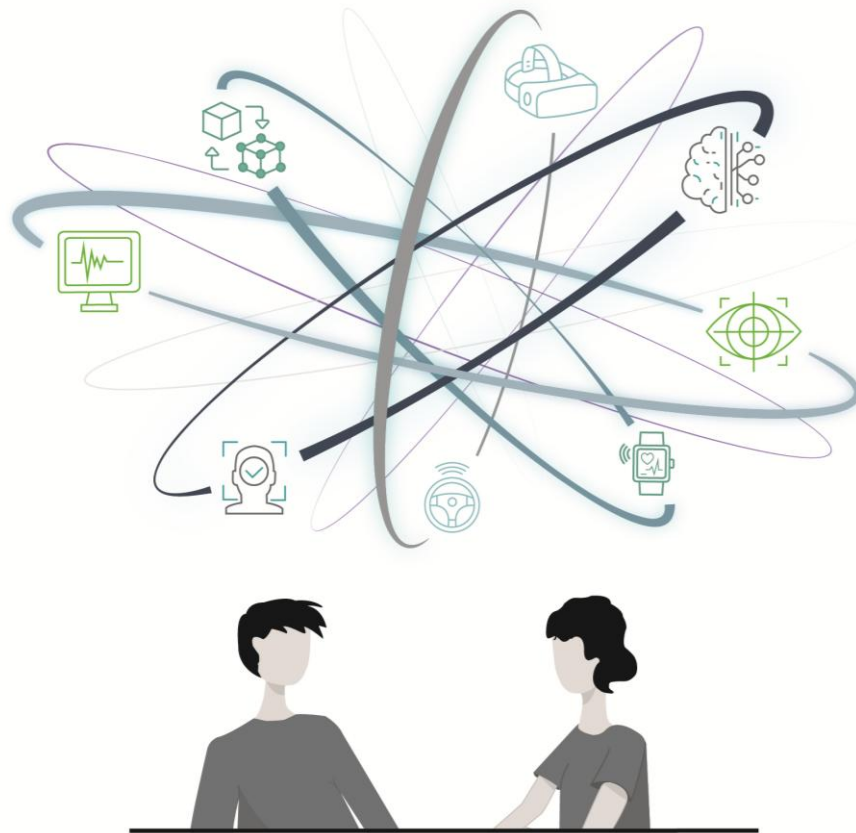




CHALMERS
UNIVERSITY OF TECHNOLOGY



The Future of User Evaluation

Exploring how emerging technology can be used in user evaluation to investigate driver interaction

Master's thesis in Industrial Design Engineering

LOVISA LJUNGDAHL
JONNA STENLUND

DEPARTMENT OF INDUSTRIAL AND MATERIALS SCIENCE

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2025

www.chalmers.se

MASTER'S THESIS 2025

The Future of User Evaluation

Exploring how emerging technology can be used in user
evaluation to investigate driver interaction

LOVISA LJUNGDAHL

JONNA STENLUND



CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Industrial and Materials Science
Division of Design and Human Factors
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025

The future of user evaluations
Exploring how emerging technology can be used in user
evaluation to investigate driver interaction

LOVISA LJUNGDAHL
JONNA STENLUND

© Lovisa Ljungdahl & Jonna Stenlund, 2025

Supervisor: Siw Eriksson
Examiner: Pontus Wallgren

Master's thesis
Department of Industrial and Materials Science
Division of Design and Human Factors
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden

Cover: An illustration of a user evaluation supported by various technologies
Cover made by Lovisa Ljungdahl & Jonna Stenlund

Print: Chalmers Digitaltryck
Gothenburg, Sweden 2025

Abstract

Technology is advancing rapidly across multiple fields and emerging technologies hold significant potential to shape the future of product development and design. This thesis explores the intersection of emerging technology and user evaluation in collaboration with Volvo Group Trucks Technology. It is particularly focused on the context of human-machine interaction and how emerging technologies can enhance future user evaluation of driver interaction. The study set out to identify what types of insights and results emerging technologies could offer for user evaluation, along with associated practical limitations and opportunities in an industrial automotive setting.

The work was structured in two phases. The first phase involved technology scouting through literature research and expert meetings to identify relevant emerging technologies. In the second phase, two selected technologies, extended reality and emotion interpretation, were tested in practice to further explore their value for user evaluation. A total of 19 expert meetings were held, an XR workshop was conducted with both a pilot and main session, and three types of emotion interpretation tests were performed.

Five technologies were identified as particularly relevant for user evaluation in the considered context: extended reality, emotion interpretation, EEG sensors, eye tracking and artificial intelligence. The results suggested that each explored technology demonstrates distinct strengths and opportunities to contribute to in-depth understanding of users and enhanced user evaluation. However, the results of the study also highlight technology limitations and the importance of complementing, rather than replacing, traditional user evaluation methods.

The thesis concluded with specific recommendations for each explored technology area, along with an overarching recommendation to actively monitor ongoing developments in emerging technology for user evaluation. Specific recommendations for user evaluation in the automotive context were also presented, as well as the suggestion that a balanced approach between new tools and established practices would be most effective. While further research in this area is needed, the findings offer promising directions for how industrial actors, like Volvo Group, can advance user evaluation practices through emerging technologies.

Acknowledgements

This master's thesis would not have been achievable without the support we have received throughout the project. We would like to express our sincere appreciation to our academic supervisor Siw Eriksson for always taking the time to assist us, showing genuine engagement, offering valuable advice and guiding us through the complexity of this study. We also extend our thanks to our examiner, Pontus Wallgren, for his helpful direction and contribution to insightful discussion throughout the thesis process.

We are deeply grateful to Volvo Group Trucks Technology for hosting our master's thesis and for the meaningful learnings and experiences we gained during our time at the company. A special thanks goes to our supervisor at Volvo, Matilda Wiklund, for her commitment to continuous supervision, valuable industry knowledge and consistently positive approach. Additionally we would like to thank the entire team at Strategy and Governance Office for their warm welcome and support whenever needed.

We would also like to thank everyone who participated in our meetings, interviews and evaluations. The time and contributed perspectives were essential to the development of this thesis and are deeply appreciated.

Lastly, we would like to extend our gratitude to Chalmers University of Technology and our Industrial Design Engineering program for an incredible five years. This journey has been truly rewarding, filled with valuable experiences and knowledge that we will carry with us forever.

Lovisa Ljungdahl & Jonna Stenlund, Gothenburg June 2025

Abbreviations

Instrument panel = The combined interface consisting of the instrument cluster, dashboard, center console and side displays in a vehicle.

HMI = Human-Machine Interaction refers to the communication and interaction between a human user and a machine or system.

XR = Extended Reality, a broad term that encompasses all forms of technology that blend real and virtual environments for immersive or interactive experiences.

VR = Virtual Reality, a form of XR that immerses users in a completely digital environment.

AR = Augmented Reality, a form of XR that overlays digital content onto the physical world.

MR = Mixed Reality, a form of XR that blends physical and digital environments.

EEG = Electroencephalography, a technique used to measure electrical activity in the brain.

BCI = Brain-Computer Interface, a direct communication pathway between the brain and an external device.

DMS = Driver Monitoring System, a technology that tracks a driver's behavior, attention and state to ensure safe driving.

AI = Artificial intelligence is the field of computer science dedicated to creating systems that can learn from data, reason and perform tasks that normally require human intelligence.

Table of contents

1. Introduction	1
1.1. Background	2
1.2. Aim	2
1.3. Research Questions	3
1.4. Objectives	3
1.5. Report Structure	3
2. Related Theory	4
2.1. Introduction to User Evaluation	6
2.2. Automotive Human-Machine Interaction (HMI)	6
2.3. Introduction to Emerging Technology	6
2.4. Emerging Technologies with Potential for User Evaluation	17
3. Study Methodology and Process	19
4. Phase One - Technology Scouting: Method	21
4.1. Technology Literature Research	21
4.2. Meetings with Experts for Technology Knowledge	23
4.3. Analysis of Technology Scouting	23
5. Phase One - Technology Scouting: Results	25
5.1. Findings from Technology Scouting	32
5.2. Phase One's Implications for Future User Evaluation	32
6. Phase Two - User Evaluation Testing: Method	34
6.1. Aim of Testing User Evaluation for HMI in Practice	34
6.2. Several Considered User Groups	35
6.3. Exploratory Ideation of User Evaluation Use Cases	35
6.4. User Evaluation Testing in Practice - XR Technology	39
6.5. User Evaluation Testing in Practice - Emotion Interpretation Technology	47
7. Phase Two - User Evaluation Testing: Results	48
7.1. Listing of User Evaluation Use Cases	48
7.2. Insights from XR Workshop Testing	54
7.3. Findings from Tests Using Emotion Interpretation Technology	57
8. Synthesized Results	59
9. Discussion	61
9.1. Aim & Research Question	61
9.2. Results	64
9.3. Process & Execution	66
9.5. Recommendations Regarding Future User Evaluations	68
10. Conclusion	69
References	70
Appendices	90

1

Introduction

Technology is advancing at a rapid pace across multiple areas. These forefront technologies are labeled *emerging technologies*. This advancement has the potential to shape the future and leveraging the opportunities they offer is in many ways crucial to ensure future competitiveness for tech-focused industries at a global scale (European Innovation Council Tech Report, 2024).

The rapid progression of technology has significant implications for product development and design and its associated evaluation practices. *User evaluation* is an integral part of such practices which involves assessing how intended users interact with, or experience, a product or system (Bertini et al., 2008). Such interaction between a user and a product or system is referred to as a human-machine interaction (HMI).

This study explores the intersection of emerging technology and user evaluation in the context of human-machine interactions. It is conducted in collaboration with Volvo Group Trucks Technology (Volvo GTT) and its Strategy and Governance Office (SGO).

1.1. Background

The following section covers emerging technology in user evaluation and provides a company description of Volvo Group. It highlights the evolving nature of user evaluation practices and how emerging technology can enhance the practices within Volvo Group's innovation strategy.

1.1.1. Technological Advancements in User Evaluation

User evaluation practices within product development and design industries have gradually evolved alongside technological advancements. New methods and tools have been incorporated to ensure the continued relevance and increased value of user evaluation. Today, a variety of methods are commonly used to conduct user evaluations within product development and design processes, with the choice of method depending on the specific context and objectives. Such methods include traditional interviews and observations, usability testing, surveys, focus groups, ethnographic studies and user journey mapping, among others.

While these methods provide valuable insights, utilizing emerging technologies in user evaluation creates the potential for new perspectives and more comprehensive understanding of users, their experiences, behaviors and perceptions. Proactively exploring possibilities and ensuring continuous evolution of user evaluation is increasingly important given the rapidly developing technology landscape. This can be achieved by identifying newly launched technologies, those that have been in use but have undergone significant improvements, or technologies that can be applied in new contexts to explore how they can be integrated into user evaluation methods.

1.1.2. Company Description: Volvo Group

As a key player in the global automotive industry, Volvo Group delivers transport and infrastructure solutions such as trucks, buses, construction equipment, as well as financing and services. This study will be conducted in collaboration with Volvo GTT and its department SGO. Within Volvo Group, Volvo GTT handles all technology and product development related to truck operations and aftermarket support. In this context, SGO is responsible for Volvo Group's strategy regarding emerging technology. They fulfill this role by identifying emerging technologies and trends with the greatest potential for value creation for Volvo Group. They further manage uncertainty and push boundaries to establish new ways of working and enhance innovation and competitiveness.

Additionally, maintaining its position as an industry-leading innovator requires Volvo Group to operate quickly and efficiently while prioritizing aspects like safety, product usability and customer satisfaction. Given the human-centered nature of the products and services offered by Volvo Group, high-quality human-machine interaction is especially important for achieving these goals, as it directly shapes how users experience and interact with the complex systems.

This underscores the importance for Volvo Group to conduct thorough user evaluations. As of now, this is done using a variety of comprehensive methods and tools. However, with the rapid pace of technological advancement and the strategic focus on future-oriented technologies led by SGO, there is growing interest in exploring how emerging technologies can be leveraged within user evaluation practices at Volvo Group. By doing so, there is potential to discover new ways of gaining deeper user insights and strengthening the company's product development results. Due to the evolving nature of the field of emerging technology, the topic is relevant not only to Volvo Group's innovation efforts but also to current research.

1.2. Aim

The aim of this study is to investigate how emerging technologies can enhance future user evaluation of human-machine interactions in the design- and product development process.

