learning tool. Study results indicate that a positive attitude toward the technology enhances the experience. Furthermore, the factors Perceived ease of use, Attitudes toward technology use, Need for technological pedagogical content knowledge, Experience with technology–Traditional, Behavioral intention–Traditional, and Behavioral intention–Innovative show a positive significant effect on the UX [55].

Lastly, Santana et al. took the UX of a learning app into account, investigating the overall satisfaction, technological acceptance, mental workload, and emotional response. Therefore, the UTAUT and the PSSUQ were applied. Observations show that the application improves learning quality. Moreover, high engagement and satisfaction could be determined [56].

In summary, AR applications in academic teaching indicate positive UX evaluation results. Therefore, both pragmatic as well as hedonic qualities are perceived as positive. Overall, learning and teaching activities can be enhanced. In general, deficiencies and errors in the system features and functionalities of the applications were cited as negative. However, a high potential is cited among all records.

C. *Theoretical Foundation concerning UX of AR*

The second research objective was the presentation of the theoretical foundation. We examined records, including models and frameworks concerning UX of AR. We further included relevant reviews or studies dealing with the theoretical foundation to provide a comprehensive overview. In summary, ***12*** records could be identified. ***6*** articles contain specific UX models and frameworks whereas ***6*** records refer to reviews or papers conducting general insights. The results are presented in the following section.

1) UX Models and Frameworks: In summary, [58]–[63] contain a UX model or framework.

Ritsos et al. developed a theoretical evaluation framework aiming to measure important aspects of the development of AR applications in a standardized way. The framework was classified into different categories input (visual, auditory, tactile, and kinæsthetic), output (visual, auditory, and haptic), context awareness, sense of immersion, health, safety and integrity, and privacy and security [58]. The framework is shown in the following Figure 7:

Irshad and Rambli presented an early framework focusing on the design and evaluation of the UX of AR in 2015 [59]. Based on this, the authors developed the multilayered conceptual framework for the enhancement of the UX of AR. The proposed framework put together the relevant factors for designing AR applications. The framework introduces the four layers *AR Product/Service Features*, *Time*, *Specific Context*, and *UX* [60]. The framework is illustrated in Figure 8

Irshad et al. further proposed a UX Design and Evaluation framework concerning MAR. The evaluation differs between instrumental and non-instrumental quality attributes. Furthermore. the specific domain of the AR application is considered [64]. Figure 9 shows the evaluation framework.

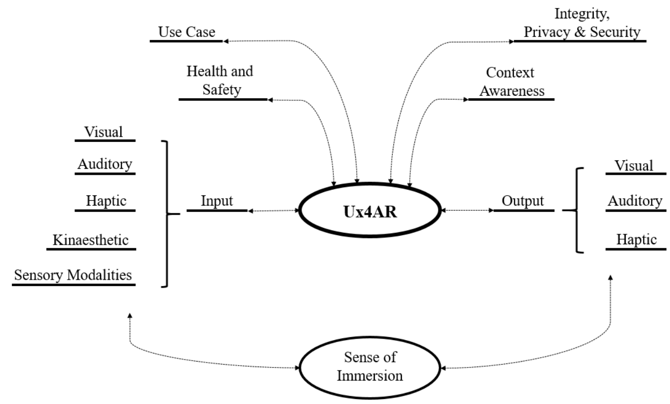

Figure 7. UX4AR Theoretical Framework Mapping by [58].

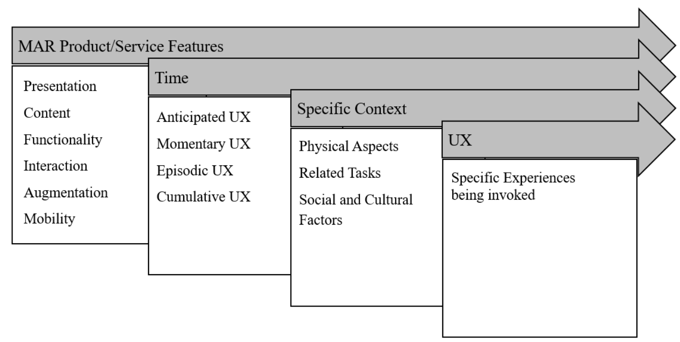

Figure 8. Multilayered conceptual framework for enhancing the UX of MAR products by [60].

Han et al. developed a UX model in the context of urban heritage tourism [65]. The authors identified relevant factors within this application field and extended the UX model by [12].

Lastly, Ahmad et al. developed a preliminary model focusing on the emotional UX applying the *Kansei* engineering approach. The so-called Augmented Reality Mobile Application Design (ARMAD) model consists of the three components *User Interface Design*, *Interaction Design*, and *Content Design*. The aim was to achieve a pleasurable design for the user [63]. The model is presented in the following Figure 10:

Besides this, [26] reviewed MAR studies from a UX perspective. As a result, the authors identified the four relevant records [12][66]–[68]. However, a distinction must be made between these papers. [12] and [67] were generally developed without reference to AR. For example, the framework developed by [12] is one of the most popular ones in UX research and, e.g., the base for the quantification of UX in some metrics. Hence, both UX frameworks can be applied to different contexts. In contrast, the frameworks developed by [66] and [68] refer specifically to MAR. Thus, we only discuss the latter two.

Perritaz et al. developed a framework focusing on the deployment of real-time concerning AR. In particular, the authors introduced the factor *Quality of Experience* consisting of the variables real-time adaption, frame rate, and image size. Their proposed framework aims to improve the *Quality of Experience* and, thus, the UX by measuring these variables

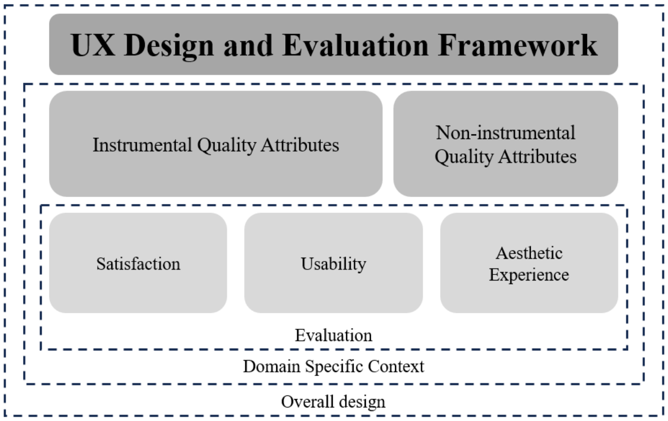

Figure 9. UX Design and Evaluation Framework for MAR by [64].

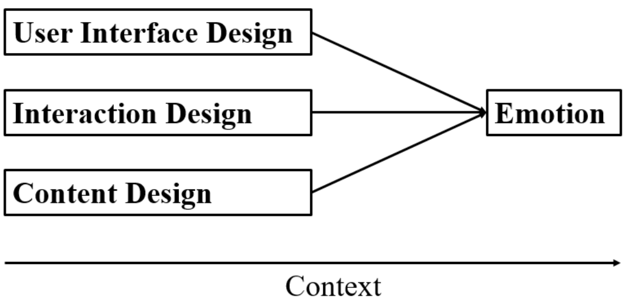

Figure 10. Augmented Reality Mobile Application Design (ARMAD) Model by [63].

Figure 11. Early framework by [66].

and improving the collaboration in real time. This can be considered a more mathematical approach [68].

Furthermore, [66] presents an early framework based on different layers of user expectations that people have with new technologies such as MAR. The work-in-progress framework was based on the four layers *Must-Be Expectations*, *Social and Societal Norms*, *Experience-Based Assumptions*, and *Desires*. The layers describe the origin of the different user expectations towards technology. Figure 11 illustrates the early framework.

As a result, only a limited number of models can be presented. The models have different characteristics in terms of focus. Concerning a common base, the UX model by [12] was applied and extended. Moreover, no model provides a combination of relevant system characteristics of AR and UX factors based on relevant literature concerning training and education. The following discusses the ***6*** relevant review records.

2) Review Paper: To sum up, [5][69]–[73] provide insights into UX of AR as a review paper.

Arifin et al. examined existing research concerning UX metrics for AR applications. Additionally, the authors specifically considered the field of Education as there are no standard measurements. Results showed that there is no metric measuring the UX from AR applications in the field of Education [69].

Irshad and Ramdbli conducted an initial study concerning the UX of MAR. The authors performed a further review presenting a comprehensive overview of the advances of MAR from a UX perspective [72]. Therefore, nine UX studies of AR and four studies in relation to a UX framework for MAR were identified (See Section IV-C1. The authors identified a lack of user research in this field [70].

Ghazwani and Smith examined the AR interaction from a UX perspective. It is argued that three aspects are relevant: (1) user type, (2) user interface, and (3) virtual content [71].

Law and Heintz conducted an SLR on the topic of AR application in education from a usability and UX perspective. Results show a lack of innovative AR-specific usability/UX evaluation methods. We will refer to this paper again in a follow-up chapter (See Section IV-B) [5].

Lastly, we want to present [73] summarizing the results of past studies in this article. The article presents an overview of User Expectations that are relevant for MAR. In summary, 16 User Expectations (categories) were identified in previous studies [74]. The Expectations are classified into the six classes *(1) Instrumental Experiences (IE)*, *(2) Cognitive and Epistemic Experiences (CEE)*, *(3) Emotional Experiences (EE)*, *(4) Sensory Experiences (SeE)*, *(5) Social Experiences (SoE)*, and *(6) Motivational and Behavioral Experiences (MBE)*. The classes and categories regarding MAR were identified in The 16 categories are listed below [73]:

- (1) Empowerment (IE)
- (2) Efficiency (IE)
- (3) Meaningfulness (IE)
- (4) Awareness (CEE)
- (5) Intuitiveness (CEE)
- (6) Amazement (EE)

- (7) Surprise (EE)
- (8) Playfulness (EE)
- (9) Liveliness (EE)
- (10) Captivation (SeE)
- (11) Tangibility and Transparency (SeE)
- (12) Collectivity and Connectedness (SoE)
- (13) Privacy (SoE)
- (14) Inspiration (MBE)
- (15) Motivation (MBE)
- (16) Creativity (MBE)

The author further provides practical instruments for the evaluation. Therefore, summative and formative measures concerning different experience categories were presented. The summative statements refer to the overall evaluation of the MAR application and are more general. In comparison, the formative statements are detailed and focus on the quality of specific features and their direct influence on the UX. For further details concerning the statements, we refer to [73].

D. User Experience Augmented Reality Questionnaire

As shown, quantitative UX measurement is the most widespread. In particular, the application of standardized questionnaires is a common way to measure the UX as it is fast, simple, and cost-efficient. Different standardized UX questionnaires can be found in scientific literature [6]. However, the questionnaires have different characteristics, structure, and focus. Thus, not every questionnaire is applicable to AR in CT. Concerning literature, three standardized UX questionnaires evaluating AR can be found. The three questionnaires are listed in the following:

- (1) **Handheld Augmented Reality Usability Scale (HARUS)** [34][33]
- (2) **Augmented Reality Immersion (ARI) Questionnaire** [75]
- (3) **Customizable Interactions Questionnaire (CIQ)** [76]

These three AR-specific questionnaires already clearly show the differences in structure and focus. The Handheld Augmented Reality Usability Scale (HARUS) developed by [33][34] focuses on the usability evaluation of handheld AR devices. To be more precise, it considers the ergonomic and perceptual issues of handheld AR issues. The structure is based on the two factors *manipulability* and *comprehensibility*. *Manipulability* refers to the ease of handling the device, whereas *comprehensibility* refers to the ease of understanding the presented information. Each factor comprises eight items based on a seven-point Likert scale (1 is “strongly disagree” and 7 is “strongly agree”) [33][34].

The Augmented Reality Immersion Questionnaire (ARI) was developed by [75] focusing on immersion in location-aware AR settings. Immersion is described as cognitive and emotional absorption during the use of interactive media such as AR. The questionnaire consists of the three factors *Engagement*, *Engrossment*, and *Total immersion*. These factors are further broken down into six sub-factors: *Interest*, *Usability*, *Emotional attachment*, *Focus of attention*, *Presence*, and *Flow* with a total of 21 items based on a seven-point Likert scale [75].

The Customizable Interactions Questionnaire (CIQ) measures the subjective impression focusing on the quality of interaction with objects in AR scenarios. The questionnaire developed by [76] contains five factors *Quality of Interactions*, *Comfort*, *Assessment of Task Performance*, *Consistency with Expectation*, and *Quality of the Sensory Enhancements* with a total of 17 items based on a five-point Likert scale [76].

Based on this, the heterogeneity of the different questionnaires emerges. Up to now, no other domain-specific standardized UX questionnaire regarding AR can be found. Moreover, no questionnaire for the Training and Education scenarios could be found.

E. User Experience and Learning Effect

We finally want to present the state of research referring to UX evaluation and learning effect in relation to AR in Training and Education. However, only a little research on this topic can be found [5][77].

Law and Heintz conducted an SLR on the usability and UX of AR applications for K12 education. Within this research, the authors also examined records referring to the relationship between UX and the learning effect. Results show that only a few studies describe the relationship between UX and learning effect qualitatively. Only [78] computed the correlation between both factors. For this, low, insignificant correlations could be identified [5].

Knowledge/skill-specific tests and systematic observations were applied as common methods for learning effect measurement.

Additionally, [77] can be stated concerning this topic. Within this study, MAR was applied as a learning solution for vocational students on a topic related to computer network devices. Results show that MAR improves learning effectiveness while implementing a positively evaluated prototype referring to the UX. However, the UX and learning effectiveness were considered separately.

No model/framework addressing the relationship between UX and learning effects could be identified. A lack of research can be stated.

V. Discussion and Conclusion

This article presents a Systematic Literature Review presenting detailed insights into the UX of AR. In particular, we aimed to provide results concerning the UX evaluation of AR in general, especially in the field of training and education, as well as the theoretical foundation. The SLR was based on a five-step approach with five defined scopes. The procedure was derived from the PRISMA guidelines [5][11]. In summary, 498 records referring to eight search terms applied in the two databases, Google Scholar and Web of Science, were identified. As a result, 71 articles regarding the UX evaluation of AR were analyzed after four exclusion stages. In the fifth stage, we identified 18 records referring to the field of Training and Education and 11 articles regarding the theoretical foundation. The main implications of this study concerning the research questions are discussed in the following.

A. Implications

In this Section, we discuss the proposed research questions (See Section II-D) by referring to the results (See Section IV of this SLR. Based on the results, we want to derive practical implications for the UX evaluation of AR. Moreover, we address topics related to UX evaluation.

1) Research Question: Which methods were applied for measuring UX in the context of AR?: No established method measuring the UX for AR in the field of Training and Education could be identified. Only three AR-specific standardized UX questionnaires could be identified. However, none of the questionnaires specifically focuses on the field of Training and Education. They are rather heterogeneous, focusing on different subsets of the UX. Thus, there is a lack of research concerning UX measurement approaches for AR in this application field.