1.3. Research Questions

The research questions this study aims to answer are as follows:

- How can emerging technology be used in future user evaluation?
- What types of insights and results can be obtained from user evaluation that incorporate emerging technologies?
- How can Volvo Group leverage these opportunities?

1.4. Objectives

The objectives this study aims to achieve are as follows:

- Provide Volvo Group with insights on how emerging technologies can enhance user evaluation practices within the context of human-machine interactions.
- Identify emerging technologies with potential applications in user evaluation methods.
- Identify future opportunities for user evaluation methods.
- Evaluate user evaluation methods that incorporate emerging technology and develop recommendations based on the findings.

1.5. Report Structure

This report presents both the process and findings of the study, divided into two phases. It begins by describing the method and results from the first phase, which focused on technology scouting. It is followed by the method and results from the second phase, which focused on user evaluation testing. A more in-depth description of the study's overall methodology and process will be provided in Chapter 3: Study Methodology and Process. Please note that some of the material included, originating from the study process, are in Swedish.

2

Related Theory

This chapter provides a theoretical framework for understanding the context of the study by covering the concepts of user evaluations, automotive HMI and emerging technologies. Additionally, it examines a range of emerging technologies relevant in user evaluations, including electroencephalography sensors (EEG sensors), extended reality (XR), artificial intelligence (AI), driver-monitoring systems (DMS), emotion interpretation, digital twins and wearable devices.

2.1. Introduction to User Evaluation

To create products, systems and services that function effectively, it is essential to first understand the prevailing circumstances, the context in which they will be implemented and, most importantly, the needs and preferences of the users (Wikberg Nilsson et al., 2015). This is where user evaluation becomes invaluable, to ensure solutions with seamless, effective and intuitive user interaction. User evaluation can also be referred to as user testing, user-based testing, user research and more. In this study it will hereafter be referred to as user evaluation. As described by Bertini et al. (2008) user evaluation is evaluation conducted together with the people intended to use the system or product being evaluated. There can be several different users or stakeholders in one project, including the client, investors, and various types of end-users (UserBit, 2025). These can be categorized as primary, secondary, tertiary users and so on. The goal of user evaluation can vary, but it is often centered around understanding the user in relation to the product (Moran, 2019).

This may involve identifying design issues, learning about user behavior and uncovering potential opportunities. By conducting evaluations along with users, designers can identify problems, improve products, reduce costs, create a competitive advantage and assure that products meet user needs (Interaction Design Foundation - IxDF, 2016). By gaining the user's perspective and hearing their insights the product's functionality, usability and overall experience can be enhanced (Moran, 2019).

2.1.1. Categories and Methods of User Evaluation

There are various methods for conducting user evaluation and the choice of method depends on various factors including the context, the participants involved, the stage of the development process, the product being evaluated and the specific criteria being assessed. Examples of user evaluation methods include usability tests, interviews, observations, and

focus groups (Wikberg Nilsson et al., 2015), as well as the NASA Task Load Index (NASA-TLX) (Agency for Healthcare Research and Quality, n.d.), Enhanced Cognitive Walkthrough (ECW) (Khasanah & Gunawan, 2023), and Wizard of Oz (Paul & Rosala, 2021). Practices of user evaluation have evolved over time alongside technological advancements. Examples of technologies that have led to the introduction of new methods for conducting user evaluation are eye tracking, driving simulators and virtual reality (VR).

User evaluation can be broadly categorized into two main types: qualitative and quantitative testing (Moran, 2019). Qualitative testing focuses on users' subjective insights and experiences, making it particularly useful for identifying issues related to user experience. This approach relies on a hermeneutical and interpreting approach (Wikberg Nilsson et al., 2015) and includes methods such as interviews, workshops, and observations. The purpose of qualitative evaluations is to develop a deep understanding of specific phenomena, user attitudes, and the reasons behind user behaviors. To enhance the chances of achieving this, researchers can employ the technique of triangulation by using multiple qualitative methods to develop a multifaceted understanding of phenomena (Carter et al., 2014). In contrast, quantitative testing collects measurable data, such as task completion rates and time-on-task data (Moran, 2019). It involves gathering data from a large number of users and is used in the design process to gain insights rather than to prove a point (Wikberg Nilsson et al., 2015). Typical quantitative methods include questionnaires, measurements related to human comfort (such as light, sound, and climate) and human dimensions (such as height, width, and reach). In both design and development processes, a combination of qualitative and quantitative evaluations is essential to achieve the most comprehensive understanding of the users and their context.

2.1.2. User Evaluation Stages in the Product Development Process

User evaluation can take place at different stages of the product development process (Wikberg Nilsson et al., 2015). Gathering information about a user is not a task that ends before the project is completed, but rather an ongoing process. In the early development stages, it is important to explore user needs and gather user insights to shape the solution. According to Ehrlich and Rohn (1994) and Noyes et al. (1996), user involvement is most effective in the early stages of development, as the cost of making changes increases significantly in later phases. As development progresses, iteratively and continuously refining the understanding of the user's situation and interaction with the solution plays a crucial role (Wikberg Nilsson et al., 2015). This repeated process is essential for creating a well-rounded and effective final product. However, the degree of user participation in the development process can vary significantly. In many cases, users are only involved late in the process to evaluate a proposed solution and confirm whether it meets the specified requirements (Bødker & Buur, 2002; Henninger et al., 2005). Some companies design *for* users, while others design *with* or *by* users (Eason, 1995).

2.1.3. Future Technology Trends in User Evaluation

The future of user evaluations is yet to be determined, but technology and user experience have always been intertwined; when one advances, it influences the other, and vice versa (Experience Haus, 2024). The key emerging technology trends predicted for future user evaluations include AI, VR, emotion-sensing technologies, biometrics and eye tracking, as identified by Patel (2024) and Experience Haus (2024). Additionally, there is an increase in interest for performing user evaluations remotely, particularly accelerated by the Covid-19 pandemic (Larsen et al., 2021).

2.2. Automotive Human-Machine Interaction (HMI)

Many products involve HMI, ranging from household appliances and computers to industrial equipment. As a result, the concept of HMI is broad, highlighting the need to clearly define its meaning in the automotive context. According to Gong (2024), automotive HMI can be described as a system that enables the exchange of dynamic information and emotions between the human user and the vehicle, excluding core driving tasks. Its primary function is to facilitate information transfer, such as user commands and vehicle feedback.

Beyond informational exchanges, emotional interaction is emerging as an important aspect of automotive HMI (Gong, 2024). This involves evoking more abstract feelings, such as a sense of luxury or comfort, through features like dynamic ambient lighting or animated visual effects. Once overlooked, these emotional interactions are now receiving greater attention in the design of in-vehicle experiences.

However, it is important to note that automotive HMI is exclusively concerned with dynamic interactions (Gong, 2024), meaning static elements like door sill text or seat stitching do not constitute HMI, as they lack active communication. Furthermore, primary driving activities, such as steering wheel positioning, pedal feel, or gear shifting in manual vehicles, are generally not considered part of HMI.

2.3. Introduction to Emerging Technology

The term emerging technology covers a broad spectrum of innovations and there has not always been a clear consensus on its definition. According to Halaweh (2013), technology is considered emerging if it is expected to radically impact businesses, industries or society. Similarly, Rotolo et al. (2015) define emerging technologies as radically novel, relatively fast-growing, coherent, prominently impactful, uncertain and ambiguous. Examples of technology trends regarded as significant and emerging are AI, machine learning, connectivity, immersive-reality technologies, quantum technologies, future of mobility and future of bioengineering (McKinsey Technology Council, 2024). As such, technology labeled as emerging does not necessarily need to be new, but can be regarded as emerging in a specific context even if well-established elsewhere.

The importance of leveraging the opportunities of emerging technology is highlighted in the European Innovation Council Tech Report (2024) and similarly by Vinnova (2024). It underscores the necessity of acknowledging the potential of emerging technology, both within tech-focused industries and on a global scale. However, when identifying the potential and leveraging the opportunities of emerging technology, it is crucial to assess whether its claims are viable and capable of delivering results to avoid falling for unrealistic “technological hype” (Gartner, n.d.). There are also additional challenges regarding emerging technology, due to its high uncertainty (Brey, 2017). Such challenges can include predicting future use patterns, impacts and consequences - both on an ethical, societal and environmental scale. Thus, the novelty of emerging technology presents both promising opportunities as well as considerable risks.

2.4. Emerging Technologies with Potential for User Evaluation

Emerging technologies can contribute to enhanced user evaluation by potentially providing new and deeper user-insights. When examining the relevance of different emerging technologies in user evaluation, key aspects to consider include technical foundations, current applications, challenges and future opportunities. This chapter presents a theoretical background for the key aspects of the examined technologies, with an overview presented in Figure 1.

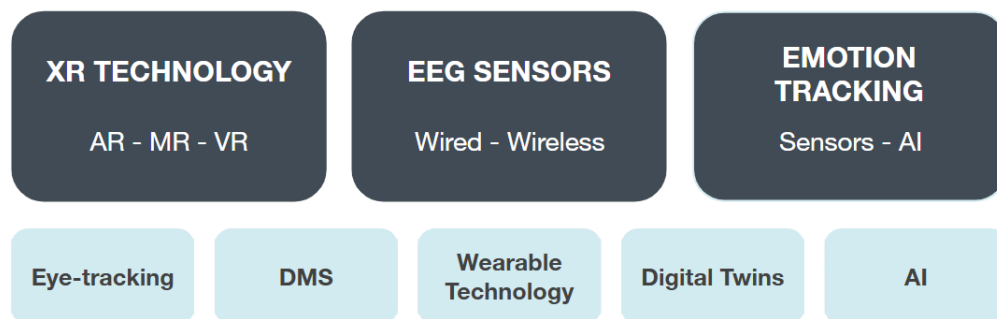


Figure 1. Examined technologies with potential for user evaluation.

2.4.1. Extended Reality (XR) Technology

XR is an umbrella term that encompasses several technologies such as virtual reality (VR), augmented reality (AR) and mixed reality (MR) (Interaction Design Foundation - IxDF, 2022). These technologies enhance reality by blending the physical and digital worlds or creating fully digital environments to deliver immersive experiences. However, XR is not limited to these specific technologies, it includes any existing or future innovations that enable the seamless fusion of physical and digital realities. Since XR covers the entire spectrum from fully real to fully virtual experiences, there are significant differences between VR, MR and AR. Virtual reality immerses users entirely in a digital environment, thereby

disconnecting them from the physical world. Augmented reality enhances the real world by overlaying digital content without replacing the existing environment. Mixed reality combines elements of both AR and VR, allowing real and virtual components to interact seamlessly in real-time. The distinctions between XR technologies are illustrated in Figure 2.

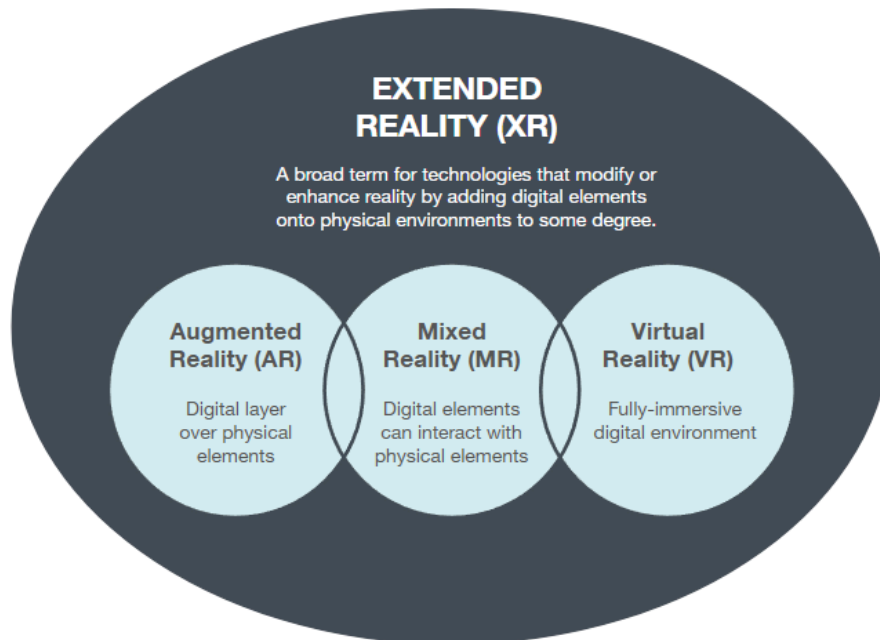


Figure 2. Distinction between XR technologies.

Available XR technology devices

The most common device when wanting an immersive XR experience is a wearable XR headset (Yord Studio, 2024). The different types of XR headsets vary from lightweight AR glasses to larger VR headsets, with each designed to meet specific use case requirements. Some headsets prioritize high resolution and exceptional quality, while others are designed with comfort as the main focus. Modern VR headsets typically fall into two categories: tethered and standalone. Tethered headsets are physically connected to a PC or PlayStation, while standalone headsets provide greater physical freedom by eliminating cables and not requiring an external device for processing.

Currently, some of the most common XR headsets on the market include the Meta Quest 3, Varjo and HTC Vive. Another XR device is The HoloLens, a pair of lightweight AR glasses that can project AR content in the real world. They have a limited field of view and are best used indoors (Skarredghost, 2021). The Meta Quest 3 supports both VR and MR experiences and is one of the most affordable standalone headsets, offering a resolution of 2064 x 2208 pixels per eye and the best overall VR experience for those new to VR and with no access to a tethered connection or PC (Greenwald, 2025). On the other hand, the HTC Vive Pro 2, a tethered headset, targets both enthusiasts and professionals with a sharp VR display and resolution of 2,448 x 2,448 pixels per eye. Varjo offers some of the most immersive virtual and mixed reality products, aimed at advanced VR users who require cutting-edge technology and resolutions of up to 3840 x 3744 pixels per eye (Varjo Technologies, 2025).

Current applications of XR technology

XR technology has its roots as far back as the 1800s, when scientist Sir Charles Wheatstone introduced the concept of stereopsis in 1838, explaining how the brain merges two images to create a single 3D perception (Marr, n.d.). This discovery led to the development of early stereoscopic devices that produced 3D images with a sense of depth. However, modern XR as we know it began to take shape in 1956, when cinematographer Morton Heilig created Sensorama, the first VR machine. Since then, XR technology has evolved rapidly and continues to grow in terms of both importance and application.

Key applications of XR today include product design, interior design, architectural representation and education (Interaction Design Foundation - IxDF, 2022). By enabling the visualization and interaction with 3D models, XR enhances prototyping and user evaluations, ultimately offering more effective design tools. Similarly, in interior design and architectural representation, designers and architects can create spatial layouts and immersive walk-throughs of rooms and buildings. This allows for better-informed decisions regarding layout, materials, and structure, which can reduce errors before construction or production. In education, XR provides immersive and engaging learning experiences, making complex concepts more tangible and interactive. However, the impact of XR goes far beyond these applications, extending into a wide range of industries. Today, XR technologies are also being adopted across diverse sectors, including manufacturing, healthcare, construction and even law enforcement (Marr, n.d.).

Challenges with using XR technology in a corporate setting

While there are many opportunities with XR technology, there are still underlying challenges that must be considered. In a study on virtual office work, Berlin and Babapour Chafi (2024) highlight some issues. One significant concern when using XR in a corporate setting, such as with Meta Quest headsets, is confidentiality. Using commercial XR headsets and associated softwares in organizational environments often results in data being processed on local servers, with files and other content uploaded to the cloud. This raises the risk of sensitive data being transmitted to unintended recipients. Additionally, confidential conversations, such as discussions of sensitive business information or details of personal health, could inadvertently compromise confidentiality and violate GDPR regulations. This poses a challenge of handling NDAs and sensitive information in XR environments.

Another issue discussed by Berlin and Babapour Chafi (2024) is the amount of required user accounts. The XR setup is described as a complex IT ecosystem, which may become inefficient for IT departments and create operational inefficiencies within organizations. Additionally, there are concerns about the reliance on external suppliers for essential applications. If a supplier decides to discontinue an XR application that a company has integrated into its workflow, this can lead to significant disruptions. It is vital for businesses to be able to maintain some control over the applications and devices used and ensure they can manage what is being used and displayed within their XR environments. Lastly, Berlin and Babapour Chafi (2024) emphasize the importance of understanding the specific value XR

technology can bring to a project. Before integrating XR, organizations should critically assess how it will enhance the work being done and what specific contributions they want the technology to make.

Future potential of XR technology

XR is rapidly growing and although its origins trace back many years, it is only in recent times that the technology has become commercially viable. This shift has led to a significant increase in usage and broader accessibility. An area associated with XR technology that has been frequently discussed in recent years and is expected to gain significant traction is the metaverse (Program-Ace, 2024). The metaverse is a vision of an interconnected extended reality made up of digital worlds, where users can seamlessly transition between experiences using virtual reality and augmented reality devices to interact, collaborate, and connect in real-time (Walsh, 2023). Although the concept was first introduced in 1992 by author Neal Stephenson in his novel *Snow Crash*, where he described a shared, immersive 3D virtual world, it is only in recent years that the metaverse has begun to emerge as a serious technological frontier. Major tech companies like Microsoft, Apple and Meta are now investing heavily to bring this digital future to life.

Other upcoming trends in XR include its increasing use for immersive educational experiences (Program-Ace, 2024). XR could potentially serve as a virtual classroom that allows users to learn and explore concepts that would be impossible or unsafe in the real world. Additionally, hyper-realism is gaining increased attention. Hyper-realism involves incorporating more senses, such as smell and touch, into the XR experience. This development blurs the line between virtual and physical realities as advancements in technology enable XR to replicate real-world sensations more convincingly.

2.4.2. Electroencephalography (EEG) Sensor Technology

An electroencephalography (EEG) sensor, also called neurosensor, is a device used to measure the brain's electrical activity by detecting small fluctuations in electrical current between the skin and sensor electrodes (Soufneyestani et al., 2020). The electrodes consist of small metal discs connected by thin wires, which are attached or adhered to the user's scalp (Johns Hopkins Medicine, n.d.). The signals are then amplified and displayed as a graph, an electroencephalogram, on a computer screen.

Available EEG sensor devices

EEG headsets are available in both wired (medical-grade) and wireless (consumer-grade) versions (Soufneyestani et al., 2020). In both cases, the recorded data is transmitted to a computer, either through a physical cable or via wireless methods such as Bluetooth. For medical-grade EEG devices, electrodes can be either saline-based or gel-based (Sabio et al., 2024). Gel-based electrodes, while more effective in terms of accuracy, require more time for application and may cause discomfort due to the gel sticking to users' hair and causing

inconvenience. Saline-based electrodes are therefore frequently preferred due to ease of use and quick setup time.

The decision between using medical-grade or consumer-grade EEG devices largely depends on factors such as budget, resources, and specific requirements. The consumer-grade devices are more lightweight, cheap and provide greater freedom of movement and ease to set up. The medical-grade options often involve a higher number of electrodes and a more precise placement in comparison. However, medical-grade EEG devices typically offer more stable connections and higher data transfer rates. As a result, medical-grade devices are more suitable for clinical use, particularly in diagnosis and treatment (Sabio et al., 2024). In contrast, consumer-grade devices are especially appealing to novice researchers or those seeking to gather data beyond the confines of a traditional lab environment.

Even though medical-grade EEG devices have traditionally been used, consumer-grade EEG devices have become more widely available and increasingly popular over the past decade (Sabio et al., 2024). According to Soufineyestani et al., commonly used EEG devices include medical-grade devices from Enobio and SMARTFONES as well as consumer-grade devices from Emotive, Imec, OpenBCI and Neurosky (Soufineyestani et al., 2020).

Current applications of EEG sensors

EEG technology has become widely used in neuroscience due to its ability to provide valuable insights into mental states, cognitive processes and even imagination (Soufineyestani et al., 2020). Researchers in various fields have thus taken advantage of this leading to EEG technology being used in many applications.

One of the most common applications of EEG technology is for brain-computer interfaces (BCIs), which use real-time EEG data to control and interact with mechanical or electronic devices (Soufineyestani et al., 2020). BCIs process brain activity and translate it into signals that can be used by external systems, enabling users to control devices using only their thoughts. This technology offers valuable support for individuals with mild to severe motor disabilities. Another primary application area for EEG technology is the medical field, where it (mostly through wired devices) is used to evaluate various brain disorders and diagnose conditions that affect brain activity, such as epilepsy seizures, Parkinson's disease, Alzheimer's disease, schizophrenia, anxiety, dyslexia and autism (Soufineyestani et al., 2020). Beyond diagnosis, EEG technology can be used to assess the brain's overall electrical activity and consequently provide valuable insights into the effects of trauma, drug intoxication or the extent of brain damage in individuals who are in a coma (Johns Hopkins Medicine, n.d.).

Another application of EEG technology is in neuroscience, encompassing both cognitive and behavioral areas (Soufineyestani et al., 2020). In cognitive neuroscience, EEG can be used to measure cognitive load, analyze brain activity during decision-making and task performance, detect sleep patterns, and explore how the brain responds to different scenarios. In behavioral neuroscience, EEG technology helps assess brain alertness during various situations and can

also measure levels of drowsiness, mental workload, and stress. A third area of neuroscience where EEG technology is applied is neurophysiology, particularly in studying fatigue and its impact on brain function. Other application areas are sport, fitness and meditation as well as for educational purposes (Soufineyestani et al., 2020).

Sabio et al. suggest that affordable and accessible neuroscientific solutions are becoming increasingly available to those outside the research community (Sabio et al., 2024). This trend is particularly relevant as technology becomes more and more integrated into daily life. According to the authors, it is in fact entirely possible that EEG devices could become a common tool in everyday life within the next few decades.

2.4.3. Emotion Interpretation Technology

Emotions can influence users' assessments of interfaces. The emotional state of users can impact their perceptions of an interface, their willingness to use it, and their overall satisfaction with the experience (Brave & Nass, 2008). Recognizing this influence, the need for emotion interpretation is becoming an emerging area of interest, particularly in understanding user behavior and decision-making.

Available emotion interpretation tools

Today there are various methods available for assessing emotions. However, most of the methods rely on self-assessment tools, where users rate their own emotional states. Examples of such tools include the Geneva Emotion Wheel, PrEmo and similar instruments (Güiza Caicedo & van Beuzekom, 2006). While there are few technologies claiming to objectively track emotions, two emerging approaches claim to interpret feelings and emotions based on specific measurements: one relying on sensors and the other utilizing artificial intelligence (AI).

One example of a company that offers sensor-based technology for emotion interpretation is Merkle (Merkle, 2019). Their system interprets emotions by analyzing the physiological responses they provoke. During testing, various sensors are placed directly on the participant's body to collect psychophysiological data of the following key indicators:

1. *Breathing Depth*: Indicates the level of active interest. Measured using a clamp sensor attached to the index finger.
2. *Electrodermal Activity*: Reflects emotional arousal, either positive excitement or negative stress, by measuring changes in hand perspiration. Recorded through electrodes placed on the palm.
3. *Heart Rate*: Provides insight into mental stress or relaxation by analyzing heart rate variability. Measured using electrodes placed on the left side of the chest.

4. *Brain Waves*: Reveal levels of mental stress and emotional engagement. Higher activity may suggest a stronger likelihood of negative emotional responses. Measured via electrodes attached to the head.
5. *Facial Muscle Activity*: Indicates emotional valence (whether responses are positive or negative) by detecting micro-movements in facial muscles. Measured using sensors at three points on the face, capturing subtle changes invisible to the naked eye.

In addition to these indicators, Merkle evaluates the participant's level of awareness, including interest and attention, to assess how appealing a product is perceived to be. To account for individual differences, baseline readings are taken before each session.

One example of a program that offers AI-based technology for emotion interpretation is Noldus FaceReader, an automated system for analyzing facial expressions (Noldus Information Technology, n.d.). It provides insights into how various stimuli affect emotional responses and identifies six recognized facial expressions: happiness, sadness, anger, surprise, fear, and disgust - along with contempt and a neutral state. Noldus FaceReader can capture dynamic, unconscious responses to stimuli or objects while the person is engaged with them. The program processes video recordings of participants to interpret emotional states based on analysis of facial muscle movements. It does this by employing a deep learning-based facial recognition system and constructing an artificial face model by analyzing approximately 500 facial key points. Facial expressions are then classified using neural networks trained on a dataset of over 100,000 labeled facial images.