In general, the most widely used methods are quantitative. More precisely, questionnaires are most commonly applied. For this, both standardized and individualized UX questionnaires were used. It is common to apply more than one questionnaire to expand the study by examining several UX subsets. However, a purely quantitative approach also entails limitations in relation to the evaluation results. When using questionnaires, purely numerical evaluation results are provided. This results in limitations in the interpretability of the results for the researchers. Thus, it is common to conduct a mixed-method approach. Quantitative approaches are often extended by qualitative methods, e.g., interviews. This is being done to obtain the most comprehensive results regarding user perception and to overcome the problem with purely numerical data.

In this regard, we want to raise another line of thought. The various quantitative and qualitative methods are, of course, valid on their own. However, there are no findings as to which specific methods work best together. Furthermore, it is also unclear how the results can be interpreted together and how they fit together. Moreover, the evaluation results are data from the user's perspective. This illustrates that the results require further interpretation by the product designers and developers to derive useful improvement suggestions. Up to now, there has been no approach addressing these points. Nevertheless, this would be an interesting idea to be considered. An approach based on combining quantitative and qualitative results might be assessed. In addition, a bridge could be created between the user's perspective and the developer's perspective. This would enable the developer to derive specific improvement options more quickly and easily based on the results from the user's perspective.

Practical Implications regarding RQ1:

As UX is a multidimensional construct, it is important to determine the respective evaluation objective. Based on this, the evaluation method must be selected. As shown in the related work (See Section II), it is common to apply standardized questionnaires. Due to the heterogeneity of both holistic and AR-specific existing questionnaires, it is essential to choose the best-fitting questionnaire concerning the evaluation objective. Against this background, customizable evaluation frameworks, such as the UEQ+ [22], could be helpful as researchers can select the relevant scales. Moreover, results should be interpretable and categorizable within the field. In other words, the result should be in the form of a UX score that allows a conclusion to be drawn about users' perceptions. For this, a questionnaire benchmark is useful [79][80]. To gather deeper insights, applying further qualitative methods is common and useful.

2) Research Question: What theoretical models and frameworks exist concerning UX and AR?: To sum up, six theoretical UX models/frameworks regarding the UX of AR could be identified. It must be noted that the different approaches refer to different focuses. Moreover, none of the frameworks refer to the same foundations. However, a common line within the research can be shown at [66][73][74] and [26][59][60][64]. [66] refers to one of the most established theoretical foundations concerning a general perspective toward UX developed by [12]. Research provided by [64] is partly based on [66]. Thus, the UX model by [12] is most commonly used as a foundation. However, no model exists combining relevant AR system characteristics and UX factors based on relevant literature regarding training and education.

Practical Implications regarding RQ2: The models can serve as an explanatory concept. Moreover, such a model can be useful in UX measurement [81]. However, no established model exists addressing AR and UX in combination. Research on both UX in general and UX of AR refers to the UX model by [12]. Established UX measurement approaches, e.g., the UEQ [18][20], are based on this model. Thus, this model can be considered as common ground towards UX theory.

3) Research Question: What results were conducted in UX research regarding AR in the domain of training and education?: Consistent results regarding the UX evaluation of AR can be seen in the field of training and education. The UX evaluation is often individualized, and methods are applied with a specific focus regarding the respective research objective. The UX is perceived consistently as positive. Both pragmatic and hedonic UX quality aspects are evaluated as highly positive. Negative evaluation results refer to errors and deficiencies concerning the system or functionality of AR. Among all records, applying AR in training and education is indicated to have great potential.

Practical Implications regarding RQ3:

The main issue is the functionality of the AR system, which was also investigated in previous research [82]. Errors and deficiencies are perceived as negative. One of the reasons for this is that, in most cases, users are using AR for the first time. Thus, practitioners should focus on an error-free running system.

4) UX Evaluation and Learning Effect: In recent years, research has been conducted to analyze the learning effects of using AR in Training and Education. As already shown, a positive UX is an essential success factor for interactive products such as AR. Thus, there could be a relation between the UX and the resulting learning effects. However, only a little research on this topic can be found [5][77].

[5] conducted an SLR on the usability and UX of AR applications for K12 education. Within this research, the authors also examined records referring to the relationship between UX and the learning effect. Results show that only a few studies describe the relationship between UX and learning effect qualitatively. Only [78] computed the correlation between both factors. For this, low, insignificant correlations could be identified.

Knowledge/skill-specific tests and systematic observations were applied as common methods for learning effect measurement.

Additionally, [77] can be stated concerning this topic. Within this study, MAR was applied as a learning solution for vocational students on a topic related to computer network devices. Results show that MAR improves learning effectiveness while implementing a positively evaluated prototype referring to the UX. However, the UX and learning effectiveness were considered separately.

No model/framework addressing the relationship between UX and learning effects could be identified, indicating a lack of research.

To sum up, research about the quantification of the potential of applying AR in Training and Education is rather rudimentary. Only a little research was done examining the correlation between a positive UX and learning effects resulting from the application of AR. [78] as the only study that computed the correlation between the UX and learning effects. As a result, a low, insignificant effect was conducted. Thus, it has not yet been researched whether a positive UX of an AR application is related to learning effects. A lack of research can be noted.

5) UX Evaluation and Generative Artificial Intelligence: Due to its current relevance, we would also like to address the topic of Generative Artificial Intelligence (GenAI). The rapid development of Large Language Models (LLMs), e.g. ChatGPT [83], impacts various research fields, including UX research. Based on their structure, LLMs show a strong ability to understand and generate natural language. Such models are useful in deep learning and natural language processing tasks. Thus, LLMs enable new opportunities to enhance, support, and automate activities along the research process. Results from the latest research show a considerable potential for applying GenAI in UX research [17][84][85]. Future research, therefore, shall further investigate the possibility of using GenAI in the UX research field.

B. Limitations

In this research, some limitations must be drawn. A severe limitation is that all data analysis is done by the researchers. Furthermore, Google Scholar was chosen as one database that has no quality control. This may result in the inclusion of irrelevant, as well as gray literature [86][87]. However, with the different scopes and stages (See Section III), we declare that we have filtered relevant records. Further databases may be investigated. However, it is questionable whether further relevant articles could be found through this, as Google Scholar is the largest database.

Moreover, we have to state the literature search conducted in July 2023. Considering the development of publications over the years, further relevant articles may be published since then. It can be also discussed whether all relevant articles were identified using the eight formulated search terms.

In addition, the Quality Assessment (See Section III-C4) must be mentioned. The record number is rather low. In contrast, the ranges between the different records concerning the quality criteria were broad. This causes the threshold to become less meaningful. Moreover, we were not able to determine a threshold for the articles regarding the theoretical foundation (See Section III-C5).

Lastly, it must be declared that all data analysis was performed by the researchers. Finally, we want to provide insights into future research.

VI. Outlook and Future Research

To conclude, all research questions were answered, and implications were drawn. Regarding the respective implications, three research gaps in the domain UX of AR were identified. This leads us to the limitations and the outlook of this SLR. Lastly, we want to derive a future research agenda based on the results of this study.

This SLR provides a comprehensive overview of AR's UX. We presented the current state of research and outlined research gaps within this field. Overall, we want to emphasize the lack of approaches within the domain of Training and Education. It is important to gather insights into the UX of the AR applications to improve both the system and the specific experience. Therefore, it is important to develop and apply suitable measurement methods. Based on this, the following aspects for future research can be concluded:

It is essential to develop and validate suitable models and frameworks that incorporate both the system characteristics of AR and UX factors. Based on such models, suitable measurement methods and metrics can be derived.

Besides this, the learning effects of the AR application should be investigated to understand the benefit of this technology in Training and Education. Bringing both together and investigating the relationship between the UX and the learning effects could be essential for designing and developing innovative teaching and learning applications. Lastly, the rapid development of GenAI must be taken into account. Applying GenAI in UX research can enhance the research process and provide benefits for researchers. These aspects should be considered in future research.

Appendix

TABLE V. 18 RECORDS IN THE FIELD OF TRAINING AND EDUCATION.

Author	Focus	Year	Source
Okimoto et al.	Corporate Training	2015	[38]
Scavo and Wild	Corporate Training	2015	[39]
Aromaa et al.	Corporate Training	2017	[40]
Helin et al.	Corporate Training	2018	[41]
Heo et al.	Corporate Training	2018	[42]
Thomaschewski et al.	Corporate Training	2020	[43]
Thomaschewski et al.	Corporate Training	2021	[44]
Alenljung and Lindblom	Corporate Training	2021	[45]
Pribeanu and Iordache	Academic Teaching	2010	[47]
Redondo et al.	Academic Teaching	2011	[48]
Sarkar and Pillai	Academic Teaching	2019	[49]
Kazanidis and Pellas	Academic Teaching	2019	[50]
Kim-Berman et al.	Academic Teaching	2019	[51]
Smaragdina et al.	Academic Teaching	2019	[52]
Vrellis et al.	Academic Teaching	2020	[53]
Domínguez Alfaro and Van Puyvelde	Academic Teaching	2021	[54]
Liu et al.	Academic Teaching	2022	[55]
Santana and Rossi	Academic Teaching	2022	[56]

TABLE VI. 71 IDENTIFIED RECORDS.

Author	Year	Source
Olsson, Thomas et al.	2011	[74]
Kerr, Steven J. et al.	2011	[88]
Redondo, Ernesto et al.	2011	[48]
Pribeanu, Costin; Iordache, Dragos Daniel	2011	[47]
Dhir, Amandeep; Al-kahtani, Mohammed	2013	[89]
Rehrl, Karl et al.	2014	[90]
Irshad, Shafaq; Rambli, Dayang Rohaya Bt Awang	2014	[91]
Li, Xiao; Xu, Bo	2014	[92]
Okimoto, Maria Lucia L. R. et al.	2015	[38]
Higgett, Nick et al.	2015	[93]
Scavo, Giuseppe; Wild, Fridolin	2015	[39]
Kamilakis, Manousos et al.	2016	[94]
Irshad, Shafaq; Rambli, Dayang Rohaya Awang	2016	[95]
Seppälä, Kaapo et al.	2016	[96]
Aromaa, Susanna et al.	2017	[40]
Rashid, Zulqarnain; Pous, Rafael	2017	[97]
Dirin, Amir; Laine, Teemu H.	2018	[98]
Cheng, Kun-Hung	2018	[99]
Han, Dai-In et al.	2018	[62]
Helin, Kaj et al.	2018	[41]
Hammady, Ramy et al.	2018	[100]
Irshad, Shafaq et al.	2018	[61]
Sekhavat, Yoones A.; Parsons, Jeffrey	2018	[101]
Heo et al.	2018	[42]
Alavesa, Paula; Pakanen, Minna	2018	[102]
Jakobsen, Christian L. et al.	2018	[103]
Stumpp, Stefan et al.	2019	[104]
Kim-Berman, Hera et al.	2019	[51]
DAVIDAVIČIENĖ,Vida et al.	2019	[105]
Marques, Bernardo; Carvalho, Raphael	2019	[106]
Andri, Chairil; Alkawaz, Mohammed Hazim	2019	[107]
Sarkar, Pratiti; Pillai, Jayesh S.	2019	[49]
Kazanidis, Ioannis; Pellas, Nikolaos	2019	[50]
Satti, Fahad Ahmed et al.	2019	[108]
Cauchi, Mattea; Scerri, Daren	2019	[109]
Smaragdina, Azhar Ahmad et al.	2019	[52]
Vrellis, Ioannis et al.	2020	[53]
Irshad, Shafaq et al.	2020	[110]
Brata, Komang Candra; Liang, Deron	2020	[111]
Wang, Lei; Lv, Meiyu	2020	[112]
Harrington, Maria C. R.	2020	[113]
Mikropoulos, Tassos A. et al.	2020	[114]
García Münzer, M.	2020	[115]
Thomaschewski, Lisa et al.	2020	[43]
Thomaschewski, Lisa et al.	2020	[44]
Domínguez Alfaro, Jessica Lizeth; Van Puyvelde, Peter	2021	[54]
DAVIDAVIČIENĖ, Vida et al.	2021	[116]
Navarro, Isidro et al.	2021	[117]
Jang, Yeonju; Park, Eunil	2021	[118]
Alenljung, Zackarias; Lindblom, Jessica	2021	[45]
Balani, Manisha Suresh; Tümler, Johannes	2021	[119]
Verhulst, Isabelle; Woods, Andy	2021	[120]
Barros et al.	2021	[121]
Syahidi, Aulia Akhrian; Tolle, Hermann	2021	[57]
Kandil, Ayman et al.	2021	[122]
Ku, Gordon Chih-Ming; Shang, I-Wie	2021	[123]
Pamparau, Christian; Vatavu, Radu-Daniel	2022	[124]
Guevara Aparicio, Ricardo Alfredo et al.	2022	[125]
Alvarez, Marina; Toet, Alexander	2022	[126]
Sudipa, Gede Iwan et al.	2022	[127]
Karimah, Hasna et al.	2022	[128]
Santana, Ronny; Rossi, Gustavo	2022	[56]
Luo, Yan et al.	2022	[129]
Liu, Enrui et al.	2022	[55]
Xue, Liangchao; Parker, Christopher J.	2022	[130]
De Paolis, Lucio Tommaso et al.	2022	[131]
Wolf, Julian et al.	2023	[132]
Hu, Shan	2023	[133]
Refae, Sema et al.	2023	[134]
Dag, Kazim et al.	2023	[135]
Gan, Quehong; Liu, Zhen	2023	[136]

TABLE VII. 11 IDENTIFIED RECORDS REGARDING MODELS, FRAMEWORKS, AND REVIEWS.

Author	Focus	Year	Source
Perritaz et al.	Model/Framework	2009	[68]
Ritsos et al.	Model/Framework	2011	[58]
Olsson	Model/Framework	2014	[66]
Irshad, Shafaq; Rambli, Dayang Rohaya Awang	Model/Framework	2015	[59]
Irshad, Shafaq; Rambli, Dayang Rohaya Awang	Model/Framework	2016	[60]
Irshad, Shafaq; Rambli, Dayang Rohaya Awang	Model/Framework	2018	[64]
Ahmad Nik Azlina et al.	Model/Framework	2023	[63]
Olsson, Thomas	Review	2013	[73]
Irshad, Shafaq; Rambli, Dayang Rohaya Awang	Review	2014	[70]
Irshad, Shafaq; Rambli, Dayang Rohaya Awang	Review	2017	[72]
Arifin, Yulyani et al.	Review	2018	[69]
Ghazwani, Yahya; Smith, Shamus	Review	2020	[71]
Law, Effie Lai-Chong; Heintz, Matthias	Review	2021	[5]

REFERENCES

[1] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators & Virtual Environments*, vol. 6, pp. 355–385, 1997.