Current applications of emotion interpretation

When it comes to application areas, emotion interpretation has found increasing relevance across various fields due to the significance of human emotional responses (Khare et al., 2024). Human emotions are dynamic cognitive and physiological states that arise in response to experiences, thoughts, or social interactions. They involve subjective feelings, cognitive processing, behavioral responses, physiological changes and communication. As a result, emotion interpretation is widely used in fields such as marketing, human-robot interaction, healthcare, mental health monitoring and security.

In the field of market research, customers cannot always accurately explain why they choose one product or design over another according to Pichardo (Merkle, 2019). While they may attempt to justify their decisions, these explanations could be unreliable due to underlying reasons not being consciously known. Emotion interpretation could offer a valuable lens in such cases, helping to uncover unconscious triggers of behavior. Pichardo further argues that emotion interpretation provides more objective insights than verbal responses alone, since emotions influence our actions but cannot be consciously controlled or fabricated.

Furthermore, emotions can provide valuable insights into conditions like fatigue, drowsiness, depression and pain (Khare et al., 2024). With advances in technology and the growing use of electronic devices, emotion interpretation is likely to expand across areas such as brain-computer interfaces, robotics, medical diagnostics, driving assistance, recruitment and patient care.

Challenges of emotion interpretation technology

Even though emotion interpretation technologies promise many benefits, psychologist and neuroscientist Lisa Feldman Barrett criticizes their scientific foundations (Barrett, 2017). She argues that the assumption that facial expressions reliably indicate emotions, such as smiling when happy or frowning when sad, is a myth. Barrett's research shows that only 30% of adults in urban cultures scowled when they were angry, with others displaying various reactions like crying, smiling, or showing no expression (Center for Law, Brain & Behavior, 2020). This low reliability suggests that facial expressions are not universally tied to specific emotions (Barrett, 2017).

Additionally, the same facial expression, such as a scowl, can indicate different emotions (e.g. confusion or concentration) highlighting the low specificity of such expressions. Despite this variability, people still categorize facial expressions as indicators of specific emotions, but these can be cultural stereotypes rather than universal truths (Barrett, 2017). Barrett emphasizes that the same emotion can manifest in radically different ways across individuals and contexts, and that identical facial movements can be seen during different emotional experiences. She also argues that individuals respond differently to emotions, and even the same person can exhibit varying physiological responses to the same emotion on different occasions. This does not imply that the body reacts randomly; rather that variation is the norm. Physical reactions have no inherent emotional meaning but instead, the brain interprets these changes in the context of a situation and constructs them into an emotional experience.

2.4.4. Additional Relevant Technologies

In addition to the previously described technologies, several other technologies that could be considered emerging in certain contexts, show potential for user evaluation. The following sections briefly outline some of these technologies: eye tracking, driver-monitoring systems (DMS), wearable technology, digital twins and AI.

Eye tracking

Eye tracking technology monitors the user's eye movements to identify where they are looking, the duration of their gaze and the path their eyes take. By analyzing the user's eye location, movement and pupil size at any given moment, the system can pinpoint areas of interest (Novák et al., 2023). While eye tracking may appear to be a recent innovation, the concept has existed since the 1800s (Leggett, 2010). However, the technology has undergone significant advancements and continues to evolve, making it increasingly popular for assessing the usability and user experience of digital interfaces and products.

A wide range of eye tracking devices are available on the market, including glasses and specialized bars that can be mounted on monitors or other surfaces (Novák et al., 2023). These devices typically operate using an infrared light beam that is invisible to the human eye, directed at the subject's face and eyes. The technology relies on two key reference points: the reflection of light from the retina and the reflection of light from the cornea. Eye tracking technology is costly and demands specialized expertise to operate effectively. Therefore, it is crucial to weigh the potential benefits against the investment required before deciding to implement it.

According to Smart Eye, eye tracking technology is continuing to grow and evolve (Smart Eye, 2025). It has expanded from being solely a research tool used in controlled laboratory settings to being applied in complex, real-world environments to study a wide range of behaviors. As demands change, so too does eye tracking technology, creating significant potential for future developments. Smart Eye identifies several key trends, including the integration of eye tracking as part of a multi-sensor approach, the move toward tracking in dynamic and unpredictable environments outside the lab, and its increasing use in safety and health monitoring.

Driver monitoring systems (DMS)

Driver-monitoring systems (DMS) are technologies designed to detect driver impairment and enable appropriate interventions (Koniakowsky et al., 2025). They monitor behaviors such as distraction and issue warnings when necessary. DMS operates by recording drivers' eye movements in real time, identifying signs of fatigue or inattention. The system determines the driver's glance location relative to predefined areas of interest using distraction detection algorithms. Glances toward the forward roadway are classified as attentive, while glances toward other areas, such as in-vehicle displays, are classified as distracted. When distraction is detected the DMS can, for instance, issue a warning to prompt the driver to refocus on the road.

Eye tracking is a key component in driver-monitoring systems, significantly expanding the range of functions these systems can perform (Anyverse, 2025). For instance, a DMS equipped with eye tracking can monitor driver drowsiness and fatigue, assess whether the driver is scanning their surroundings effectively and detect interactions with pedestrians. Additionally, it can evaluate passenger emotions, identify the presence of children in the vehicle to enhance their safety and even support in-cabin health monitoring. By analyzing eye behavior and interaction patterns, the system may also help detect early symptoms of medical conditions.

DMS is currently used in millions of vehicles, and looking ahead, the goal is to expand its capabilities to detect impaired or intoxicated drivers, with studies on this already underway (Lyrheden, 2022). From July 7th, 2024, all new vehicles are required to be equipped with a Driver Drowsiness and Attention Warning (DDAW) system, which is designed to detect signs of driver fatigue and inattention (European Commission, 2022). However, starting in 2026, all new and existing vehicles, both passenger cars and commercial vehicles, must be

equipped with an Advanced Driver Distraction Warning (ADDW) system. To comply with this regulation, all new cars will be required to incorporate a camera-based DMS (Lyrheden, 2024).

Wearable technology

Wearable technology refers to electronic devices that are comfortably worn on the body or attached to clothing (Shen et al., 2017). This type of technology goes beyond simply counting steps, measuring heart rate, or tracking sleep patterns (KMS Healthcare, 2024). Wearable devices today offer far more advanced capabilities. As one of the fastest-growing segments in the tech market, they are widely recognized as an emerging technology (Burmaoglu et al., 2018). These devices encompass a broad range of applications that collect and display real-time data related to health, movement and other sensory inputs.

Wearable technology has had a significant influence on the healthcare sector, where they play a key role in enabling continuous monitoring of vital health metrics (KMS Healthcare, 2024). Equipped with integrated health sensors, these devices allow for remote observation and early detection of potential medical issues, features that are particularly useful in the management of chronic conditions. As a result, they help facilitate more responsive care, improve access to personal health data and have the potential to lower healthcare costs. Beyond healthcare, wearable technology is also becoming increasingly popular in the commercial sector. Fitness trackers and smartwatches, for example, are now widely used by consumers seeking to monitor and improve their overall well-being through real-time feedback (Burmaoglu et al., 2018).

Other industries adopting wearable technology include aviation and space research. In aviation, wearables are used to monitor pilots' vital metrics, helping to identify and prevent abnormal physiological conditions such as impaired blood flow, oxygen deficiency or fatigue-related symptoms, which can affect performance across different types of aircraft (Bellamy III, 2020). In the space sector, wearable technology is widely used to monitor parameters for health research on astronauts. On the International Space Station, for instance, astronauts have worn specialized devices such as watches, headbands, and vests to help scientists study sleep quality, exercise effectiveness, cardiovascular health, and more (Gaskill, 2024).

Digital twins

The foundational idea of digital twins dates back to NASA's use of mirrored systems during the Apollo program (Glaessgen & Stargel, 2012). The term "digital twin" was officially introduced in 2002 as a virtual model of a physical system that is updated with real-time data. It can be a virtual representation of a physical object, system, person or process, intended for real-time monitoring, simulation, and analysis (McKinsey & Company, 2024). A digital twin works by a physical object being paired with sensors that monitor key aspects like energy output, temperature and weather conditions (IBM, 2021). These sensors transmit data to a processing system, which then updates the digital replica of the object in real-time. With this

data, the digital twin can run simulations, identify performance issues and suggest improvements. The goal is to leverage insights from the digital model to optimize the performance of the physical object.

Digital twins are used extensively in various industries such as manufacturing, automotive, aviation, energy, healthcare and logistics (Willige, 2022). Applications span across areas like establishing time frames, improving designs, tracking health indicators and improving product performance (IBM, 2021). As stated by RISE (RISE Research Institutes of Sweden, n.d.), digital twins have the potential to be the next major advancement in the development of a digitalized society. This is further emphasised by IBM (2021) that states the future of digital twins to be nearly limitless in a rapidly expanding market driven by increased demand. However, it is important to note that not all projects would benefit from digital twin technology, as it may place a high demand on resources and not be suitable for less complex products.

Artificial intelligence (AI)

AI can function independently but also serve as part of a solution, for example, in technologies such as eye tracking, emotion interpretation, XR and more. AI is expected to significantly shape the future of such technologies, for example by improving data processing and enhancing overall capabilities.

While artificial intelligence may seem like a relatively recent innovation, the concept dates back to early work by pioneers like Alan Turing and John McCarthy in the 1950s (Lawrence Livermore National Laboratory, 2021). The term artificial intelligence was chosen as a name for a project aimed at developing ideas on thinking machines. Defining AI at the moment is no easy task and there are numerous different definitions used (Sheikh et al., 2023). A common and broad definition of AI is as a technology that allows machines to replicate and simulate complex human abilities. Other definitions delve deeper into what such “complex human abilities” entails with suggestions like the capacity for environmental awareness, goal-directed behavior, action initiation and learning through feedback. The High-Level Expert Group on Artificial Intelligence (AI HLEG) of the European Commission defines AI as: Systems that display intelligent behaviour by analysing their environment and taking actions - with some degree of autonomy - to achieve specific goals.

AI operates by processing large volumes of data through algorithms, often based on machine learning or neural networks, to identify patterns and make predictions or decisions without being explicitly programmed for each task (Dwivedi et al., 2021). It is already widely used across various industries including healthcare, finance and consumer technology, and it is further enhancing the performance of already established technologies. Dwivedi states that as AI continues to evolve rapidly, with new opportunities constantly emerging, its significance and potential to impact multiple industries and society as a whole become increasingly evident. Stige et al. (2023) suggest that AI can be leveraged in key areas such as gaining insight into the context of use, identifying user needs, informing solution design, evaluating usability and supporting the development process.

Although AI presents vast opportunities as described above, the way forward is not clear and its implementation requires careful consideration. Challenges include high data and computing requirements, a lack of transparency in decision-making (the “black box” problem) and ethical concerns around privacy, fairness and accountability (Floridi et al., 2018).

3

Study Methodology and Process

The study was carried out in two main phases: *Phase one - Technology Scouting* and *Phase two - User Evaluation Testing*. The phases overlapped iteratively as research of the technologies and their potential applications continued throughout the study. The process is illustrated in Figure 3.

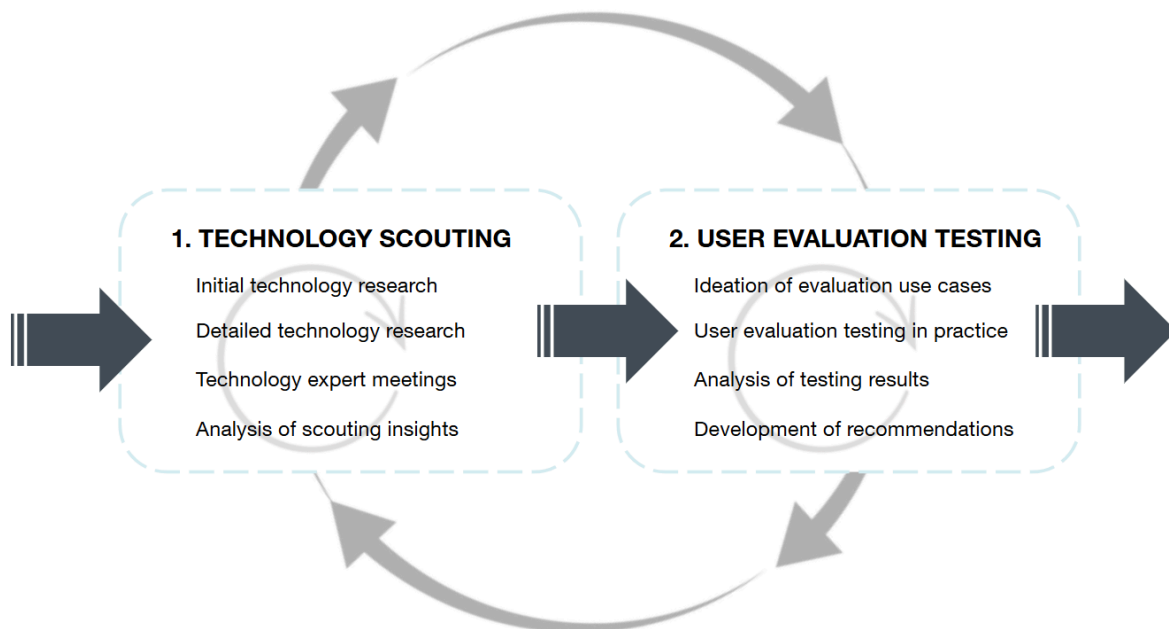


Figure 3. The iterative project process of Phase one and Phase two.

Phase one served as the foundation of the project, with phase two building upon its insights and findings. The first phase began with an initial research analysis that included background research and meetings with the team at Volvo, as well as relevant contacts at Chalmers. This work helped define the project scope, identify available resources, determine what could realistically be achieved during the semester and clarify the definition of emerging technologies.

Following this, the study entered a deeper research phase, focusing on a more detailed exploration of emerging technologies and their potential role in shaping future user evaluation methods. This also included a series of expert interviews with both internal Volvo stakeholders and external experts. The aim was to gain insight into current and upcoming technologies, their potential applications for Volvo and the challenges they may pose. Key

outcomes from phase one included a benchmarking of technologies, identification of opportunities and barriers as well as refined focus for the continued work.

Building on the findings from phase one, phase two explored how both new and existing technologies could be included in user evaluations. A central task was developing user evaluation methods that incorporated emerging technologies to uncover the types of insights they could generate. Two types of user evaluations were conducted: an XR workshop simulating early-stage product development and a driving simulator study incorporating emotion interpretation technology. The outcomes from phase two included concepts for user evaluation methods and insights from testing in practice.

The combined findings from phases one and phase two resulted in a set of recommendations to guide Volvo Group's future efforts and adoption in this area.

4

Phase One - Technology Scouting: Method

The aim of phase one was to mainly address the first research question: *How can emerging technology be used in future user evaluation methods?* Accordingly, this phase focused on exploring which technologies could be considered emerging, which have recently seen notable advancements and which hold significant potential for user evaluation. Throughout the entire phase, the perspective of Volvo Group and the implementation for user evaluations was kept in mind. The phase involved two main parts: data collection and data analysis. Iterations between the two parts took place continuously. The data collection involved expert meetings and extensive technology literature research, while the data analysis involved mindmapping, KJ-analysis and internal group discussions.

4.1. Technology Literature Research

The data collection included an initial research stage that developed into a targeted technology literature research on relevant technologies. In parallel, existing literature on user evaluation methods and prior studies in related fields was reviewed. The initial research stage involved a broad literature scan to identify emerging technologies relevant to the study. Based on insights from this scan, along with input from expert meetings conducted in parallel, a selection of technologies for further exploration was made. This led to the more targeted technology literature research, in which the selected technologies were examined in greater depth. This part of the literature research involved online searches using keywords such as “emerging technology in user evaluation”, “emerging technology in HMI evaluation”, “XR user evaluation” and “future applications for eye tracking”. The searches aimed to identify relevant information about the technologies and their potential applications and selected articles along with references were reviewed in greater depth to inform the study.

4.2. Meetings with Experts for Technology Knowledge

In parallel with the literature research, numerous expert meetings were conducted to gain deeper insights into practical usage of emerging technologies. Meetings were held with both internal Volvo stakeholders and representatives from external companies. The overall objective with the expert meetings was to assess how each technology is currently used and to explore its potential applications in future user evaluations, including associated challenges, opportunities and considerations. Internal meetings were conducted to gather insights into Volvo’s current practices regarding user evaluations and the integration of emerging technologies. External meetings with stakeholders from other organizations were carried out to provide benchmarking input and inspiration for potential opportunities and best

practices. Table 1 shows the distribution of meetings by technology, with parentheses indicating the number of internal meetings out of the total number of meetings.

Table 1. Overview of meetings in each technology area.

Technologies	Total # of meetings
XR	11 (5)
Emotion interpretation	2
EEG-sensors	2
Eye tracking	1
AI	1
User evaluations	4 (3)
Other meetings	2 (2)

The internal meetings were held with experts from the departments Volvo GTT Complete Vehicle, Volvo Group Design Advanced Design, Volvo Group Digital & IT Visualization & Global Support, Volvo Penta Digital Strategy & Business Office, Volvo Group Digital & IT Mobile & XR, Volvo Group Digital & IT Mobile, 3D and XR as well as Volvo GTT Human Behaviours & Usability. The meetings were conducted with senior roles specialized in areas such as design, strategy, management and engineering.

External meetings included employees and researchers from Chalmers University of Technology, specifically from the Design and Human Factors department and FUSE makerspace. Additionally, external meetings were conducted with the companies Smart Eye, Imitera, AI Sweden, RISE, VTI, Innobrain, Noldus and Oulu University in Finland. The meetings were conducted with roles such as tech leads, system engineers, chief officers, researchers and co-founders. Beyond the meetings, participation in an XR-related event hosted by XR Sweden was undertaken.

4.2.1. Recruitment of Experts for Technology Meetings

The expert recruitment aimed to include a diverse range of knowledge and expertise, covering different roles, assignments, and work experience. They represented different genders, ages, and nationalities. Some had general knowledge of XR, eye tracking, or user evaluations, while others specialized in areas such as Volvo’s use of XR, emotion interpretation applications, or EEG sensors.

A total of 19 experts or teams were consulted, with several participating in recurring meetings that either focused on the same technology or area or showcased expertise across more than one area. Of these, 10 were internal experts or teams from various Volvo Group

departments. Experts were selected based on their knowledge in areas such as user evaluation methods, specific emerging technologies or Volvo's internal processes. Experts were identified using a combination of sampling strategies. In some cases, experts available for meetings responded to inquiries sent to selected companies' general information addresses. This approach combined purposive and convenience sampling. In other cases, referral sampling was applied, drawing on recommendations from earlier meetings and input from the SGO team at Volvo. At other times, specific experts were deliberately selected based on their area of expertise, representing purposive sampling.

4.2.2. Technology Meeting Structure

An interview guide was tailored for each expert meeting, based on the interviewee's area of expertise and the specific objectives of the meeting. While the guides were adjusted slightly between participants, they all had a similar structure and followed a semi-structured approach, allowing for flexibility to explore interesting topics or perspectives during the conversations. In the initial stages of the study, the questions were broad and exploratory, aiming to gather foundational insights. Over time, as the understanding of each explored technologies deepened and knowledge gaps became clearer, the interview guides were refined accordingly to include more targeted and specific questions aligned with the evolving focus of the study. See Appendix A for an example of the full interview guide.

While several meetings were conducted in person, others were held via Microsoft Teams due to geographical distance or similar constraints. In-person meetings often included guided tours of relevant facilities and opportunities to interact with, or test, the technologies discussed. Meeting durations ranged from 30 minutes to two hours, with one or more experts participating. One member of the study's project team facilitated the discussion, while the other was responsible for taking notes. The meetings were primarily conducted in Swedish, although English was used occasionally when necessary. No audio or video recordings were made during these sessions.

4.3. Analysis of Technology Scouting

Given the wide range of technologies identified during the initial stage of data collection, an iterative analysis process was adopted. This involved a screening and prioritization process, ensuring that attention remained focused on the most promising and applicable technologies. The screening and prioritization process was conducted by continuous discussions within the project team. Technologies that, based on research or insights, were deemed no longer relevant to the study's objectives were excluded from further exploration. Conversely, technologies that demonstrated potential were selected for more in-depth analysis.

During the analysis, it became apparent that certain information was lacking, preventing informed and accurate decisions regarding the potential of some technologies. As a result, a mind map was created to highlight the corresponding knowledge gaps. The mindmap visualization led to the initiation of an additional technology scouting aimed at addressing those gaps.

Following the different research phases, the insights collected were analyzed using the KJ method. Insights and reflections from both the expert meetings and the literature review were transcribed onto color-coded post-it notes (representing internal or external insights) and grouped into thematic categories. The categories included the technologies themselves, user evaluation, opportunities, needs, challenges and other relevant themes (see Figure 4), facilitating the identification of patterns and emerging insights.

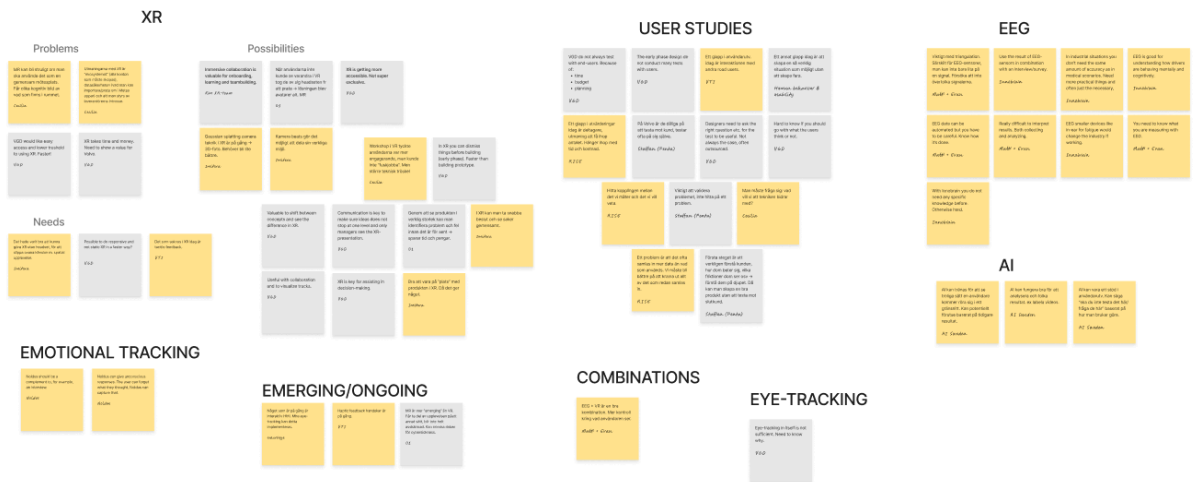


Figure 4. Overview of the KJ-analysis structure.

5

Phase One - Technology Scouting: Results

This chapter presents the results of the study's first phase, the technology scouting. The technology scouting consisted of technology literature research as well as expert meetings and covered a wide range of technologies across various industries. The presented result provides insights into the value emerging technologies could provide for user evaluations at Volvo Group.

5.1. Findings from Technology Scouting

The following technologies were initially identified as relevant in the study's context: extended reality (XR) technology, electroencephalography (EEG) sensors, emotion interpretation, eye tracking, artificial intelligence (AI), driver monitoring systems (DMS), wearable technology and digital twins.

In phase one, DMS, wearable technology and digital twins were excluded for further exploration despite their emerging nature and future potential in the automotive industry. This decision was made since DMS and wearable technology were considered less suitable for short-term user experience or HMI evaluations, but show greater potential for long-term driver monitoring, especially in health-related contexts. Similarly, digital twins were deemed as less suitable for short-term user evaluations and additionally classified as better suited as complements to user evaluations rather than a stand-alone solution. Moreover, there are two main technologies for emotion interpretation: sensor-based and AI-based. Given that sensor-based techniques (e.g., heart rate monitoring) are less emerging in this context and that AI-based emotion interpretation software is more easily accessible, the decision was made to proceed with the latter. Consequently, sensor-based emotion interpretation was excluded from further exploration.

5.1.1. Key Aspects and Potential Applications of the Identified Technologies

Following the exclusion of DMS, wearable technology, digital twins and sensor-based emotion interpretation, the remaining technologies were deemed relevant for further exploration in the next stages of phase one. Table 2 presents these technologies in terms of the key aspects measurements, current use, challenges and possible combinations. Table 3 outlines potential future applications of each technology across diverse areas related to user involvement in the automotive industry.

Table 2. The identified technologies: metrics, applications, challenges and combinations.