[2] D. Chatzopoulos, C. Bermejo, Z. Huang, and P. Hui, "Mobile augmented reality survey: From where we are to where we go," *IEEE Access*, vol. 5, pp. 6917–6950, 2017. DOI: 10.1109/ACCESS.2017.2698164.

[3] M. Billinghurst and A. Duenser, "Augmented reality in the classroom," *Computer*, vol. 45, no. 7, pp. 56–63, 2012. DOI: 10.1109/MC.2012.111.

[4] S. Criollo-C, D. Abad-Vásquez, M. Martic-Nieto, F. A. Velásquez-G, J. L. Pérez-Medina, and S. Luján-Mora, "Towards a new learning experience through a mobile application with augmented reality in engineering education," *Applied Sciences*, vol. 11, p. 4921, 2021.

[5] E. L.-C. Law and M. Heintz, "Augmented reality applications for k-12 education: A systematic review from the usability and user experience perspective," *International Journal of Child-Computer Interaction*, vol. 30, p. 100 321, 2021, ISSN: 2212-8689. DOI: https://doi.org/10.1016/j.ijcci.2021.100321.

[6] M. Schrepp, "A comparison of ux questionnaires - what is their underlying concept of user experience?" In *Mensch und Computer 2020 - Workshopband*, C. Hansen, A. Nürnberger, and B. Preim, Eds., Bonn: Gesellschaft für Informatik e.V., 2020. DOI: 10.18420/muc2020-ws105-236.

[7] M. Schrepp *et al.*, "On the importance of UX quality aspects for different product categories," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. In Press, pp. 232–246, Jun. 2023. DOI: 10.9781/ijimai.2023.03.001.

[8] I. O. for Standardization 9241-210:2019, *Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems*. ISO - International Organization for Standardization, 2019.

[9] M. Rauschenberger, M. Schrepp, M. P. Cota, S. Olschner, and J. Thomaschewski, "Efficient measurement of the user experience of interactive products. how to use the user experience questionnaire (ueq).example: Spanish language version," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 2, pp. 39–45, 2013.

[10] M. Schrepp, R. Otten, K. Blum, and J. Thomaschewski, "What causes the dependency between perceived aesthetics and perceived usability?," pp. 78–85, 2021.

[11] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and T. P. Group, "Preferred reporting items for systematic reviews

and meta-analyses: The prisma statement," *PLOS Medicine*, vol. 6, no. 7, pp. 1–6, Jul. 2009. DOI: 10.1371/journal.pmed.1000097. [Online]. Available: https://doi.org/10.1371/journal.pmed.1000097 [retrieved: 08/2024].

[12] M. Hassenzahl, "The thing and i: Understanding the relationship between user and product," in *Funology: From Usability to Enjoyment*, M. A. Blythe, K. Overbeeke, A. F. Monk, and P. C. Wright, Eds. Dordrecht: Springer Netherlands, 2004, pp. 31–42, ISBN: 978-1-4020-2967-7. DOI: 10.1007/1-4020-2967-5_4.

[13] M. Schrepp, "User experience questionnaires: How to use questionnaires to measure the user experience of your products?," 2021.

[14] W. B. Albert and T. T. Tullis, *Measuring the User Experience. Collecting, Analyzing, and Presenting UX Metrics*. Morgan Kaufmann, 2022.

[15] C. Rohrer, "When to use which user-experience research methods," 2022. [Online]. Available: https://www.nngroup.com/articles/which-ux-research-methods/ [retrieved: 08/2024].

[16] S. Graser and S. Böhm, "Quantifying user experience through self-reporting questionnaires: A systematic analysis of the sentence similarity between the items of the measurement approaches," in *HCI International 2023 – Late Breaking Posters*, C. Stephanidis, M. Antona, S. Ntoa, and G. Salvendy, Eds., Cham: Springer Nature Switzerland, 2024, pp. 138–145, ISBN: 978-3-031-49212-9.

[17] S. Graser, S. Böhm, and M. Schrepp, "Using ChatGPT-4 for the identification of common UX factors within a pool of measurement items from established ux questionnaires," *CENTRIC 2023, The Sixteenth International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services*, 2023. [Online]. Available: https://www.thinkmind.org/index.php?view=article&articleid=centric_2023_1_30_38008 [retrieved: 08/2024].

[18] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *HCI and Usability for Education and Work*, A. Holzinger, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 63–76, ISBN: 978-3-540-89350-9.

[19] I. Díaz-Oreiro, G. López, L. Quesada, and Guerrero, "Standardized questionnaires for user experience evaluation: A systematic literature review," *Proceedings*, vol. 31, pp. 14–26, Nov. 2019. DOI: 10.3390/proceedings2019031014.

[20] T. UEQ, "Ueq user experience questionnaire," 2018. [Online]. Available: https://www.ueq-online.org/ [retrieved: 08/2024].

[21] M. Schrepp and J. Thomaschewski, "Design and validation of a framework for the creation of user experience questionnaires," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. InPress, pp. 88–95, Dec. 2019. DOI: 10.9781/ijimai.2019.06.006.

[22] M. Schrepp, "Ueq+ a modular extension of the user experience questionnaire," 2019. [Online]. Available: http://www.ueqplus.ueq-research.org/ [retrieved: 08/2024].

[23] D. Pajić, "On the stability of citation-based journal rankings," *Journal of Informetrics*, vol. 9, no. 4, pp. 990–1006, 2015, ISSN: 1751-1577. DOI: https://doi.org/10.1016/j.joi.2015.08.005.

[24] J. Mañana-Rodríguez, "A critical review of SCImago Journal amp; Country Rank," *Research Evaluation*, vol. 24, no. 4, pp. 343–354, 2014.

[25] A. Dey, M. Billinghurst, R. W. Lindeman, and J. E. Swan, "A systematic review of 10 years of augmented reality usability studies: 2005 to 2014," *Frontiers in Robotics and AI*, vol. 5, p. 37, 2018.

[26] S. Irshad and D. R. A. Rambli, "Advances in mobile augmented reality from user experience perspective: A review of studies," in *Advances in Visual Informatics*, H. Badioze Zaman *et al.*, Eds., Springer International Publishing, 2017, pp. 466–477, ISBN: 978-3-319-70010-6.

[27] J. Brooke, "Sus: A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, Nov. 1995.

[28] J. Chin, V. Diehl, and K. Norman, "Development of an instrument measuring user satisfaction of the human-computer interface," *ACM CHIi'*, Jan. 1988.

[29] M. Hassenzahl, M. Burmester, and F. Koller, "Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität," in *Mensch & Computer 2003: Interaktion in Bewegung*, G. Szwillus and J. Ziegler, Eds. Wiesbaden: Vieweg+Teubner Verlag, 2003, pp. 187–196, ISBN: 978-3-322-80058-9. DOI: 10.1007/978-3-322-80058-9_19.

[30] K. S. B. Robert S. Kennedy Norman E. Lane and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *The International Journal of Aviation Psychology*, vol. 3, no. 3, pp. 203–220, 1993. DOI: 10.1207/s15327108ijap0303_3.

[31] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Human Mental Workload*, ser. Advances in Psychology, P. A. Hancock and N. Meshkati, Eds., vol. 52, North-Holland, 1988, pp. 139–183. DOI: https://doi.org/10.1016/S0166-4115(08)62386-9.

[32] M. Lombard, T. Bolmarcich, and L. Weinstein, "Measuring presence: The temple presence inventory," Jan. 2009.

[33] M. E. C. Santos, J. Polvi, T. Taketomi, G. Yamamoto, C. Sandor, and H. Kato, "Toward standard usability questionnaires for handheld augmented reality," *IEEE Computer Graphics and Applications*, vol. 35, no. 5, pp. 66–75, 2015. DOI: 10.1109/MCG.2015.94.

[34] M. E. C. Santos, T. Taketomi, C. Sandor, J. Polvi, G. Yamamoto, and H. Kato, "A usability scale for handheld augmented reality," ser. VRST '14, Edinburgh, Scotland: Association for Computing Machinery, 2014, ISBN: 9781450332538. [Online]. Available: https://doi.org/10.1145/2671015.2671019 [retrieved: 08/2024].

[35] J. Lewis, "Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ," vol. 2, Jan. 1992, pp. 1259–1263.

[36] F. Davis, "A technology acceptance model for empirically testing new end-user information systems: Theory and results," *Massachusetts Institute of Technology*, 1985.

[37] V. Venkatesh, M. Morris, G. Davis, and F. Davis, "User acceptance of information technology: Toward a unified view," *MIS Quarterly*, vol. 27, pp. 425–478, Sep. 2003. DOI: 10.2307/30036540.

[38] M. L. L. Okimoto, P. C. Okimoto, and C. E. Goldbach, "User experience in augmented reality applied to the welding education," *Procedia Manufacturing*, vol. 3, pp. 6223–6227, 2015, 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015, ISSN: 2351-9789. DOI: https://doi.org/10.1016/j.promfg.2015.07.739.

[39] G. Scavo, F. Wild, and P. Scott, "The ghosthands UX: Telementoring with hands-on augmented reality instruction," Jul. 2015. DOI: 10.13140/RG.2.1.5069.1682.

[40] S. Aromaa, A. Väätänen, M. Hakkarainen, and E. Kaasinen, "User experience and user acceptance of an augmented reality based knowledge-sharing solution in industrial maintenance work," Jul. 2017, pp. 145–156, ISBN: 978-3-319-60491-6. DOI: 10.1007/978-3-319-60492-3_14.

[41] K. Helin, T. Kuula, C. Vizzi, J. Karjalainen, and A. Vovk, "User experience of augmented reality system for astronaut's manual work support," *Frontiers in Robotics and AI*, vol. 5, p. 106, Sep. 2018. DOI: 10.3389/frobt.2018.00106.

[42] M.-H. Heo, D. Kim, and J. Lee, "Evaluating user experience of augmented reality-based automobile maintenance content - mobile device and hololens comparison-," *International Journal of Control and Automation*, vol. 11, pp. 187–196, Apr. 2018. DOI: 10.14257/ijca.2018.11.4.17.

[43] L. Thomaschewski, B. Weyers, and A. Kluge, "A two-part evaluation approach for measuring the usability and user experience of an augmented reality-based assistance system to support the temporal coordination of spatially dispersed teams," *Cognitive Systems Research*, vol. 68, pp. 1–17, Dec. 2020. DOI: 10.1016/j.cogsys.2020.12.001.

[44] L. Thomaschewski, B. Weyers, and A. Kluge, "Using usability and user experience scores to design an augmented reality-based ambient awareness interface to support spatially dispersed teams," Oct. 2020. DOI: 10.1109/ICHMS49158.2020.9209564.

[45] Z. Alenljung and J. Lindblom, "User experience in augmented reality: A holistic evaluation of a prototype for assembly instructions," in Jul. 2021, pp. 139–157, ISBN: 978-3-030-78226-9. DOI: 10.1007/978-3-030-78227-6_11.

[46] H. Xue, P. Sharma, and F. Wild, "User satisfaction in augmented reality-based training using microsoft hololens," *Computers*, vol. 8, p. 9, Jan. 2019. DOI: 10.3390/computers8010009.

[47] C. Pribeanu and D. Iordache, "From usability to user experience: Evaluating the educational and motivational value of an augmented reality learning scenario," *Affective, Interactive and Cognitive Methods for E-Learning Design: Creating an Optimal Education Experience*, pp. 244–259, Nov. 2011. DOI: 10.4018/978-1-60566-940-3.ch013.

[48] E. Redondo, I. Navarro, A. Riera, and D. Fonseca, "Visual interfaces and user experience: Augmented reality for architectural education: One study case and work in progress," vol. 166, Jun. 2011, pp. 355–367, ISBN: 978-3-642-21983-2. DOI: 10.1007/978-3-642-21984-9_31.

[49] P. Sarkar and J. S. Pillai, "User expectations of augmented reality experience in indian school education," in *Research into Design for a Connected World*, A. Chakrabarti, Ed., Singapore: Springer Singapore, 2019, pp. 745–755, ISBN: 978-981-13-5977-4.

[50] I. Kazanidis and N. Pellas, "Developing and assessing augmented reality applications for mathematics with trainee instructional media designers: An exploratory study on user experience," *Journal of Universal Computer Science*, vol. 25, pp. 489–514, Jun. 2019. DOI: 10.3217/jucs-025-05-0489.

[51] H. Kim-Berman, E. Karl, J. Sherbel, L. Sytek, and V. Ramaswamy, "Validity and user experience in an augmented reality virtual tooth identification test," *Journal of Dental Education*, vol. 83, s1–s8, Aug. 2019. DOI: 10.21815/JDE.019.139.

[52] A. Smaragdina, G. Ningrum, A. Mursyidun Nidhom, N. Rahmawati, M. Rusdiansyah, and A. Putra, "The user experience analysis of computer graphics educational comics (grafmic) based on markerless augmented reality," Oct. 2019, pp. 220–225. DOI: 10.1109/ICEEIE47180.2019.8981439.

[53] I. Vrellis, M. Delimitros, P. Chalki, P. Gaintatzis, I. Bellou, and T. A. Mikropoulos, "Seeing the unseen: User experience and technology acceptance in augmented reality science literacy," in *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)*, 2020, pp. 333–337. DOI: 10.1109/ICALT49669.2020.00107.

[54] J. L. Dominguez Alfaro and P. Puyvelde, "Mobile augmented reality apps in education: Exploring the user experience through large-scale public reviews," in Sep. 2021, pp. 428–450, ISBN: 978-3-030-87594-7. DOI: 10.1007/978-3-030-87595-4_32.

[55] E. Liu, S. Cai, Z. Liu, and L. Liu, "Webart: Web-based augmented reality learning resources authoring tool and its user experience study among teachers," *IEEE Transactions on Learning Technologies*, vol. PP, pp. 1–14, Jan. 2022. DOI: 10.1109/TLT.2022.3214854.

[56] R. Santana *et al.*, "Studying the user experience of an educational ar-based app for smart glasses," in May 2022, pp. 266–275, ISBN: 978-3-031-04825-8. DOI: 10.1007/978-3-031-04826-5_26.

[57] A. Akhrian Syahidi and H. Tolle, "Evaluation of user experience in translator applications (banjar-indonesian and indonesian-banjar) based on mobile augmented reality technology using the UX honeycomb method," *International Journal of Game Theory*, vol. 6, pp. 7–13, Jun. 2021. DOI: 10.21512/jggag.v6i1.7430.

[58] P. D. Ritsos, D. P. Ritsos, A. S. Gougoulis, and S. Toys, "Standards for augmented reality: A user experience perspective," 2011.

[59] S. Irshad and D. R. A. Rambli, "Preliminary user experience framework for designing mobile augmented reality technologies," in *2015 4th International Conference on Interactive Digital Media (ICIDM)*, 2015, pp. 1–4. DOI: 10.1109/IDM.2015.7547833.

[60] S. Irshad and D. R. Awang Rambli, "Multi-layered mobile augmented reality framework for positive user experience," in *Proceedings of the 2nd International Conference in HCI and UX Indonesia 2016*, ser. CHIuXiD '16, Jakarta, Indonesia: Association for Computing Machinery, 2016, pp. 21–26, ISBN: 9781450340441. [Online]. Available: https://doi.org/10.1145/2898459.2898462 [retrieved: 08/2024].

[61] S. Irshad, D. Awang Rambli, N. Intan Adhani, S. R. Mohd Shukri, and Y. Omar, "Measuring user experience of mobile augmented reality systems through non-instrumental quality attributes," in Jan. 2018, pp. 349–357, 5th International Conference, i-USEr 2018, Puchong, Malaysia, August 28–30, 2018, Proceedings, ISBN: 978-981-13-1627-2. DOI: 10.1007/978-981-13-1628-9_31.

[62] D.-I. Han, M. C. Tom Dieck, and T. Jung, "User experience model for augmented reality applications in urban heritage tourism," *Journal of Heritage Tourism*, vol. 13, Feb. 2017. DOI: 10.1080/1743873X.2016.1251931.

[63] N. A. N. Ahmad, M. Abdullah, A. M. Lokman, and A. I. H. Suhaimi, "Preliminary emotional user experience model for mobile augmented reality application design: A kansei engineering approach," *International Journal of Interactive Mobile Technologies*, vol. 17, pp. 32–46, 2023.

[64] S. Irshad and D. R. B. Awang, "A UX oriented evaluation approach for mobile augmented reality applications," in *Proceedings of the 16th International Conference on Advances in Mobile Computing and Multimedia*, ser. MoMM2018, Yogyakarta, Indonesia: Association for Computing Machinery, 2018, pp. 108–112, ISBN: 9781450364522. DOI: 10.1145/3282353.3282357. [Online]. Available: https://doi.org/10.1145/3282353.3282357 [retrieved: 08/2024].

[65] D.-I. D. Han, M. C. tom Dieck, and T. H. Jung, "User experience model for augmented reality applications in urban heritage tourism," *Journal of Heritage Tourism*, vol. 13, pp. 46–61, 2018.

[66] T. Olsson, "Layers of user expectations of future technologies: An early framework," in *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '14, Toronto, Ontario, Canada: Association for Computing Machinery, 2014, pp. 1957–1962, ISBN: 9781450324748. DOI: 10.1145/2559206.2581225. [Online]. Available: https://doi.org/10.1145/2559206.2581225.

[67] V. Jääskö and T. Mattelmäki, "Observing and probing," in *Proceedings of the 2003 International Conference on Designing Pleasurable Products and Interfaces*, ser. DPPI '03, Pittsburgh, PA, USA: Association for Computing Machinery,

2003, pp. 126–131, ISBN: 1581136528. [Online]. Available: https://doi.org/10.1145/782896.782927 [retrieved: 08/2024].

[68] D. Perritaz, C. Salzmann, and D. Gillet, "Quality of experience for adaptation in augmented reality," *2009 IEEE International Conference on Systems, Man and Cybernetics*, pp. 888–893, 2009.

[69] Y. Arifin, T. G. Sastria, and E. Barlian, "User experience metric for augmented reality application: A review," *Procedia Computer Science*, vol. 135, pp. 648–656, 2018, The 3rd International Conference on Computer Science and Computational Intelligence (ICCSCI 2018) : Empowering Smart Technology in Digital Era for a Better Life, ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2018.08.221.

[70] S. Irshad and D. Rohaya Bt Awang Rambli, "User experience of mobile augmented reality: A review of studies," in *2014 3rd International Conference on User Science and Engineering (i-USEr)*, 2014, pp. 125–130. DOI: 10.1109/IUSER.2014.7002689.

[71] Y. Ghazwani and S. Smith, "Interaction in augmented reality: Challenges to enhance user experience," in *Proceedings of the 2020 4th International Conference on Virtual and Augmented Reality Simulations*, ser. ICVARS '20, Sydney, NSW, Australia: Association for Computing Machinery, 2020, pp. 39–44, ISBN: 9781450376945. DOI: 10.1145/3385378.3385384. [Online]. Available: https://doi.org/10.1145/3385378.3385384 [retrieved: 08/2024].

[72] S. Irshad and D. R. A. Rambli, "Advances in mobile augmented reality from user experience perspective: A review of studies," in *International Visual Informatics Conference*, 2017.

[73] T. Olsson, "Concepts and subjective measures for evaluating user experience of mobile augmented reality services," New York, 2013. DOI: https://doi.org/10.1007/978-1-4614-4205-9_9.

[74] T. Olsson, E. Lagerstam, T. Kärkkäinen, and K. Väänänen, "Expected user experience of mobile augmented reality services: A user study in the context of shopping centres," *Personal and Ubiquitous Computing*, vol. 17, Feb. 2011. DOI: 10.1007/s00779-011-0494-x.

[75] Y. Georgiou and E. A. Kyza, "The development and validation of the ari questionnaire: An instrument for measuring immersion in location-based augmented reality settings," *International Journal of Human-Computer Studies*, vol. 98, pp. 24–37, 2017, ISSN: 1071-5819. DOI: https://doi.org/10.1016/j.ijhcs.2016.09.014.

[76] M. Gao and D. Boehm-Davis, "Development of a customizable interactions questionnaire (CIQ) for evaluating interactions with objects in augmented/virtual reality," *Virtual Reality*, vol. 27, pp. 1–18, Aug. 2022. DOI: 10.1007/s10055-022-00678-8.

[77] Subandi, Joniriadi, A. A. Syahidi, and A. Mohamed, "Mobile augmented reality application with multi-interaction for learning solutions on the topic of computer network devices (effectiveness, interface, and experience design)," *2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE)*, pp. 1–6, 2020.

[78] H.-C. Lin, M.-C. Chen, and C.-K. Chang, "Assessing the effectiveness of learning solid geometry by using an augmented reality-assisted learning system," *Interactive Learning Environments*, vol. 23, pp. 1–12, Dec. 2013. DOI: 10.1080/10494820.2013.817435.

[79] M. Schrepp, J. Thomaschewski, and A. Hinderks, "Construction of a benchmark for the user experience questionnaire (UEQ)," vol. 4, no. 4, 2017. [Online]. Available: https://reunir.unir.net/bitstream/handle/123456789/11754/ijimai20174_4_5_pdf_94297.pdf?sequence=1&isAllowed=y [retrieved: 08/2024].

[80] A.-L. Meiners, M. Schrepp, A. Hinderks, and J. Thomaschewski, "A benchmark for the UEQ+ framework: Construction of a simple tool to quickly interpret UEQ+ kpis," *International Journal of Interactive Multimedia and Artificial Intelligence*, 2023. [Online]. Available: http://dx.doi.org/10.9781/ijimai.2023.05.003 [retrieved: 08/2024].

[81] E. L.-C. Law and P. van Schaik, "Modelling user experience – An agenda for research and practice," *Interacting with Computers*, vol. 22, no. 5, pp. 313–322, Sep. 2010, ISSN: 0953-5438. DOI: 10.1016/j.intcom.2010.04.006. eprint: https://academic.oup.com/iwc/article-pdf/22/5/313/1988154/iwc22-0313.pdf. [Online]. Available: https://doi.org/10.1016/j.intcom.2010.04.006 [retrieved: 08/2024].

[82] S. Graser, L. H. Nielsen, and S. Böhm, "Factors influencing the user experience of mobile augmented reality apps: An analysis of user feedback based on app store user reviews," in *Digital Disruption and Media Transformation: How Technological Innovation Shapes the Future of Communication*, A. Godulla and S. Böhm, Eds. Cham: Springer International Publishing, 2023, pp. 109–129, ISBN: 978-3-031-39940-4. DOI: 10.1007/978-3-031-39940-4_9. [Online]. Available: https://doi.org/10.1007/978-3-031-39940-4_9.

[83] OpenAI, "GPT-4 technical report," *ArXiv*, vol. abs/2303.08774, 2023. [Online]. Available: https://arxiv.org/abs/2303.08774 [retrieved: 08/2024].

[84] S. Graser, M. Schrepp, and S. Böhm, "Identifying semantic similarity for UX items from established questionnaires using ChatGPT-4," *International Journal On Advances in Systems and Measurements*, vol. 17, no. 1/2, 2024. [Online]. Available: https://www.iariajournals.org/systems_and_measurements/tocv17n12.html [retrieved: 08/2024].

[85] S. Graser, A. Snimshchikova, M. Schrepp, and S. Böhm, "Enhancing ux research activities using genai – potential applications and challenges," *CENTRIC 2024, The Seventeenth International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services*, 2024, accepted for publication in 2024-11. [Online]. Available: https://www.thinkmind.org/library/CENTRIC.

[86] G. Halevi, H. Moed, and J. Bar-Ilan, "Suitability of google scholar as a source of scientific information and as a source of data for scientific evaluation—review of the literature," *Journal of Informetrics*, vol. 11, no. 3, pp. 823–834, 2017, ISSN: 1751-1577. DOI: https://doi.org/10.1016/j.joi.2017.06.005.

[87] A. P. Siddaway, A. M. Wood, and L. V. Hedges, "How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses," *Annual Review of Psychology*, vol. 70, no. 1, pp. 747–770, 2019. [Online]. Available: https://doi.org/10.1146/annurev-psych-010418-102803 [retrieved: 08/2024].

[88] S. J. Kerr *et al.*, "Wearable mobile augmented reality: Evaluating outdoor user experience," ser. VRCAI '11, Hong Kong, China: Association for Computing Machinery, 2011, pp. 209–216, ISBN: 9781450310604. [Online]. Available: https://doi.org/10.1145/2087756.2087786.

[89] A. Dhir and M. Al-kahtani, "A case study on user experience (ux) evaluation of mobile augmented reality prototypes," *Journal of Universal Computer Science*, vol. 19, pp. 1175–1196, Jan. 2013.

[90] K. Rehrl, E. Häusler, S. Leitinger, and D. Bell, "Pedestrian navigation with augmented reality, voice and digital map: Final results from an in situ field study assessing performance and user experience," *Journal of Location Based Services*, vol. 8, pp. 75–96, Apr. 2014. DOI: 10.1080/17489725.2014.946975.

[91] S. Irshad and D. R. A. Rambli, "User experience evaluation of mobile AR services," *Proceedings of the 12th International*

Conference on Advances in Mobile Computing and Multimedia, 2014.

[92] X. Li, B. Xu, Y. Teng, Y.-t. Ren, and Z.-m. Hu, "Comparative research of ar and vr technology based on user experience," *International Conference on Management Science and Engineering - Annual Conference Proceedings*, pp. 1820–1827, Oct. 2014. DOI: 10.1109/ICMSE.2014.6930456.

[93] N. Higgett, Y. Chen, and E. Tatham, "A user experience evaluation of the use of augmented and virtual reality in visualising and interpreting roman leicester 210ad (ratae corieltavorum)," *Athens Journal of History*, vol. 2, pp. 7–22, Dec. 2015. DOI: 10.30958/ajhis.2-1-1.

[94] M. Kamilakis, D. Gavalas, and C. Zaroliagis, "Mobile user experience in augmented reality vs. maps interfaces: A case study in public transportation," Jun. 2016, pp. 388–396, ISBN: 978-3-319-40620-6. DOI: 10.1007/978-3-319-40621-3_27.

[95] S. Irshad and D. R. A. Rambli, "Design implications for quality user experience in mobile augmented reality applications," 2016.

[96] K. Seppälä *et al.*, "Examining user experience in an augmented reality adventure game: Case luostarinmäki handicrafts museum," Sep. 2016, pp. 257–276, ISBN: 978-3-319-44804-6. DOI: 10.1007/978-3-319-44805-3_21.

[97] Z. Rashid, R. Pous, and C. Norrie, "An independent shopping experience for wheelchair users through augmented reality and rfid," *Assistive technology: the official journal of RESNA*, vol. 31, May 2017. DOI: 10.1080/10400435.2017.1329240.

[98] A. Dirin and T. H. Laine, "User experience in mobile augmented reality: Emotions, challenges, opportunities and best practices," *Computers*, vol. 7, no. 2, 2018. DOI: 10.3390/computers7020033.

[99] K.-H. Cheng, "Parents' user experiences of augmented reality book reading: Perceptions, expectations, and intentions," *Educational Technology Research and Development*, vol. 67, Jun. 2018. DOI: 10.1007/s11423-018-9611-0.

[100] R. Hammady, M. Ma, and A. Powell, "User experience of markerless augmented reality applications in cultural heritage museums: 'museumeye' as a case study," in Jul. 2018, pp. 349–369, 5th International Conference, AVR 2018, Otranto, Italy, June 24–27, 2018, Proceedings, Part II, ISBN: 978-3-319-95281-9. DOI: 10.1007/978-3-319-95282-6_26.

[101] Y. Sekhavat and J. Parsons, "The effect of tracking technique on the quality of user experience for augmented reality mobile navigation," *Multimedia Tools and Applications*, vol. 77, pp. 1–34, May 2018. DOI: 10.1007/s11042-017-4810-y.

[102] P. Alavesa *et al.*, "Mobile augmented reality client as a ux method for living lab's user involvement tool," Oct. 2018, pp. 135–142, ISBN: 978-1-4503-6589-5. DOI: 10 . 1145 / 3275116.3275147.

[103] C. Jakobsen, J. Larsen, M. Nørlem, and M. Kraus, "Improving user experience for lost heritage sites with a user-centered indirect augmented reality application," in Jan. 2018, pp. 54–63, ISBN: 978-3-319-76907-3. DOI: 10 . 1007 / 978 - 3 - 319 - 76908-0_6.

[104] S. Stumpp, T. Knopf, and D. Michelis, "User experience design with augmented reality (AR)," Nov. 2019. DOI: 10. 34190/ECIE.19.019.

[105] V. Davidavičienė, J. Raudeliuniene, and R. Viršilaitė, "User experience evaluation and creativity stimulation with augmented reality mobile applications," vol. 12, pp. 34–48, Mar. 2019. DOI: 10.3846/cs.2019.3576.

[106] B. Marques, R. Carvalho, P. Dias, and B. Santos, "Pervasive augmented reality for indoor uninterrupted experiences: A user study," Sep. 2019, pp. 141–144, ISBN: 978-1-4503-6869-8. DOI: 10.1145/3341162.3343759.

[107] C. Andri, M. Alkawaz, and S. Waheed, "Examining effectiveness and user experiences in 3d mobile based augmented reality for msu virtual tour," Jun. 2019, pp. 161–167. DOI: 10.1109/I2CACIS.2019.8825054.