Technology	Key aspects
XR	<p>Measures: Mixed.</p> <p>Current use: User evaluation in immersive controlled settings for contextual insight. Also used in training, teambuilding, onboarding, collaboration, entertainment, monitoring, marketing, sales and as a virtual shared meeting space. Factory layout design. Reviews of technical product aspects and design concepts. Quick prototyping.</p> <p>Challenges: Potential struggles with stability and drifting of virtual objects (especially with “heavy” models). Precision of tools (eg. pointing and selecting). Not always seamless or quick enough to go from virtual 3D model into review in XR. Distracting due to low-fidelity (resolution). Lacking in haptic feedback. Influenced by the suppliers’ interest (eg. software updates removing certain tools). Costly to create detailed and high-quality 3D-models. Requires resources such as time and money. Potentially complicated and time-consuming setup, especially with multiple users. Initial learning-curve for tools and applications. Gamified feeling could lead to the experience being too engaging or non-professional. Experienced nausea. Ergonomic issues due to bulky hardware. Issues regarding hygiene when sharing headsets. Confidentiality issues if external application is used, GDPR.</p> <p>Possible combinations: Eye tracking, EEG sensors, emotion interpretation technology, wearable technology, AI.</p>
EEG sensors	<p>Measures: Brain imaging technique that uses scalp electrodes to measure brain activity.</p> <p>Current use: Primarily used in medical fields for diagnosis, but also applied in research on user experience and human-machine interaction for identifying cognitive load, fatigue, interest etc.</p> <p>Challenges: Involves lengthy setup, complex result interpretation and potentially bulky equipment. Almost impossible to interpret the results as a non-professional. Limited in detecting factors such as emotions and user preferences/experience. Need a lot of data for accurate measurements. Many factors can disturb the signals and affect results (e.g. blinking and clenching jaw). Some wireless devices are not accurate enough for user evaluation. May cause discomfort or distrust due to the feeling of being observed. Costly equipment.</p> <p>Possible combinations: Eye tracking, emotion interpretation technology, XR, wearable technology, AI, digital twins.</p>

<p>Emotion interpretation (using AI)</p>	<p>Measures: Muscle movements in facial expressions to infer emotions.</p> <p>Current use: To observe how individuals behave in interactions with others and to understand the subconscious feelings influencing their choices.</p> <p>Challenges: Users react differently to emotional stimuli (some may laugh when scared while others may appear neutral) which highlights a limitation in using AI, such as facial recognition, to interpret emotions. Needs precise camera set-up, dependent on good lighting. Extensive data output and complex result interpretation.</p> <p>Possible combinations: Eye tracking, wearable technology, AI, EEG sensors.</p>
<p>Eye tracking</p>	<p>Measures: Eye-movements, which can expose human behaviours often expressed subconsciously.</p> <p>Current use: To understand where users focus their attention, how they process information, make decisions and more. Supports the improvement of digital interfaces and the analysis of user and consumer behavior. Used in cars (DMS), simulators for tests and training, Face-ID, Face-Authenticator and in usability testing, among others.</p> <p>Challenges: Relies on the assumption that there is a correlation between eye fixations and cognitive processes which may not hold true for all users or products. Extensive data output and complex result interpretation.</p> <p>Possible combinations: Emotion interpretation technology, AI, XR.</p>
<p>AI</p>	<p>Measures: Mixed (can be part of several technologies and methods).</p> <p>Current use: Supports coding and testing, makes sense of large amounts of data, chatbots, data analysis, ideation feedback, inspiration.</p> <p>Challenges: AI can “hallucinate” by generating false information based on learned statistical patterns. If AI has been trained “wrongly” it can make misinterpretations. Bias can occur if only trained on a certain type of data.</p> <p>Possible combinations: Mixed (can be part of several technologies and methods).</p>

Table 3. The identified technologies: potential future applications.

Technology	Potential future applications
XR	<ul style="list-style-type: none"> ● Review early-stage designs. ● Enable quick concept comparison. ● Experience “pre-feasible” and future-oriented designs. ● Design in XR, eg. use AR to overlay sketches on physical objects. ● Evaluate non-static digital models responding to what users do. ● Evaluate anthropometry in a digital truck. ● Co-simulate interaction between truck and other road users to better understand traffic safety. Combine with motion tracking (suit, stripes or full-body estimation). ● Conduct global digital user evaluation. ● Conduct design workshops. ● Conduct virtual focus group discussions. ● Share physical environments (immersive feeling) with participants at different locations. ● Create XR experiences without headsets, via projections. ● Create “3D photographs” with Gaussian splatting camera technique (photorealism). ● Replace screens (in-vehicle) with lightweight interactive AR glasses. ● Provide customer support and service guidance. ● Create interactive marketing experiences.
EEG sensors	<ul style="list-style-type: none"> ● Evaluate and compare different design concepts. ● Observe user reaction to simulated scenarios in an XR environment. ● Evaluate cognitive load, mental workload, frustration. ● Evaluate audio or visual alerts in terms of user attention. ● Evaluate different ways of altering other road users (eg. pedestrians). ● Detect low driver engagement and trigger real-time alerts through headbands. ● Evaluate fatigue with lightweight in-ear devices. ● Assess pre- and post-drive performance.
Emotion interpretation (using AI)	<ul style="list-style-type: none"> ● Evaluate and compare different design concepts. ● Observe user reaction to simulated scenarios in an XR environment. ● Gather unconscious responses in user experience evaluations. ● Gather unconscious responses when a person is exposed to multiple stimuli. ● Replace surveys with camera data using emotion interpretation. ● Use as a basis for discussion in evaluation (complementary), after a test.

<p>Eye tracking</p>	<ul style="list-style-type: none"> ● Iris recognition in vehicles for secure access and unlocking. ● Adjust settings according to driver preference through driver recognition. ● Track the entire cabin and all passengers by improved monitoring systems. ● Analyze in-vehicle behavior and adjust settings as well as issue alerts needed. ● Detect driver fatigue and use an engaging AI companion to provide support. ● Detect user focus and automatically adjusts settings accordingly, eg. different parts of interface lighting up during driving in the dark. ● Provide interactive HMI feedback to guide driver attention and prevent false alerts or wrongful driving behaviour. ● Evaluate and compare different design concepts. ● Observe user reaction to simulated scenarios in an XR environment.
<p>AI</p>	<ul style="list-style-type: none"> ● Suggest relevant questions and identify gaps in user evaluations execution. ● Analyze and interpret evaluation results. ● Analyze data based on driver performance in real driver contexts. ● Detect health issues by face-analysis. ● Predict how users will interact with a product, “participate” in evaluations. ● Allow personalized user experiences.

5.1.2. Emerging Trends and Considerations in XR

XR is becoming more broadly accessible, moving away from its previous status as an exclusive or high-end tool and several aspects of the technology could currently be considered emerging. This includes improved avatars, increasingly realistic virtual environments, improved wireless headsets, the ability to simulate walking long distances within a confined physical space and the growing use of natural hand gestures instead of hand-held controllers. Opinions on hand gestures vs. hand-held controllers vary. While some value the immersive and intuitive feel of using their own hands, others prefer hand-held controllers due to their precision, lower learning curve and haptic feedback. A promising development in the area of haptic feedback is the emergence of tactile gloves, which aim to deliver a more natural interaction experience without sacrificing feedback quality. Another critical, yet often overlooked, aspect of immersive feel in XR is the development of realistic sound design. Some argue that high-fidelity audio plays a significant role in enhancing the sense of presence and overall realism in XR experiences.

While XR holds great promise, it is also important to acknowledge its current limitations, as considerations that can guide more purposeful and value-driven use. Thus, XR should be used selectively when it truly adds value. The value mostly lies in enabling immersive collaboration with virtual representations of products or spaces. This type of interaction can lead to faster decisions and clearer communication than traditional tools such as PowerPoint presentations, Teams meetings or Excel-based analysis. Full immersion, where all participants use headsets, is essential for realizing these benefits and partial participation through desktop interfaces often diminishes the collaborative and spatial advantages of XR. As such, participants in an XR experience joining in via computer should be an exception.

5.1.3. Practical Considerations in Applying EEG Technology

EEG technology is advancing, but its application remains complex, especially outside professional research. Both industry and academic voices agree that EEG signals are difficult to interpret without specialized expertise. Despite this, EEG technology is becoming more accessible, with wireless devices and systems that simplify data collection and interpretation entering the market. Using such innovations should be done with caution since some argue that they are sometimes marketed with exaggerated claims of resolution needed for serious research. On the contrary, there are arguments made that EEG intended for industrial use does not need as high resolution or wide range measurements of brain activity as medical research.

Since industry-focused research typically wants tools that are fast, reliable, and easy to deploy, lengthy and complex setups of EEG sensors remain a challenge. High-quality signals still depend on wet active electrodes, which provide the best data but require time-consuming preparation and are sensitive to variables like hair products or skin conditions. Semi-wet alternatives such as saltwater-dipped sensors offer a middle ground, reducing setup time but still requiring careful handling. While dry electrodes are the most convenient in these aspects, their signal quality is as mentioned often less sufficient for detailed analysis.

Another constraint of EEG technology is comfort, since wearing EEG sensors for over an hour can become uncomfortable, leading to reduced participant engagement and data quality. This creates a trade-off where longer sessions may provide more data (necessary for reliable testing), but only if participants remain attentive and at ease. There is also a risk of misconceptions among participants, on things such as that EEG technology can "read thoughts." These misunderstandings can cause discomfort or hesitation so to build trust, researchers can demonstrate how brain activity is measured by showing the signals generated from simple actions like blinking or jaw clenching. In most cases, this helps participants feel more at ease and often sparks fascination and a positive testing experience.

A promising but still-developing area is the integration of EEG with VR. While some manufacturers are exploring this, usability challenges remain. For the setups to work, the VR gear must be comfortable during extended use and avoid interfering with the EEG signals.

5.1.4. Balancing Emotion Input from User and System

When evaluating HMI by emotion interpretation, it is important to consider both user expression and the system (machine) context, since tracking facial expressions alone is not enough. Capturing what the machine is doing at any given moment, for example by using an additional camera, enables more accurate correlation between system behavior and user response.

Asking participants about their experience also adds depth and context that technology alone might miss. But, it is also worth noting that users might forget or misremember how they felt at some stages of an evaluated interaction and that their final emotional impression can overwrite initial reactions. For this reason, combining direct user input with objective data

(video analysis) can provide a more nuanced understanding. Thus, these methods are complementary and can together possibly offer a more complete picture of emotional engagement.

5.1.5. Opportunities and Concerns when Implementing AI

As AI technologies become more accessible and affordable, their use has expanded beyond traditional tech-centric fields. This has opened up new possibilities of AI but also raised important concerns around data transparency and ethics. Automated systems should not be implemented blindly, especially if it is unclear how data is being processed or analyzed.

In the context of user evaluation, there is some hesitation around using talking or guiding AI assistants. While they can streamline tasks, there is a risk of over-reliance and there are points made that AI could not replace the value of independent human thought (also applicable in AI-contexts outside of user evaluations). As one person put it: *"Nothing beats having to think for yourself"*.

One of the most promising applications of AI is the growing field of digital assistants, both for professional and personal applications. There is also untapped potential in personalizing experiences. However, some highlight that automated AI-personalization is often skipped by companies due to high development costs and limited perceived value. Looking ahead, AI tools that can collect and analyze large volumes of user data may fundamentally reshape how industries operate. By enabling deeper and faster insights, such tools could revolutionize how we understand behavior and develop products – if used responsibly.

5.1.6. Combining Eye tracking with Complementary Methods

While eye tracking is a well-established technology, it could still be considered emerging in certain contexts. The technology has only recently become more affordable, enabling its adoption across various industries. Furthermore, continuous advancements in eye tracking technology contribute to its ongoing relevance and potential. However, eye tracking on its own has its limitations. While it provides a visual mapping, such as heat maps showing where users look, it does not reveal the reason behind the gaze pattern. To truly understand user intent or emotion, eye tracking should be combined with other data sources, technologies or follow-up questions. Eye tracking is as such best utilized as part of a triangulation approach, complementing other user testing and observational methods rather than replacing them.

Some companies are focusing on non-intrusive eye tracking technology to observe user behavior. By designing tools that do not make users feel actively monitored, they can capture more natural and authentic interactions. There is also future potential in combining eye tracking with other non-intrusive methods, such as radar-based sensing, to gather data from a distance without the need of wearable equipment.