[108] F. Satti *et al.*, "Holistic user experience in mobile augmented reality using user experience measurement index," Sep. 2019, pp. 1–6. DOI: 10.1109/NEXTCOMP.2019.8883528.

[109] M. Cauchi and D. Scerri, "Enriching tourist ux via a location based ar treasure hunt game," Sep. 2019, pp. 199–204. DOI: 10.1109/ICCE-Berlin47944.2019.8966141.

[110] S. Irshad, D. Awang Rambli, and S. Sulaiman, "Design and implementation of user experience model for augmented reality systems," in *Proceedings of the 18th International Conference on Advances in Mobile Computing & Multimedia*, ser. MoMM '20, Chiang Mai, Thailand: Association for Computing Machinery, Nov. 2021, pp. 48–57, ISBN: 9781450389242. DOI: 10.1145/3428690.3429169. [Online]. Available: https : / / doi . org / 10 . 1145 / 3428690 . 3429169 [retrieved: 08/2024].

[111] K. Brata and D. Liang, "Comparative study of user experience on mobile pedestrian navigation between digital map interface and location-based augmented reality," *International Journal of Electrical and Computer Engineering*, vol. 10, pp. 2037–2044, Apr. 2020. DOI: 10.11591/ijece.v10i2.pp2037-2044.

[112] L. Wang and M. Lv, "Study on assessing user experience of augmented reality applications," in *Virtual, Augmented and Mixed Reality. Design and Interaction: 12th International Conference, VAMR 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I*. Copenhagen, Denmark: Springer-Verlag, Jul. 2020, pp. 208–222, ISBN: 978-3-030-49694-4. DOI: 10 . 1007 / 978 - 3 - 030 - 49695 - 1_14. [Online]. Available: https://doi.org/10.1007/978-3-030-49695-1_14.

[113] M. C. R. Harrington, "Connecting user experience to learning in an evaluation of an immersive, interactive, multimodal augmented reality virtual diorama in a natural history museum the importance of story," in *2020 6th International Conference of the Immersive Learning Research Network (iLRN)*, 2020, pp. 70–78. DOI: 10.23919/iLRN47897.2020.9155202.

[114] T. A. Mikropoulos *et al.*, "Acceptance and user experience of an augmented reality system for the simulation of sensory overload in children with autism," in *2020 6th International Conference of the Immersive Learning Research Network (iLRN)*, 2020, pp. 86–92. DOI: 10.23919/iLRN47897.2020. 9155113.

[115] M. Münzer, "How can augmented reality improve the user experience of digital products and engagement with cultural heritage outside the museum space?" *IOP Conference Series: Materials Science and Engineering*, vol. 949, p. 012 040, Nov. 2020. DOI: 10.1088/1757-899X/949/1/012040.

[116] V. Davidavičienė, J. Raudeliuniene, and R. Viršilaitė, "Evaluation of user experience in augmented reality mobile applications," *Journal of Business Economics and Management*, vol. 22, pp. 467–481, Feb. 2021. DOI: 10.3846/jbem.2020. 13999.

[117] I. Navarro *et al.*, "Heritage augmented reality applications for enhanced user experience," in *Learning and Collaboration Technologies: Games and Virtual Environments for Learning*, P. Zaphiris and A. Ioannou, Eds., Cham: Springer International Publishing, 2021, pp. 302–312, ISBN: 978-3-030-77943-6.

[118] Y. Jang and E. Park, "Satisfied or not: User experience of mobile augmented reality in using natural language processing techniques on review comments," *Virtual Reality*, vol. 26, Nov. 2021. DOI: 10.1007/s10055-021-00599-y.

[119] M. Balani and J. Tümler, "Usability and user experience of interactions on VR-PC, hololens 2, VR cardboard and AR smartphone in a biomedical application," in Jul. 2021, pp. 275–287, ISBN: 978-3-030-77598-8. DOI: 10.1007/978-3-030-77599-5_20.

[120] I. Verhulst, A. Woods, L. Whittaker, J. Bennett, and P. Dalton, "Do vr and ar versions of an immersive cultural experience engender different user experiences?" *Computers in Human Behavior*, vol. 125, p. 106 951, Jul. 2021. DOI: 10.1016/j.chb.2021.106951.

[121] V. Barros, E. Oliveira, and L. Araújo, "Analysis of improved user experience when using AR in public spaces," in *Intelligent Computing*, K. Arai, Ed. Cham: Springer International Publishing, 2021, pp. 1169–1189, ISBN: 978-3-030-80126-7.

[122] A. Mohammed, B. Al-Jumaah, and I. Abu Doush, "Enhancing user experience of interior design mobile augmented reality applications," Jan. 2021, pp. 101–108. DOI: 10.5220/0010630400003060.

[123] G. Ku, I.-W. Shang, and M.-F. Li, "How do location-based augmented reality games improve physical and mental health? evaluating the meanings and values of pokémon go users' experiences through the means-end chain theory," *Healthcare*, vol. 9, p. 794, Jun. 2021. DOI: 10.3390/healthcare9070794.

[124] C. Pamparău and R.-D. Vatavu, "The user experience of journeys in the realm of augmented reality television," in *Proceedings of the 2022 ACM International Conference on Interactive Media Experiences*, ser. IMX '22, Aveiro, JB, Portugal: Association for Computing Machinery, 2022, pp. 161–174, ISBN: 9781450392129. DOI: 10.1145/3505284.3529969. [Online]. Available: https://doi.org/10.1145/3505284.3529969 [retrieved: 08/2024].

[125] R. Aparicio, J. Aliaga, and D. Velasco, "Mobile application for the recommendation of furniture and appliances through augmented reality to improve the user experience in the online shopping process," Sep. 2022, pp. 1–6. DOI: 10.1145/3545897.3545898.

[126] M. Alvarez, A. Toet, and S. Dijkstra-Soudarissanane, "Virtual visits: UX evaluation of a photorealistic ar-based video communication tool," in *Proceedings of the 1st Workshop on Interactive EXtended Reality*, ser. IXR '22, Lisboa, Portugal: Association for Computing Machinery, 2022, pp. 69–75, ISBN: 9781450395014. DOI: 10.1145/3552483.3556462. [Online]. Available: https://doi.org/10.1145/3552483.3556462 [retrieved: 08/2024].

[127] I. G. I. Sudipa, P. Aditama, and C. Yanti, "Evaluation of lontar prasi bali application based on augmented reality using user experience questionnaire," *East Asian Journal of Multidisciplinary Research*, vol. 1, pp. 1845–1854, Oct. 2022. DOI: 10.55927/eajmr.v1i9.1531.

[128] H. Karimah, M. Suryani, and A. Hidayat, "Implementation of design thinking in part 30 of the qur'an augmented reality based learning application to improve user experience and cognitive ability," vol. 2433, Oct. 2022, p. 020 019. DOI: 10.1063/5.0100383.

[129] Y. Luo, F. Liu, Y. She, and B. Yang, "A context-aware mobile augmented reality pet interaction model to enhance user experience," *Computer Animation and Virtual Worlds*, vol. 34, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:251830348.

[130] L. Xue, C. Parker, and C. Hart, "How augmented reality can enhance fashion retail: A ux design perspective," *International Journal of Retail Distribution Management*, vol. 51, Sep. 2022. DOI: 10.1108/IJRDM-09-2021-0435.

[131] L. De Paolis, C. Gatto, L. Corchia, and V. De Luca, "Usability, user experience and mental workload in a mobile augmented reality application for digital storytelling in cultural heritage," *Virtual Reality*, vol. 27, pp. 1–27, Nov. 2022. DOI: 10.1007/s10055-022-00712-9.

[132] J. Wolf, D. Luchmann, Q. Lohmeyer, M. Farshad, P. Fürnstahl, and M. Meboldt, "How different augmented reality visualizations for drilling affect trajectory deviation, visual attention, and user experience," *International Journal of Computer Assisted Radiology and Surgery*, vol. 18, pp. 1–9, Feb. 2023. DOI: 10.1007/s11548-022-02819-5.

[133] S. Hu, L. J. Rong, J. Han, D. Zhang, and W. Jiang, "The effects of interaction mode and individual differences on usability and user experience of mobile augmented reality navigation," *IEEE Access*, vol. 11, pp. 41 783–41 795, 2023. [retrieved: 08/2024].

[134] S. Refae, T. Ragab, and H. Samir, "Augmented reality (ar) for urban cultural heritage interpretation: A user experience evaluation," in Mar. 2023, pp. 283–298, ISBN: 978-3-031-19559-4. DOI: 10.1007/978-3-031-19560-0_23.

[135] K. Dağ, S. Çavuşoğlu, and Y. Durmaz, "The effect of immersive experience, user engagement and perceived authenticity on place satisfaction in the context of augmented reality," *Library Hi Tech*, vol. 42, no. 4, pp. 1331–1346, Feb. 2023. DOI: 10.1108/LHT-10-2022-0498.

[136] Q. Gan, Z. Liu, T. Liu, Y. Zhao, and Y. Chai, "Design and user experience analysis of AR intelligent virtual agents on smartphones," *Cognitive Systems Research*, vol. 78, pp. 33–47, 2023, ISSN: 1389-0417. DOI: https://doi.org/10.1016/j.cogsys.2022.11.007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389041722000699 [retrieved: 08/2024].

Designing For and On Wall-Sized Displays: a Preliminary Study with Figma

Lou Schwartz
Luxembourg Institute of Science and Technology (LIST)
Esch-sur-Alzette, Luxembourg
email: lou.schwartz@list.lu
orcid: 0000-0002-8645-5326

Valérie Maquil
LIST
Esch-sur-Alzette, Luxembourg
email: valerie.maquil@list.lu
orcid: 0000-0002-0198-3729

Mohammad Ghoniem
LIST
Esch-sur-Alzette, Luxembourg
email: mohammad.ghoniem@list.lu
orcid: 0000-0001-6745-3651

***Abstract*—Wall-Sized Displays (WSDs) have several spatial characteristics that are difficult to address during user interface design. The design at scale 1:1 could be part of the solution. One designer explored the feasibility to use a well-known prototyping tool, Figma, on two different WSDs, by relying on three different interaction methods: touch, a keyboard and touchpad, and a tablet. We observed that designing at scale 1:1 was appreciated and that interaction with the tablet proved to be the most comfortable interaction method, but Figma seems not adapted for this usage.The physical environment also had an impact.**

Keywords-wall-sized display ; design ; figma ; large scale display.

I. INTRODUCTION

Wall-Sized Displays (WSDs) are also referred to as vertical Large Interactive Displays (LIDs) or Large High-Resolution Displays (LHRDs). However, the notion of 'large' is not defined and can be subjective [1]. Belkacem et al. defined a LHRDs as a display that "*creates a coherent physical view space that is at least of the size of the human body and exhibits a significantly higher resolution than a conventional display*" [2]. According to Chen et al., WSDs improve user performance and satisfaction for tasks, such as model design, analysis, and visual data mining [3]. However, these new ways of viewing, collaborating and interacting differ from desktop and smartphone applications [1], because of their size, their resolution, the collaboration they foster, and the so-called *natural* interactions used (mainly tactile and gestural) [4]. Several papers have voiced the challenges posed by WSDs [1] [2] [5]. As a result, it remains a challenge to support designers with the right tools and methods for designing applications for WSDs. The design of these kinds of systems raises several questions regarding the User Interface (UI).

In this paper, we seek to address the *designer support* challenge, i.e., the need for design and testing tools and methods [1] [5]. Therefore, we look into the design of a UI prototype and seek to understand whether an existing online design tool can be used to prototype UIs in the WSD environment at 1:1 scale. With 'UI prototyping', we mean the prototyping of interface, functionalities, screen layouts and behaviors at the mid-fidelity level. We seek to understand whether a popular UI design tool, Figma [6] can be used to prototype UIs in the WSD environment at 1:1 scale.

Related work about methods and tools to prototype for WSDs are reported in SectionII, then the study is described in Section III, the results are presented in Section IV and discussed in Section V.

II. RELATED WORK

Several tools and methods have been proposed in the literature to support design for WSDs. The interest of **Paper prototyping** [7] is to explore, communicate and evaluate early interface designs with end-users or within the design team. In this case, a designer often plays the role of the computer to simulate the behavior of the system by changing the pieces of paper shown to the participants. A number of studies have used paper prototyping to design applications on a WSD (e.g., [8] [9]). Another option is the use of **role play** and actors to prototype and test interactions with the WSD. For instance, Avellino et al. asked actors to act out some interactions and played them back during tests with participants to simulate a controlled remote collaboration situation on a WSD [10]. Furthermore, **prototype development** is a common practice [11], but there is no indication on how the applications were designed (e.g., [12] [13]). Finally, **mixing mock-up techniques** exist, which are mainly used to prototype ubiquitous computing systems, but can be also used to mock-up WSDs. For instance, *Mini-studio*, which consists of a physical paper model of the system and projection of the content, could be used to prototype for WSDs [14]. Another system is *SketchStudio*, a 2.5D (devices in 2D and characters in 3D) animated scenario design tool for rapid prototyping of systems involving multiple users and multiple components or devices [15]. These methods and tools advantage is that they also enable the interactions to be played around the WSDs. However, they are not accurate enough for a prototype of the screen layout and content, especially in contexts where large amounts of data and high resolutions are required [2]. Overall, the prototyping method is frequently used for designing WSD applications, but how the design was achieved is usually not described. In those cases it is documented, paper prototyping is the most widely used method. We did not find any studies covering the design of a UI prototype on a WSD at 1:1 scale.

III. STUDY

Prototyping for WSDs in actual size, directly onto the targeted support, could reduce complexity, give a sense of scale, and ensure that the target resolution is correctly achieved and exploited. It could also help to check the visibility of the designed UI at various distances and viewing angles [2] [5]. But, as noted by Lischke et al., "*[it] is often not possible to prototype in the original size*" [1].

A. Research question

To study the design at scale 1:1 on a WSD, we decided to use Figma, a mid-fidelity web-based prototyping tool. The research question addressed by this preliminary study is: **can Figma**, as a desktop optimized tool, **be used in a WSD environment to design at 1:1 scale?** What are the problems and the opportunities arising from using Figma to design a UI for, and directly on a WSD ?

B. Protocol

Figma was tested on a WSD by one designer under several experimental conditions: two WSD settings (WSD-IA, and WSD-VW) with different physical conditions, and three different interaction methods: a Bluetooth keyboard with a touchpad, direct touch on the WSD, and a synchronized tablet. The **participant** was an expert in UI design and has participated in the design of several UIs for WSDs, but had never used Figma before. She was free to stop the experiment whenever she wanted (e.g., when it became too difficult) or when having finished the design. Since this preliminary study's aim was to verify the feasibility of using Figma under these conditions before carrying out more in-depth studies, we judged that only one user was necessary.

The **system** consisted of a touch WSD displaying Figma in the Chrome internet browser in full-screen mode. Two WSDs were used: WSD-IA (curved, diameter: 3.64m, height: 2m, composed of 12 4K screens in portrait mode, 8 of which are touch-enabled using infrared frames, completed by a height-adjustable table and a keyboard/touchpad as shown on Figure 2.c) and WSD-VW (flat, width: 7m, height: 2m, total resolution 13152×3872 pixels, composed of 24 HD screens with infrared frames enabling touch, completed by three fixed-height tables with two mobile extended-height chairs at each end of the WSD, a large fixed table opposite the middle of the WSD, touch keyboard displayed in the middle at the bottom of the WSD as shown on Figure 4.c).

Concerning the **task**, the participant first discovered Figma on a computer for two hours. Then, she used Figma to reproduce a UI previously developed for both WSDs as shown on Figure 1. Figma was chosen because of its practicality and popularity. This UI was chosen because it comprises different UI elements (text, sliders, graph, a social media feed). Reproducing an existing UI ensures that it is feasible, well adapted to the WSD environment, and allows observation to be focused on the Figma manipulation rather than the process of creating a new design.

Video cameras and microphones recorded the tests. For the WSD-IA, three video cameras were used, at the top, front (middle of the WSD) and back (at top of the opening). For the WSD-VW, two cameras were used at the back, positioned at the ends of the WSD. Comments and actions **were analyzed thematically** to identify encountered issues.

IV. Results

In general, the participant appreciated the ability to design at a 1:1 scale, regardless of the interaction method and the WSD used, with the main advantage of being able to see the final rendering on the destination screen in real time. Several difficulties can be ascribed to the participant's lack of familiarity with Figma, as the use of widgets, components and plugins was deemed complicated, and the participant was unable to use them successfully.

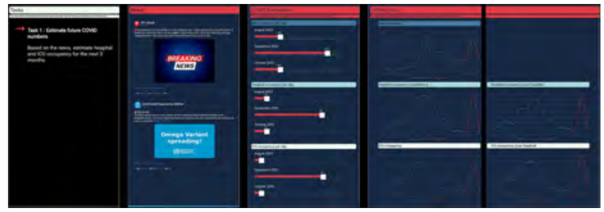

Figure 1. Initial prototype to copy, for more information see [16].

In addition, the configuration of the **Figma environment** was not always adapted for WSDs. For instance, the properties of a selected object are displayed on the right-hand side of the display (see Figure 3.a), the main menu is displayed at the very top (see Figure 3.b) and dialogue boxes open in the middle of the display. The user must also scroll with the cursor or walk through the entire display to modify, e.g., the elements properties (see Figure 2.d), which is tiring over a long period of time. In the following, we will discuss in more detail the issues related to each interaction method.

A. Interacting with a Bluetooth keyboard with a touchpad

The session lasted one hour for the WSD-IA and ten minutes for the WSD-VW. On both WSDs, the participant would sometimes look for the cursor, which was not easy to find on the large display.

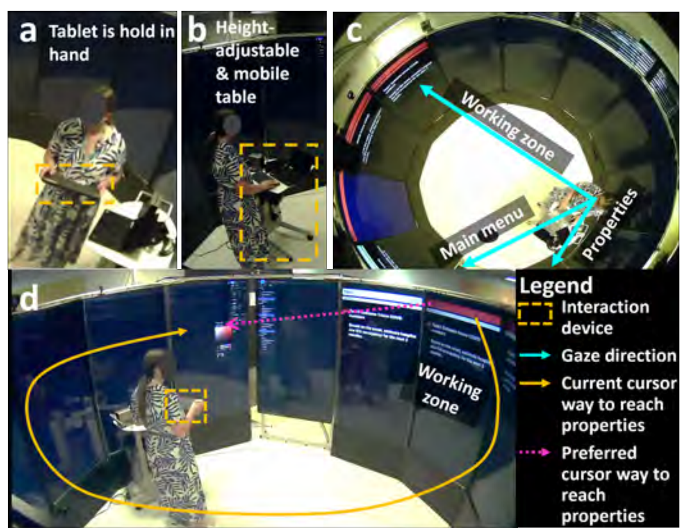

Figure 2. Observations made when interacting with the keyboard+touchpad. a) At the beginning, the participant held the keyboard. b) Use of a table to put down the keyboard. c) Lots of head rotations to see all the important areas. d) WSD-IA does not allow you to cross directly from the left screen to the right screen.

Concerning the *WSD-IA*, the menu and items list were displayed on the screen used as the left door, and the selected item's properties on the right door screen. To avoid turning her head from the extreme left to the extreme right too often, the participant closed the doors to look at them both at the same time. She also worked at an angle to see the menu, properties and working area at a glance (see Figure 2.c). As the session was short, and all UI elements were tightly grouped on the left, the position was acceptable. But the position could not be maintained when the user was working in the middle. In this configuration (menu on the left and properties on the right), the participant turned her head and body a lot, which could possibly be painful and exhausting. At first, the participant carried the interaction device, see Figure 2.a. After 15 minutes, she felt physically tired and placed it on a table, see Figure 2.b. Another problem was the impossibility to switch easily from the WSD's leftmost to the rightmost side with the cursor: the participant must move the cursor all the way around the WSD, which is tiring, see Figure 2.d. To avoid turning her head too much, the participant did not follow the cursor with her eyes when it was behind her back.

In the condition of *WSD-VW*, the Figma interface text size was an issue. Due to the size of WSD-VW and flatness, the text could not be read on the opposite side of the display. So, when the user wanted to modify a property's value, she must move to the properties area. The participant walked a lot across the WSD-VW and rested on a table next to it. Then the user leaned on the middle table for comfort and stayed at a certain distance from the WSD-VW to see everything at once. The fatigue caused by moving around, eye strain due to the text size, and carrying the keyboard led the participant to stop the test after ten minutes. Although the WSD-VW and the WSD-IA have nearly the same length, moving the cursor felt less painful here because it was always visible and the experiment duration was shorter.

B. Interacting using direct touch on the WSD

The session with WSD-IA lasted twenty minutes and the session with WSD-VW was stopped after ten minutes.

In the condition *WSD-IA*, to manage physical fatigue (neck strain and gorilla arm), the user tried to work at a lower scale by zooming in on the work area without minimizing the Figma window. Even if the menu remained too high and properties too far away, objects could be moved with smaller movements and were better positioned in the vision field, generating less neck pain. Only the middle eight screens of the WSD-IA support touch, so they were used to display Figma, but its interface elements (list of objects and properties) took up space. So, the designed prototype, which should start on the first touchable screen, was moved to the right, and the UI elements where no more aligned with the tiles, see Figure 3.a. With touch, it was hard to move an object from one tile to another. The Bluetooth keyboard was used to input text or values. It was held in the hand or placed on the table.

On *WSD-VW*, the properties panel was too far away from the work area, but unlike WSD-IA, when a property was changed on WSD-VW, the result was not visible from the user's position. So, she stepped back to check, e.g., whether the

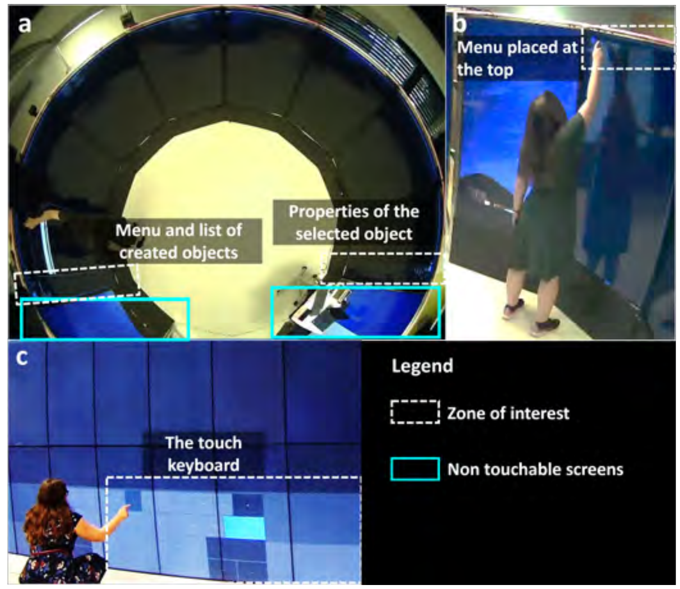

Figure 3. Observations done when interacting with direct touch. a) The touch space is occupied by the Figma interface on the left (list of created objects) and the right (properties). b) The menu is too high. c) The WSD-VW's touch keyboard is not comfortable.

font size is large enough. The top menu was out of reach, and the WSD-VW's virtual keyboard was not suitable for entering more than one word due to its design (position at the bottom and large size, see Figure 3.c). After ten minutes of use, the participant complained from the gorilla arm.

C. Interacting on a synchronized tablet

The same Figma project was loaded onto the tablet and onto the WSD. The UI elements were created, moved and adjusted on the tablet. We observed that the participant mainly looked at the tablet to add UI elements, move them around and set parameters, see Figure 4.a. Then, the participant looked at the WSD to check, e.g., the position and size of the UI elements, the readability of text, and colors, see Figure 4.b. A main issue was the impossibility to select several UI elements at the same time on the tablet, as they are superimposed. The session with the WSD-IA lasted ninety minutes, whereas the session with the WSD-VW was stopped after twenty minutes.

On *WSD-IA*, the user had difficulties to position the UI prototype on the WSD correctly, as the position on the WSD was not synchronized with the tablet, although the modification of UI elements was kept in sync. This required the use of the extra Bluetooth keyboard's touchpad. The participant also used the touchpad to select a group of UI elements to save them as a new reusable UI element. She placed the tablet on the height-adjustable and mobile table. She felt that the WSD-IA and tablet configuration was the most comfortable.

On *WSD-VW*, the participant sat down and placed the tablet on the table, see Figure 4.c. But, as the table was not well positioned and too heavy to be moved, she preferred to hold the tablet in her hand, which was tiresome.

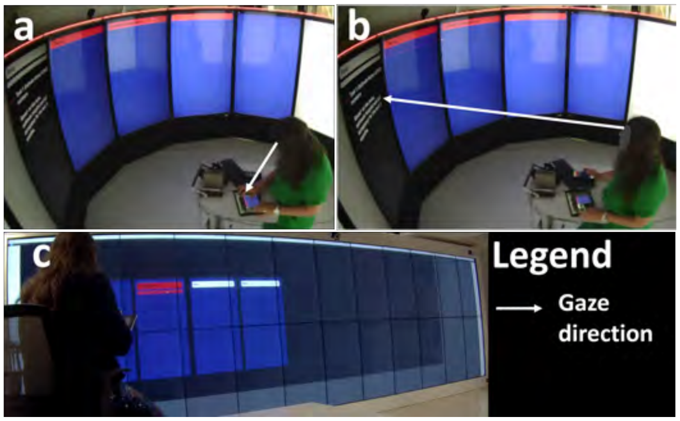

Figure 4. Observations done when interacting with a synchronized tablet. a) The participant modified the prototype on the tablet. b) Then, the participant checked the result on the WSD. c) The participant sat on a chair.

V. Discussion

The duration of the test sessions varied widely, from ten to ninety minutes. The most comfortable condition seems to be the WSD-IA with a tablet and a height-adjustable and mobile table. But the problem of multiple selection and correct positioning of the prototype on the WSD needs to be solved. Overall, the main issues were: (I1) physical fatigue, (I2) accessibility of Figma elements, (I3) readability of the Figma interface, (I4) the hugeness of the interaction surface, (I5) when a project is reopened, objects are moved to the middle, (I6) that a part of the WSD is covered by the Figma elements, which is not a perfect 1:1 scale, and (I7) that dialogue boxes open in the middle of the display.

(I2), (I3), (I5), (I6) and (I7) show that Figma seems not adapted to prototype at 1:1 scale on WSD. We propose some design ideas for each issue. (I1) could be reduced by managing the physical environment and providing a height-adjustable and movable table to place the interacting devices or by interacting at a distance. (I2) could be improved by offering floating contextualized menus and value input fields, by opening dialogue boxes close to the work area or by using a smaller interaction device as a tablet or a laptop. For (I3) and (I4) the size of the Figma elements should be adapted. For (I4) a bigger cursor should be used as well as accelerated scrolling to reach the opposite end of the WSD. (I5) could be solved by fixing the UI elements in their positions and reloading them in exactly the same position. To achieve 1:1 scale (I6) the Figma interface should be concealable or movable. (I7) open the dialogue boxes near of the working zone.

Our findings come with the following limits. The task was limited to the reproduction of an existing UI prototype for a WSD environment. The study involved a single participant who had never used Figma. The advantage was that the user had no prior habits, e.g., using specific shortcuts and was not frustrated by not being able to work as quickly as an expert would on a familiar software.

VI. Conclusion

We presented a preliminary study on the use of Figma to design at scale 1:1 on two different WSDs using three different interaction methods: touch, a keyboard and touchpad, and a tablet. The main study results are that (i) prototyping at 1:1 scale and being able to see the final rendering in real time is appreciated, (ii) interaction with a tablet seems to be the most comfortable, (iii) the design of the physical environment is of utmost importance, and (iv) Figma seems not adapted to this usage in its current form.

In the future, we will expand our sample with a random order of our conditions, to verify our initial observations, and we will explore other prototyping tools.

References

[1] L. Lischke *et al.*, “Challenges in designing interfaces for large displays: The practitioners’ point of view,” in *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. ACM, 2020, pp. 1–6.

[2] I. Belkacem *et al.*, “Interactive visualization on large high-resolution displays: A survey,” in *Computer Graphics Forum*. Wiley Online Library, 2022, p. e15001.

[3] L. Chen, H.-N. Liang, J. Wang, Y. Qu, and Y. Yue, “On the use of large interactive displays to support collaborative engagement and visual exploratory tasks,” *Sensors*, vol. 21, no. 24, p. 8403, 2021.

[4] C. Ardito, P. Buono, M. F. Costabile, and G. Desolda, “Interaction with large displays: A survey,” *ACM Computing Surveys (CSUR)*, vol. 47, no. 3, pp. 1–38, 2015.

[5] L. Schwartz, V. Maquil, and M. Ghoniem, “The challenges of designing for large interactive displays,” in *IHM23 Adjunct*. ACM, 2023, pp. 1–6.

[6] https://www.figma.com/design/, [Online; accessed 17-September-2024].

[7] C. Snyder, *Paper prototyping: The fast and easy way to design and refine user interfaces*. Morgan Kaufmann, 2003.

[8] B. Bailey, J. Biehl, D. Cook, and H. Metcalf, “Adapting paper prototyping for designing user interfaces for multiple display environments,” *Personal and Ubiquitous Computing*, vol. 12, pp. 269–277, 2008.