5.1.7. Considerations for Comprehensive User Understanding

In user research, quantitative methods like eye tracking, EEG, stress measurements and heart rate monitoring offer precise measurements, but they need to be combined with qualitative methods such as interviews and observations to create meaningful insights. Due to this, there is a need to triangulate methods to capture both objective and subjective dimensions. When doing so, starting with open-ended methods often helps identify the core challenges and make sure that additional measurements provide the sought insights. Without clarity on *what* is to be understood, even the most reliable data risks being irrelevant.

A major industry challenge regarding user evaluation is that large amounts of data that are collected risks remaining underused. This is highlighting the argument that there is potential in utilizing the data that is already being collected to a large extent, rather than collecting more. If the analysis of large amounts of data (eg. data collected from users driving their own vehicles) could be streamlined and made more efficient, it could have a transformative impact on the human factors field.

Another area of concern is the tendency to avoid real end-user testing due to confidentiality, logistical challenges, scheduling constraints, limited budgets and lack of expertise. As a result, global business decisions may be based on insights from a small sample or rely too heavily on assumptions rather than honest feedback. Additionally, the individuals conducting user evaluations are in some cases not the ones analyzing the results or making product design decisions. This disconnect can lead to important insights being lost in translation. To avoid this, designers should ideally be closely involved in user evaluations, observing behaviors and interpreting feedback in context. Ultimately, deep user understanding and strong early-stage research create the foundation for taking bold design risks later in the process. However, there is an added challenge in interpreting user reactions since truly innovative design can feel unfamiliar at first, but appreciated over time.

5.2. Phase One's Implications for Future User Evaluation

The analysis in phase one revealed several gaps in user evaluations relevant to this study. These insights are categorized below by identified needs and areas for improvement. The findings have important implications for future user evaluations and directly informed the ideation activities carried out in phase two of the project. The identified needs are:

- In-depth understanding of users
- Understanding of driver interaction with other road users
- A high feeling of reality in lab-simulations
- Effective and efficient collection and analysis of data

There is in some cases a lack of conducted user studies with end users, sometimes due to several resource issues. Important questions to address are: *How many users are targeted in*

evaluations? Who are the users targeted in evaluations? When in the process are user evaluations performed? Furthermore, the results of user studies are not always taken into account in design decisions and designers would prefer to be present in evaluations more often. Important questions to address are: What are the reliability of user evaluations? Is there a lack of trust in the results from user evaluations, why? Are user evaluations representative of the targeted user group?

Take-aways from Phase One: Technology Scouting

- Each emerging technology has the potential to offer significant value in future applications.
- Some emerging technologies still face significant challenges and require further refinement, while others are more mature and ready for immediate implementation.
- Combining and complementing different methods and emerging technologies can unlock benefits for user evaluation.
- Current uses of XR in user evaluation and various industries offer diverse benefits, from training to product design, but challenges include technical limitations, cost, setup complexity, and potential user discomfort, emphasizing the need for careful implementation.
- EEG technology is valuable for diagnosing medical conditions and researching user experience. Depending on the chosen device, the complexity, signal interference and potential discomfort pose challenges for effective and reliable use.
- Emotion interpretation technology is used to observe and understand subconscious emotional responses, but its limitations, such as varied user reactions and complex data interpretation, highlight the challenges in accurately interpreting emotions.
- Eye tracking helps analyze user behavior but faces challenges in data interpretation and the assumption that eye fixations always correlate with cognitive processes. For best results, it may be combined with other methods.
- AI supports coding, testing, and data analysis and has the potential to transform user research. Issues like "hallucination", misinterpretations and bias require careful implementation.

6

Phase Two - User Evaluation Testing: Method

The second phase of the project built upon the insights and outcomes from Phase One - Technology Scouting, with the primary aim of addressing the second research question: *What types of insights and results can be obtained from user evaluation methods that incorporate emerging technologies?* To explore this, the phase focused on ideating and testing emerging technologies in practice as part of user evaluation of human-machine interactions. This was done in order to better understand the types of results such evaluations could generate. Several types of user evaluations were designed, conducted and analyzed to assess their potential and effectiveness. An overview of the tests in practice are presented in Figure 5.

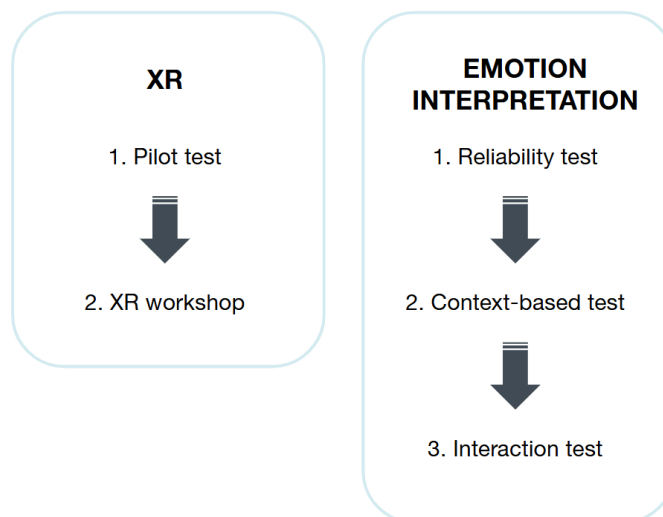


Figure 5. Overview of the user evaluations tested in practice.

6.1. Aim of Testing User Evaluation for HMI in Practice

Building on the gaps identified from the result in phase one, a more refined aim and objective for this specific phase of user evaluation testing were established. The aim of the user evaluation testing was to explore how emerging technologies can enhance user evaluation by enabling *earlier testing* in the design process, increasing *accessibility* for designers and allowing for *broader user participation*. Additionally, the study sought to investigate how these technologies can capture more *nuanced user insights* to ensure that valuable aspects are not overlooked. This was summarized in the overarching objective of assessing:

- How can the understanding of truck drivers as end user's needs and preferences be increased?

As more insights and information emerged, ideation sessions continued and evolved. A second round of ideation was conducted with a stronger emphasis on feasibility, potential value and emerging technology relevance of the use cases within the context of Volvo Group. As a result, the earlier ideas underwent a filtering process that eliminated irrelevant concepts, refined and developed promising ones and in some cases broke them down into sub-ideas or entirely new concepts. This process was informed by the *Technology Scouting Matrix* and the potential application areas identified in phase one. The methods employed during this round of ideation included brainstorming, mindmapping, brainwriting and mindmapping.

Two ideas were selected based on their potential, relevance, feasibility for immediate testing as well as alignment with the project's objectives. The first type of user evaluation consisted of an XR workshop simulating remote collaboration during early-stage product development, while the second type of user evaluation was a driving simulator study incorporating emotion interpretation technology. These evaluations were designed to have similarities between them, facilitating easier comparison. As a result, the focus of each evaluation was on assessing the user experience of interacting with both digital screens and physical buttons, aligning with the earlier goal of comparing low-tech and high-tech interfaces in the user evaluation process.

Moreover, since the project team was less familiar with emotion interpretation technology, this technology underwent preliminary testing to assess its reliability. In contrast, Volvo was already using XR technology, so no such uncertainty existed around its application.

6.4. User Evaluation Testing in Practice - XR Technology

The first type of user evaluation chosen for practical testing consisted of an XR workshop. The goal with the test was to observe user responses, determine whether the evaluation would generate new insights and evaluate its overall effectiveness. The following section outlines the process of employing and testing this type of evaluation using XR technology.

6.4.1. Development of the XR Workshop

The user evaluation employed XR technology in a workshop setting with end users to explore the potential for generating new insights. This method required the use of an XR-software, 3D files compatible with XR, separate rooms, XR-equipment and a detailed manuscript. Several different XR collaboration programs were considered before the choice was made on using Campfire 3D, an XR collaboration platform designed to facilitate intuitive, spatial communication around 3D models. It enables teams to interact with and review complex designs in a shared virtual space, regardless of their physical location or device type. The program features collaborative elements, enabling interaction through commenting, sketching, laser pointing, measuring and more. In Campfire, each participant had their own account and were represented as avatars with names displayed below. The 3D-file used consisted of a car instrument panel which was sourced from an Open Source 3D-model website (see Figure 7).



Figure 7. 3D-model of car instrument panel used in XR workshop (Urveshk2623, 2022).
CC BY 4.0

The workshop was conducted at FUSE, Chalmers, using standalone Meta Quest 3 headsets. The chosen technology for the workshop was mainly MR with some quick scenes showcased in VR.

6.4.2. User Recruitment for the XR Workshop

A pilot test was conducted prior to the XR workshop to assess its functionality and identify any potential technical issues. The pilot test involved two male Industrial Design Engineering students, both 25 years old, experienced drivers but with no experience with similar user evaluations. One had some prior experience with XR through gaming, while the other was unfamiliar with the technology. They were recruited through convenience sampling, with a preference for participants with an educational background in design, as their perspective could provide valuable insights into how designers - who may also be considered end users of the developed user evaluation methods - would perceive the workshop.

After the pilot test, the XR workshop was conducted once with two truck driver students from a school in Gothenburg. The chosen number of participants were based on the possibility for the project team to efficiently facilitate the workshop as well as the number of available XR equipment. The drivers, aged 22 and 29, were both male and had been studying to become truck drivers for seven months. They were recruited by contact with the school, which provided contact information of interested students, thereby employing a combination of convenience and purposive sampling.

6.4.3. Evaluation Approach and Execution of the XR Workshop

In the XR workshop, one team member acted as the facilitator, while the other provided technical support. Participants, as well as the facilitator and technician, were placed in individual study rooms at FUSE, Chalmers, to simulate remote testing conditions. The facilitator used an XR headset and joined the participants in the session virtually, while the technician joined via computer. The evaluation was recorded from the facilitator's headset.

The technician assisted with setup and monitored the session, offering support in case of any technical issues. The workshop followed a manuscript, found in Appendix B, and was conducted in six steps as listed and explained below.

1. *First introduction (physical environment)*: The workshop began with participants signing GDPR consent forms, followed by an introduction of the project team and the study. Each participant stated their name, age, truck-driving experience and prior use of XR technologies. The test procedure was explained in detail, including how to use the XR-headset and controller. Participants were assured that there were no right or wrong answers and that the goal was to understand their perceptions and experiences.
2. *Second introduction (XR)*: The XR-session began in an empty virtual room where participants and the facilitator introduced themselves again, this time as avatars. The facilitator explained the available tools within the program and how to access them.
3. *Providing context (XR)*: To set the context for the workshop, a VR environment featuring a truck exterior and interior was shown (see Figure 8). Participants were asked about their familiarity with Volvo trucks and similar panels.



Figure 8. Contextual scenes in XR (Volvotrucks, 2025; Volvotrucks, 2020).

4. *Brainstorming (XR)*: The session continued in the empty virtual room (same as second introduction) and participants brainstormed ideas for an ideal instrument panel.
5. *Evaluating instrument panel (XR)*: The participants were shown the 3D car instrument panel and encouraged to share their opinions verbally as well as by using the program's comment function.
 - a. The participants explored the instrument panel interactively. They sat in front of the steering wheel to assess the layout, reachability and overall feel. One by one, they conducted the task of changing the radio station via different screens and buttons, while technical support played music through the speaker.

- b. The participants provided feedback on various layouts, placements and color schemes related to the instrument panel. The facilitator adjusted the placement of elements in real time while participants remained seated in front of the virtual dashboard, see Figure 9 below.
- c. At the end of the XR session, participants engaged in a brief summary reflection on what had been explored and discussed.