[9] I. Avellino, C. Fleury, W. E. Mackay, and M. Beaudouin-Lafon, “Camray: Camera arrays support remote collaboration on wall-sized displays,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017, pp. 6718–6729.

[10] I. Avellino, C. Fleury, and M. Beaudouin-Lafon, “Accuracy of deictic gestures to support telepresence on wall-sized displays,” in *Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 2393–2396.

[11] M. Chegini, S. Lin, D. J. Lehmann, K. Andrews, and T. Schreck, “Interaction concepts for collaborative visual analysis of scatterplots on large vertically-mounted high-resolution multi-touch displays,” in *Forum Media Technology*, 2017, pp. 90–96.

[12] A. Prouzeau, A. Bezerianos, and O. Chapuis, “Towards road traffic management with forecasting on wall displays,” in *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*. ACM, 2016, pp. 119–128.

[13] J. Wall and M. Bertoni, “The decision arena: A model-centric interactive workspace for product-service system design,” in *Proceedings of NordDesign 2020*. Design Society, 2020, pp. 1–10.

[14] H.-J. Kim, J.-W. Kim, and T.-J. Nam, “Ministudio: Designers’ tool for prototyping ubicomp space with interactive miniature,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016, pp. 213–224.

[15] H.-J. Kim, C. M. Kim, and T.-J. Nam, “Sketchstudio: Experience prototyping with 2.5-dimensional animated design scenarios,” in *Proceedings of the 2018 Designing Interactive Systems Conference*. ACM, 2018, pp. 831–843.

[16] V. Maquil, D. Anastasiou, H. Afkari, A. Coppens, J. Hermen, and L. Schwartz, “Establishing awareness through pointing gestures during collaborative decision-making in a wall-display environment,” in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, 2023, pp. 1–7.

The Balance Between Surveillance and Privacy

Adjusting to a Changing Threat Landscape

Lasse Berntzen
School of Business
University of South-Eastern Norway
Horten, Norway
e-mail: lasse.berntzen@usn.no

Craig Marais
Cybersecurity Research Group
Gokstad University College
Sandefjord, Norway
e-mail: cm@gokstadakademiet.no

***Abstract*—The cybersecurity threat landscape is growing dramatically, and digital surveillance and its consequence of losing privacy are among the top 10 threats. However, there are contradictory discourses regarding the purpose of surveillance, whether it is for safety or for breaching the privacy of individuals and threatening the security of society. In this paper, we unpack the concept of surveillance and its various forms and purposes. We further look at the factors that advanced surveillance practice, such as technology and the electronic footprint, in addition to the growing group of threat actors. We discuss the matter of balancing surveillance and privacy and draw insights into key measures to deal with surveillance practices by various parties to breach individuals' privacy and the security of society.**

Keywords-surveillance; privacy; human rights; electronic footprints; threat landscape; measures.

I. INTRODUCTION

This paper is a follow-up to an earlier paper coauthored by one of the authors. In that paper, Berntzen and Karamagioli focused on human rights in the context of the digital society [1]. As they observed, privacy is a fundamental human right recognized in all major international agreements regarding human rights, such as Article 12 of the Universal Declaration of Human Rights [2]. The authors emphasized the growing importance of privacy in the context of the digital society. They pointed out that citizens are possible subjects of new and powerful systems of surveillance, personal data collection, and other sophisticated Internet-based techniques, such as the use of "tracking cookies," leaving users completely unaware of such privacy breaches taking place. They also observed a change in government policies where the current political situation in the world and the threat of terrorist attacks have led to governmental proposals in the European Union requiring Internet service providers to store personal information, such as data relating to Internet traffic, e-mails, the geographical positioning of cellular phones and similar, for more extended periods than currently required [3]." They concluded that "ICT offers the technical possibilities of embedded privacy protection obtained by making technology trustworthy and legitimate by design. This includes incorporating options for socially acceptable behavior in technical systems and making privacy protection rights and responsibilities transparent to the user. Therefore, privacy should be a major concern when designing future regulatory mechanisms addressing the digital society."

The paper by Berntzen and Karamagioli [1] was written in 2008. Since then, society has changed. First, the number of electronic footprints has grown exponentially. Second, the threat landscape has changed dramatically; in a recent Delphi study report by the European Union Agency for Cybersecurity (ENISA) [4] on foresight cybersecurity threats for 2030, the threat "Rise of Digital Surveillance Authoritarianism / Loss of Privacy" is ranked number five among the top 10 prioritized threats.

Surveillance has been given several definitions, many of which fall outside the digital scope that we are concerned with, such as defining surveillance as "a systematic social practice" or "watching over and listening to personal details of people." All forms of surveillance have been used for various purposes, such as national security, policing, marketing, epidemiology, and public health [4].

This paper focuses on *digital surveillance using information technology*. This is typically concerned with the collection of personal data; this can be termed "data surveillance" or simply "dataveillance" [5]. A recent example of public health dataveillance is the mobile application *Smittestopp* (Stop the Infection) [6], developed by the Norwegian Institute of Public Health in collaboration with the Simula Research Laboratory. The app was used during the COVID-19 epidemic to track the spread of the virus within Norway and research the effect of the preventive measures applied to deal with the pandemic. However, the Norwegian Data Protection Authority "Datatilsynet" banned the processing of personal data collected by the application, rendering it practically useless.

Furthermore, the widely known Chinese social credit system [7] is an example of dataveillance, implemented as a means of building trust in society through rewards and punishments to fight corruption, telecom scams, tax evasion, academic plagiarism, and pollution, among others.

Privacy has been cited along with surveillance in various discourses, such as seeing surveillance as breaching privacy, using privacy to regulate surveillance, or using surveillance for marketing, which breaches the consumers' privacy but empowers them [5]. According to the International Association of Privacy Professionals (IAPP) [8], privacy is the right to be let alone or freedom from interference or intrusion. In contrast, information privacy is the right to have

 ISBN: 978-1-68558-193-0

control over how your personal information is collected and used.

The question: "Would you prefer privacy or safety?" is relevant for most people. In a society where citizens feel unsafe due to criminal activities, terrorist attacks, and ongoing wars, many citizens welcome surveillance as a safety measure. But surveillance can be abused. One thing is government agencies surveilling public spaces or doing surveillance of criminal suspects after seeking court approval. Another thing is when private companies use the same technologies and tools to profile citizens. This prompts another question: "What is the acceptable use of surveillance?" Many citizens install surveillance equipment in their homes for safety. But that is not the same as giving others access to their homes.

This paper focuses on individuals, but individuals are mostly targets because they are part of an organization. Therefore, the organization plays an essential role in protecting its individuals, and individuals should be considered based on the organizations in which they participate.

The next section discusses the growing electronic footprints, followed by a section reviewing the current threat landscape with new threat actors in the context of surveillance. Section IV discusses how to balance surveillance with privacy. Section V concludes the paper.

II. Growing Electronic Footprints

Surveillance is more than video cameras on street corners or eavesdropping on conversations. It is also about digital footprints caused by advances in technology and new ways to collect and analyze such footprints. The following paragraphs discuss some of these footprints.

A. *Smartphones*

Smartphones have become an integrated part of modern life. The users may perform an increasing number of sensitive and critical tasks, making them a very lucrative target for attackers. Beretas [9] presents an overview of smartphone surveillance methods. Smartphones collect a lot of information, such as geographical position and user behavior. Positioning data is shared with service providers. Smartphones also collect other types of information, like video, photos, and speech, which can be compromised.

B. *Electronic payments*

Cash is less and less used. Electronic payments through smart cards or smartphone payment solutions are taking over. Each transaction is stored with, amongst other data, a timestamp, location, and amount. Lauer [10] discusses surveillance using credit and payment cards, while Martin [11] addresses digital footprints generated by mobile money.

The payment data is valuable for analyzing customer behavior and leaves digital footprints.

C. *Smart Cars*

Smart cars with built-in communication capabilities bring some advantages to their users. The vehicle can report on maintenance status and alert the repair facility about the problem. If the smart car is involved in an accident, the vehicle can alert emergency services automatically. But smart cars also generate comprehensive digital footprints. Claypoole [12] discusses how vehicles will continue to be more intrusive in our lives. Automatic toll stations using plate recognition or RFID technology add to the amount of information generated.

D. *Surveillance cameras*

The number of surveillance cameras has grown exponentially. According to Jha [13], 122.1 million households globally use security cameras. Household cameras can be hacked and may be a severe threat to privacy.

Law enforcement uses video surveillance to monitor public spaces, while companies use video surveillance to protect their properties. Video surveillance is both preventive and valuable for criminal investigations. Ashby [14] analyzed 251,195 crimes recorded by British Transport Police that occurred on the British railway network between 2011 and 2015. CCTV was available to investigators in 45% of cases and judged to be useful in 29% (65% of cases in which it was available).

E. *Internet of Things (IoT) devices*

According to the Norwegian newspaper Aftenposten [15], the number of IoT devices is growing and is estimated to be more than 14 billion devices globally. IoT devices represent a challenge to privacy since they may be used for surveillance of individuals and households to map behavioral patterns [15].

F. *Artificial Intelligence*

The vast amount of information generated by the electronic footprints can be utilized more efficiently due to the implementation of artificial intelligence to analyze numerous data streams at the same time. Feldstein [16] reported on the global expansion of AI surveillance. He mentions new possibilities to analyze digital information in smart cities/safe cities, facial recognition, and smart policing.

III. New Threat Actors and Surveillance

Since George Orwell's novel 1984, surveillance is mainly connected to governments, the "Big Brother" kept track of its citizens. However, today, the threat landscape is more complex. Threat actors are not only the government but also criminals, industrial spies, private companies, individuals, hacktivists, and foreign governments.

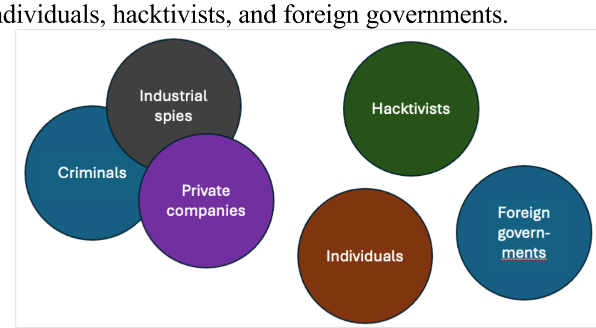

Figure 1. New threat actors.

Copyright (c) IARIA, 2024. ISBN: 978-1-68558-193-0

As shown in Figure 1, criminals, industrial spies, and private companies have overlapping motivations for using surveillance. This overlap signifies a shared motivation of financial gain, to the individual or the company. In contrast, the other actors can be seen to have motivations that differ significantly. We elaborate on these new threat actors and their differences in the following paragraphs.

A. Criminals

Criminals are driven by financial gain. They use surveillance to get information that can be used for blackmailing individuals and companies. Blackmailing may be directly connected to money, but also to gain control of or compromise individuals.

In the case of surveillance by criminals, the target victims are wealthy individuals, especially elderly wealthy individuals. Those wealthy individuals are typically lured through various social engineering techniques to gain access to their personal information and, subsequently, steal their money. Such types of threats result from the lack of awareness among those lured individuals. Consequently, this indicates the need for more regulatory and organizational measures that aim at raising public awareness.

To mitigate risks, such as threats, regulatory measures need to be in place or strengthened through creating awareness campaigns by government agencies, such as the Norwegian Center for Information Security (NorSiS), which is now part of the Norwegian National Security Authority (NSM). Examples of awareness campaigns are the banks' efforts to warn customers about phishing attacks. To support the regulatory and organizational measures, advanced intrusion detection technologies should be in place as technological measures to detect abnormal behavior based on the behavioral patterns of bank customers.

B. Industrial spies

Industrial spies are a subset of criminals with some specific goals. They want to get access to classified information that can be sold to competitors. This can be information on designs, patents, trade secrets, and marketing plans. Hou and Wang [17] observed that techniques generated by rapid developments in IoT and Data Science are enabling a massive increase in both frequency and power of industrial espionage-related activities.

Industrial spy threats are a common surveillance practice whereby an agent aims to steal trade secrets or gain a competitive advantage. This practice has gone through developments from using humans to using technology. The original form of it used to be that spies get hired as employees at the target victim organization. Those spies (under employee cover) can occupy jobs from executive management to janitors at the victim organization. The digital form of industrial spy threat is to break into computers and monitor network traffic for valuable data.

Successful industrial spy threats are a result of needing more robust internal routines for protecting company confidential information.

To mitigate the risks from such threats, organizations should establish measures to protect their infrastructure, but also do relevant background checks on employees trusted with corporate secrets.

Regulatory measures can support the protection of trade secrets through patent registrations and trade secrets laws. Technological measures could help detect break-ins and monitoring attempts.

C. Private companies

Also, private companies may engage in surveillance to get access to privacy-related information. Hinds, Williams, and Johnson [18] addressed privacy concerns and perspectives following the Cambridge Analytica scandal. Cambridge Analytica inappropriately collected data from approximately 87 million users' Facebook profiles to create psychographically tailored advertisements that allegedly aimed to influence people's voting preferences in the 2016 US presidential election [18]. The Norwegian Consumer Council expressed serious concerns about how toy manufacturers are violating privacy by collecting conversations between kids and the toys [19]. They pointed out that the toys fail at several points: lack of security, illegal user terms, kids' secrets being shared, and that kids are subject to hidden marketing.

Private companies are motivated by financial profits and can gain competitive advantages through various forms of surveillance, such as profiling of existing and potential clients, gathering intelligence on competitors, exploitive employee performance measurements, and predatory marketing techniques. Many of these techniques have existed for centuries but have become far more effective in recent years through the use of novel or improved technologies.

Counteracting these actions will require a combination of public engagement and regulatory enforcement. Governing bodies should adopt regulatory requirements for private companies, such as the General Data Protection Regulation (GDPR), in the EU and partner states.

From the consumers' perspective, they should be informed of their rights and how to enforce them when dealing with private companies. Violation of these regulations must impose meaningful penalties on the violators.

Alongside these actions, companies should be encouraged to adopt standards and compliance models that demonstrate their commitment to consumer privacy. One method for companies to establish accountability is to publish transparency reports voluntarily, these reports would help the public and third-party experts to understand how the collected data are being used.

D. Individuals

Individuals may also engage in illegal surveillance activities. Their motivation is to get access to information for personal reasons beyond blackmail and fraud. Examples can be to obtain information about the actions and whereabouts of partners, film nudity or sexual activities, or eavesdrop on conversations in the workplace based on suspicions that coworkers are badmouthing.

The availability of cheap surveillance equipment, such as hidden cameras or audio recorders, lowers the barriers for individuals to indulge in such surveillance. During Arendalsuka (Arendal week) 2024, an annual gathering for politicians, influencers, media, organizations, and other stakeholders, one of the participants is investigated after placing a hidden camera in the bathroom of a flat shared with one politician, two female colleagues, and a journalist. The camera was hidden in a portable audio speaker [20].

This example is not unique, several other incidents have been reported where individuals have placed cameras in toilets of libraries and schools, as well as locker rooms and showers in school facilities.

The legislation clearly forbids secret recordings in public spaces. However, some individuals attempt this kind of illegal behavior driven by their personal motives.

E. Hacktivists

Hacktivists often have ideological or ethical reasons for engaging in surveillance. Their motivation comes from a desire to oppose politicians and decision-makers on specific causes. They aim to reveal hidden truths, mobilize public opinion, or disrupt the operations of their targets. Examples are documenting animal abuse [21], monitoring industrial plants for contamination, or keeping surveillance of child molesters.

Hacktivists engage in surveillance to expose perceived injustices or advance their social or political agendas, often targeting governments or corporations. Many large multinational companies could be considered targets in the eyes of hacktivists; this could include large pharmaceutical companies or oil manufacturers, but hacktivists can also operate at the community level where their actions may have more immediate personal effect; they may target minorities in their community or challenge local government decisions. The common thread for the target of hacktivists is the decision-making power of the individual in the targeted organization or the social impact that ‘the hack’ will produce.

Hacktivist threats happen because of the lack of organizational measures, whether in government organizations or private organizations. Such organizations may need more plans to respond to this type of threat as well as policies to regulate similar threatening activities. In the case of hacktivists who are part of the organization, there is a high risk that they are aware of the vulnerabilities in the organization’s technological infrastructure. Therefore, the hacktivists can exploit those vulnerabilities for their own gains.

To mitigate the risks from such threats, a set of organizational measures should be in place, such as response plans that include alternative actions to handle hacktivists’ threats. Possible actions could be to engage in a dialogue with the hacktivists and try to address their legitimate concerns to resolve the conflict. Another organizational measure could be to have a policy that regulates the conduct of activist activities in a civilized way. Technological measures should be in place as well, such as advanced intrusion detection technologies in addition to cameras to observe the behavior of the hacktivists. To support the organizational and technological measures, some regulatory measures could also be employed to protect the rights of the hacktivists and the affected companies or organizations.

F. Foreign governments

Due to the geopolitical situation, surveillance by foreign governments has become more common. Foreign governments possess highly advanced technology that can be used for surveillance. Norway has seen several monitoring attempts by foreign actors targeting politicians, researchers, and industry leaders. Certain foreign technologies, especially within the telecommunications sector, have been banned due to suspicion of being used for surveillance. In March 2023, the Norwegian Ministry of Justice and Public Security decided to ban TikTok and Telegram from the work mobiles of government employees [22]. The decision was based on a recommendation from the Norwegian National Security Authority (NSM). Within a few days, other public institutions, like the Norwegian Parliament, did the same. The surveillance aims to gather intelligence to be used to blackmail or control individuals and to contribute to the destabilization of the government and political system.

To mitigate the risks of foreign government surveillance and maintain the national security of society, national governments should have in place or strengthen the organizational measures regarding counterintelligence activities to detect and disturb foreign government surveillance. Organizational measures could also include maintaining international collaboration with allied countries to share intelligence information to identify, prevent, or respond to foreign government surveillance. Regulatory measures can further support the mitigation of foreign governments’ surveillance threats through sanctions or other penalizing measures.

IV. How to Balance Surveillance and Privacy?

So far, the discussion has focused on new threat actors in the surveillance area. This section will discuss different categories of measures that can be used to balance the need for surveillance and the need for privacy. Some measures are relevant for individuals, some are relevant for organizations, and some are relevant for society at large. Figure 2 shows the organization of the society. Individuals may or may not be members of an organization, depending on the context. If they are in an organization, the organization may play a role in protecting the privacy of its individuals.

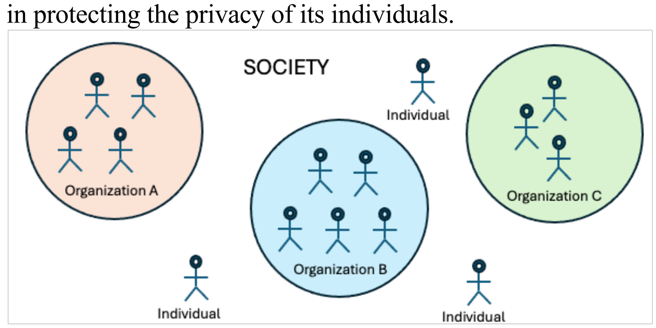

Figure 2. Different levels in the society.

The measures may be regulatory, managerial, or technical. Regulatory measures are about laws and regulations defining the limits on surveillance and the rights to privacy for society. Managerial measures are what an organization does to protect its members. Finally, technical measures are about the use of technology to detect surveillance and protect privacy. Figure 3 shows how the different measures contribute to balancing surveillance needs with privacy protection.

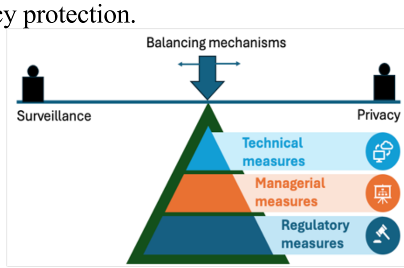

Figure 3. Balancing mechanisms.

The following subsections will describe these three categories of measures, followed by a discussion of data collection practices.

A. *Regulatory measures*

Regulatory measures regulate the use of surveillance and the protection of privacy. On the national level (society), this includes data protection laws and regulations. When deciding on the level of surveillance to be conducted, government actors must weigh up the cost (to privacy) with the proportionality of the threat. Under all circumstances, there must be a clear legal basis for the surveillance. In any circumstance where the cost is considered high, the actors must seek additional judicial oversight (such as a court order). This judicial oversight will help act as a check against potential abuse of surveillance powers. State-conducted surveillance should be done within clearly defined constraints and must stand up to legal scrutiny. New threat actors conducting illegal surveillance activities should be punished based on the severity of the misconduct. The penal code should address unlawful surveillance. On the managerial level, the organizations should comply with the regulations. Individuals should comply with the regulations and use the necessary legal mechanisms available to protect their privacy. Individuals may restrict the use of tracking cookies, not volunteer private information, be careful about using social media, and use the rights of GDPR to remove personal data.

B. *Managerial measures*

Society should protect against the new threat actors, and national authorities and agencies should be established. The police should have the necessary tools and competence to investigate illegal surveillance claims. The personnel conducting surveillance activities should be trained in both the technical aspects and the ethical considerations of their work. Ethical standards should be clearly defined, emphasizing respect for privacy and the importance of adhering to legal and procedural safeguards. On the managerial level, organizations should work on creating policies and guidelines. The most essential part for organizations is to increase their competence. The goal should be to create a security culture where they are aware of the threat landscape, potential actors, and countermeasures. Individuals also play an essential role. They should be aware of possible threats and know how to report suspicious behavior or activities (including illegal surveillance cameras).

C. *Technological measures*

The society should monitor threats and infrastructure on the national level. The national level should also communicate possible threats to organizations and, where relevant, individuals. The organizational level has an important role. Most individuals are targets because of their affiliation to an organization. The organization must secure its infrastructure by establishing relevant access control. The organization should also monitor its infrastructure with intrusion detection and intrusion prevention systems. Incident response handling should be in place, and necessary recovery mechanisms should be established. Individuals can install antivirus software and personal firewalls. They can also use secure communication through Virtual Private Networks (VPN) and avoid unsecured networks.

D. *On Data Collection*

Surveillance technologies should be designed to minimize the collection of data that is not directly relevant to the identified threat or objective of the surveillance. As it can be challenging to predict how data may be used in the future, only the bare minimum of data should be recorded during surveillance activities. Surveillance data should only be retained as long as necessary for legitimate purposes, and access to this data should be restricted to authorized personnel. At all levels, detailed records must be kept of the collection and access of surveillance data. This metadata should be easier to audit and report on and thus not jeopardize the privacy of those under surveillance. The surveillance data should be anonymized and de-identified before storage whenever possible; this helps to protect the privacy of bystanders in the case of public surveillance.

Furthermore, this data should be securely encrypted when in a storage state, further protecting the data in the event of a breach. Breaches can and do happen, and a violation of access to a system should not inherently provide complete access to all data stored therein. Technology should, by default, provide as much privacy as possible and as little access as possible; the control of one's privacy should remain in the hands of the user.

V. CONCLUSION

The previously mentioned new threat actors find the opportunity to use surveillance for their various gains, causing a breach of privacy because of the lack of one or more of the regulatory, managerial, and technological measures. Such a lack of measures causes an imbalance between surveillance and privacy. Surveillance will always happen, but the most important thing is to ensure the protection of privacy.

The balancing mechanisms will take place on different levels of the society. The society consists of individuals that often belong to organizations. Individuals need protection, but when they are part of an organization, the organization also has responsibilities to protect both its own and their privacy. Before engaging in surveillance, the state should do a thorough assessment to ensure that surveillance mechanisms/measures are within policy, proportionate to the threat, necessary, and the potential impact on privacy. This assessment should also account for the storage and accessing of the surveillance data generated. The public must be engaged in matters relating to the surveillance policies and practices. Public oversight bodies play a crucial role in holding their own governments and government agencies accountable for their actions and practices. They should provide guidance and possibly legal assistance to individuals who have been subject to unlawful or wrongful surveillance and thus had their privacy rights violated.

These mechanisms aim to create a balanced approach where the government can effectively protect public safety through necessary surveillance while maintaining strong safeguards to protect individual privacy. The goal is to ensure that surveillance is conducted within a framework that respects human rights and is subject to appropriate checks and balances but also protects citizens from illegal surveillance by new threat actors.

Social values do differ across borders and some populaces may be willing to allow more surveillance if they feel the benefits outweigh the cost to privacy. Our findings are influenced by a Norwegian perspective, characterized by high trust in the government alongside concerns about privacy.

REFERENCES

[1] L. Berntzen and E. Karamagioli, “Human Rights in the Context of the Digital Society,” Proceedings 2nd International Conference on the Digital Society (IARIA). IEEE Computer Society, pp 129-133, 2008.

[2] United Nations, “Universal Declaration of Human Rights (1948),” [Online]. Available from: https://www.un.org/sites/un2.un.org/files/2021/03/udhr.pdf 2024.08.15

[3] Commission of the European Communities, “Proposal for a Directive of the European Parliament and of the Council on the retention of data processed in connection with the provision of public electronic communication services and amending Directive 2002/58/EC,” COM(2005)438 final, 2005 [Online]. Available from: https://www.statewatch.org/media/documents/news/2005/sep/com-438-data-retention.pdf 2024.08.15

[4] European Union Agence for Cybersecurity, “Foresight Cybersecurity Threats For 2030 - Update 2024: Executive Summary,” [Online]. Available from: https://www.enisa.europa.eu/publications/foresight-cybersecurity-threats-for-2030-update-2024-executive-summary 2024.08.15

[5] D. Lyon. “Surveillance,” Internet Policy Review, 11(4) [Online]. Available from: https://policyreview.info/concepts/surveillance 2024.08.15

[6] J. Lund-Tønnesen, “Smittestopp (Stop the infection),” Store Norske Leksikon [Online]. Available from: https://snl.no/Smittestopp 2024.08.15

[7] Z. Yang, “China just announced a new social credit law. Here’s what it means,” MIT Technology Review. [Online]. Available from: https://www.technologyreview.com/2022/11/22/1063605/china-announced-a-new-social-credit-law-what-does-it-mean/ 2024.08.15

[8] International Association of Privacy Professionals (IAPP), “What does privacy mean?” [Online]. Available from: https://iapp.org/about/what-is-privacy/#:~:text=Broadly%20speaking%2C%20privacy%20is%20the,information%20is%20collected%20and%20used. 2024.08.15

[9] C. P. Beretas, “Smart Phones Surveillance Methods,” Journal of Clinical and Biomedical Recent Advances. 1(1), pp. 1-5, 2022.

[10] J. Lauer, “Plastic surveillance: Payment cards and the history of transactional data, 1888 to present,” *Big Data & Society*, *7*(1). https://doi.org/10.1177/2053951720907632

[11] A. Martin, “Mobile Money Platform Surveillance,” Surveillance & Society, 17(1-2), pp. 213-222, 2019.

[12] T. F. Claypoole, “How Your Car Became a Surveillance Weapon,” National Law Review, XIV(243) [Online]. Available from: https://natlawreview.com/article/how-your-car-became-surveillance-weapon 2024. 2024.08.15

[13] R. Jha, “Surveillance cameras in cities: A threat to privacy?” Observer Research Foundation. [Online]. Available from: https://www.orfonline.org/expert-speak/surveillance-cameras-in-cities-a-threat-to-privacy 2024.08.15

[14] M. P. J. Ashby, “The Value of CCTV Surveillance Cameras as an Investigative Tool, An Empirical Analysis,” European Journal on Criminal Policy and Research, 23, pp. 441-459, 2017.

[15] V. S. Wiken, «Hvor sikre er de nye smartdingsene dine? (How safe are your new smart devices),” Aftenposten. 01.08.2023. pp 4-5.

[16] S. Feldstein, “The Global Expansion of AI Surveillance,” Carnegie Endowment for International Peace (Working Paper). [Online]. Available from: https://blog.fdik.org/2019-09/WP-Feldstein-AISurveillance_final1.pdf 2024.08.15

[17] T. Hou and V. Wang, “Industrial espionage – A systematic literature review (SLR),” Computers & Security, 98, 102019, https://doi.org/10.1016/j.cose.2020.102019, pp. 1-12, 2020.

[18] J. Hinds, E. J. Williams, and A. N. Joinson, “’It wouldn't happen to me’: Privacy concerns and perspectives following the Cambridge Analytica scandal,” International Journal of Human-Computer Studies, 143, 102498, https://doi.org/10.1016/j.ijhcs.2020.102498, 2020.

[19] F. Myrstad, “Connected toys violate European consumer law,“ Norwegian Consumer Council, 2016. [Online]. Available from: https://www.forbrukerradet.no/siste-nytt/connected-toys-violate-consumer-laws/ 2024.08.15

[20] M. Rønning and K. Skårdalsmo, “Næringstopp siktet for snikfilming under Arendalsuka (Business leader charged with surreptitious filming during the Arendal week),” [Online]. Available from: https://www.nrk.no/norge/naeringstopp-siktet-for-snikfilming-under-arendalsuka-1.17022547 2024.08.30

[21] Animal Justice Project (website) [Online]. Available from: https://www.animaljusticeproject.com/undercover-investigations 2024.08.15

[22] A. Staalesen, “Norway calls on state officials to delete Telegram,” The Barents Observer [Online]. Available from: https://thebarentsobserver.com/en/democracy-and-media/2023/03/norway-calls-state-officials-delete-telegram 2024.08.15