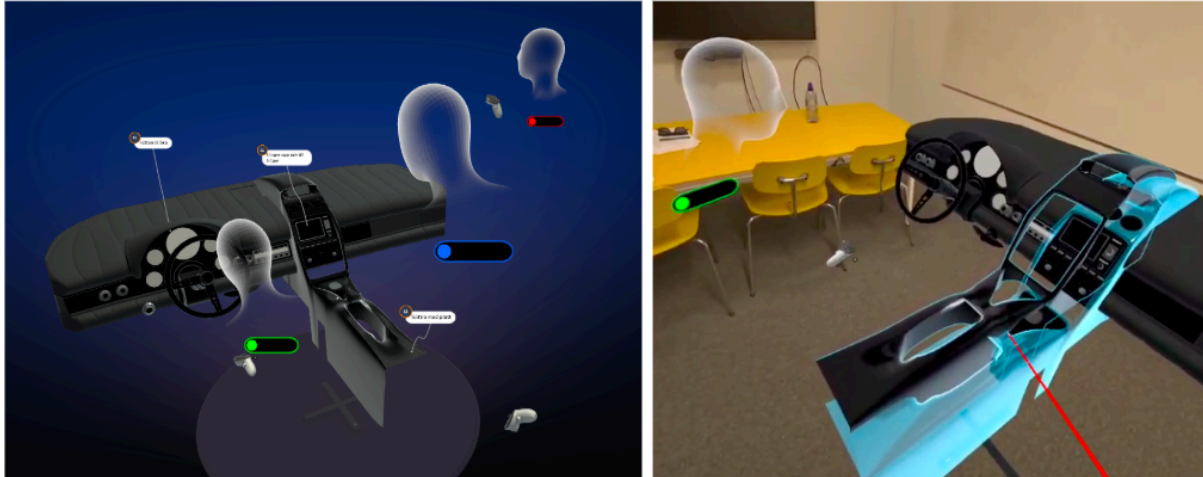


Figure 9. Adjusting placement of elements in XR workshop.

6. *Reflection (physical environment)*: The XR session ended and the participants and project team gathered for a joint reflective discussion, where participants were asked questions like how they felt about the workshop, their thoughts on the outcome, how they believed the results would differ if the workshop had been held physically and what changes they would suggest for a future session. The workshop concluded with the distribution of thank-you gifts.

6.4.4. Analysis of Outcome from the XR Workshop

The pilot test was analyzed through discussions within the project team. Technical issues and participant feedback were reviewed in the evaluation setup. Similarly, the tests conducted with truck driver students were analyzed using team-based discussions. The video recording from the facilitator's headset was transcribed and the transcripts were reviewed and highlighted for noteworthy quotes, observations and participant perceptions. The findings were then discussed within the team, focusing on potential sources of error, direct participant feedback, additional insights or observations noted by the team during the sessions and suggested improvements.

6.5. User Evaluation Testing in Practice - Emotion Interpretation Technology

The second technology selected for practical testing was emotion interpretation technology. The process began with a reliability test and a context-based test to evaluate how trustworthy the technology was and how well it functioned in an automotive setting. This was followed by a user evaluation focused on interaction with a truck instrument panel. Similar to the XR workshop, the aim was to observe user responses, determine whether the evaluation generated new insights and assess the overall effectiveness of the method.

6.5.1. Reliability Testing of Emotion Interpretation Technology

Due to concerns regarding the reliability of AI-based emotion interpretation technology, as well as its value in the context of user evaluations, a reliability test was conducted using the AI-based software Noldus FaceReader. The test aimed to evaluate whether Noldus could accurately identify emotions experienced by users in controlled scenarios. To achieve this, a video composed of three distinct clips was created, each intended to evoke a specific emotion: sadness, happiness, and disgust. The video had a total duration of 2:30 minutes. The first clip portrayed a teacher struggling during the COVID-19 pandemic, later surprised by a thoughtful gesture from students via Zoom. The second clip featured humorous and clumsy behavior from babies and young children. The final clip showed a time-lapse of food decaying, aimed at inducing disgust. The participants watched the video by themselves on a computer in a quiet, undisturbed room. Their facial expressions were recorded using the computer's built-in webcam. During screening of the video, the participants were aware that they were being recorded. After viewing the video, participants were asked to self-assess their emotional responses using the PrEmo instrument (Laurans & Desmet, 2017) and explain their feelings (see Figure 10 below). A detailed overview of the PrEmo instrument and self-assessment tool can be seen in Appendix C and Appendix D respectively.

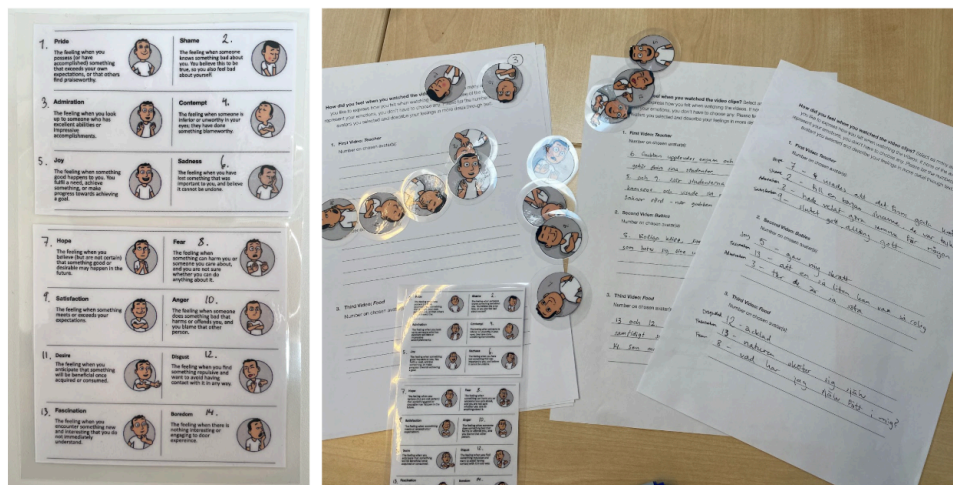


Figure 10. PrEmo instrument and self-assessment for reliability test (Laurans & Desmet, 2017). CCBY-NC-ND4.0

User recruitment for the reliability test

In the reliability test, four participants were involved, all of whom were students in the Industrial Design Engineering master's program at Chalmers University of Technology. Their ages ranged from 23 to 27 and they represented different ethnic backgrounds. This group was selected through convenience sampling.

6.5.2. Context-Based Testing of Emotion Interpretation Technology

The second test, a context-based evaluation, aimed to explore the applicability of emotion interpretation technology in an automotive setting, specifically with truck drivers. It sought to determine whether emotional responses could be effectively triggered and detected during a simulated driving scenario. The participants in this test were shown a video with three clips from the game *Euro Truck Simulator*. The clips simulated the driver's perspective in three different traffic scenarios: a calm scenic drive, a sudden crash and a congested rush-hour situation. The total duration of the video was 2:12 minutes. The video was displayed on a large TV screen while participants sat by themselves in a quiet, undisturbed room. Their facial expressions were recorded and analyzed using Noldus Facereader via a portable camera placed directly in front of the participant (see the complete setup in Figure 11 below). The participants were aware that they were being recorded and that their facial expressions would be analyzed. After watching the video, participants were asked to self-assess the emotions they experienced by using the Geneva Emotion Wheel (Scherer, 2005; Scherer et al., 2013; see Appendix E). During the self-assessment, participants were encouraged to think aloud and explain their choices.



Figure 11. Setup for context-based test.

User recruitment for the context-based test

The context-based test included two participants, both male, who held military truck-driving licenses. Both participants were experienced drivers and 25 years old. Since the test was conducted in the context of truck driving, it was important for participants to have some experience with trucks and, as such, these participants were recruited through a combination of convenience and purposive sampling.

6.5.3. Interaction Testing of Emotion Interpretation Technology

Following the reliability and context-based test, a more comprehensive final method for user evaluation incorporating emotion interpretation technology was developed and tested. This evaluation focused on interpreting emotions during human-machine interactions in a driving simulator. It required several components, including an interactive simulator with a screen and instrument panel, emotion interpretation software, cameras and a detailed evaluation plan. A basic driving simulator was used in a lab at Chalmers. It consisted of a seat, steering wheel, and pedals connected to a computer as shown in Figure 12.



Figure 12. The original driving simulator setup.

A suitable instrument panel for the evaluation was identified during a visit to a Volvo Truck Center in Kungälv. The selected instrument panel, shown in Figure 13, had previously been part of a test vehicle of model FM Electric. This model was released for sale in December 2022, with production starting in March 2023. To adapt the panel to the simulator, modifications were required, as the original steering wheel opening was too small. As a result, the panel was partially dismantled by the project team (see Figure 14), then to some extent reconstructed using foamboard and glue, and ultimately mounted using foamboard, glue, wire, wooden parts and clamps (see Figures 15).



Figure 13. FM Electric instrument panel. Figure 14. Partially dismantled instrument panel.



Figure 15. The assembly process of the driving simulator.

To simulate a functioning instrument panel, a Wizard of Oz method was employed, in which the system appeared autonomous but was actually controlled manually behind the scenes. An iPad was mounted on the instrument panel and served as a stand-in touchscreen, displaying a pre-designed interface created in Figma. A Bluetooth-enabled wireless mouse was utilized to simulate screen interactions while the participant used the physical buttons on the instrument panel. The simulator was connected to a TV screen and ran Euro Truck Simulator 2, a vehicle simulation game that allows users to drive trucks across realistic European road networks. The game features detailed environments, traffic systems and vehicle controls. The fully assembled driving simulator is shown in Figure 16.



Figure 16. Fully assembled driving simulator.

User recruitment for the interaction test

In the interaction test, a pre-test was conducted within the project team to evaluate the procedure and identify any potential issues prior to the main testing phase. During this test, it became clear that one person in the project team needed to serve as technical support, managing the camera, initiating recordings and playing music. The other person in the project team assumed the role of facilitator, responsible for asking questions, assigning tasks and overseeing the process.

Following this, four tests were conducted with students at Chalmers University of Technology. As the primary objective was to evaluate the effectiveness of the method and the insights it could generate, the inclusion of truck drivers was considered advantageous but not essential. Due to time constraints, four students from various academic programs, age groups, and ethnic backgrounds were recruited through convenience sampling. The main inclusion criterion was possession of a valid driver's license and a relatively high degree of driving experience. The participants, aged 24–28, comprised three males and one female. Their driving experience ranged from 5 to 10 years, and all participants held a category B driver's license, with one participant additionally holding a truck license. The majority of the participants had no prior experience with driving simulators or similar evaluations, although one individual reported limited experience with Formula 1 games using a basic home simulator setup.

Evaluation approach and execution of the interaction test

The interaction test was conducted with one participant at a time, with each session lasting approximately 45 to 55 minutes. All tasks performed in the interaction test were recorded using a Smart Eye eye tracking camera (see Figure 17), connected to a dedicated computer and controlled from a terminal. The recording began as soon as each driving session in the test started, and time-stamps for when each task started was noted.

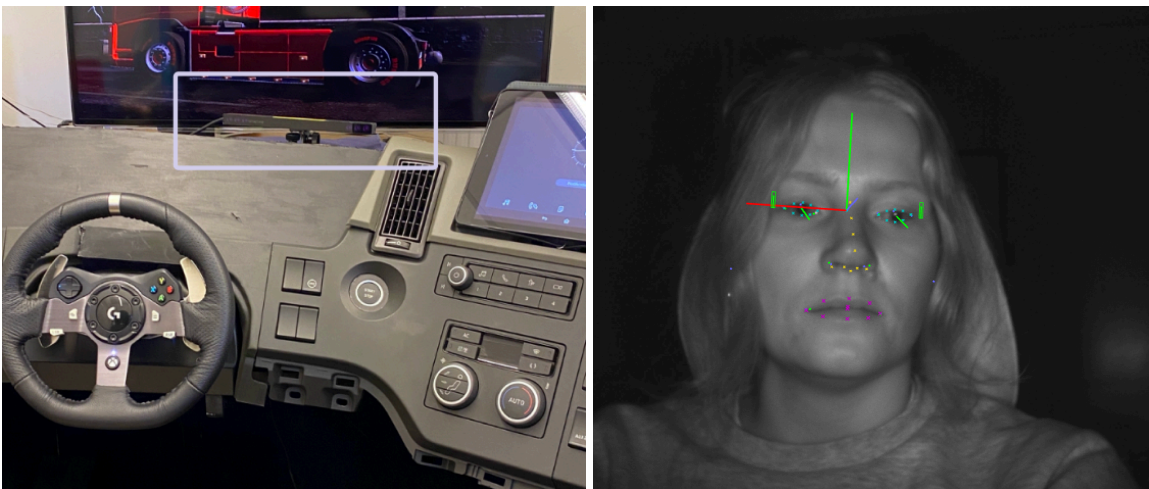


Figure 17. Placement and appearance of camera.

One member in the project team acted as technician in charge of time and camera procedures, whereas the other team member facilitated the test by providing participants with instructions and support during self-assessment. The test followed a manuscript, found in Appendix F, and was conducted in four steps as listed and explained below.

1. *Introduction*: Each session began with the participant signing GDPR consent forms, followed by a brief introduction of the project team and the study. The participants were informed that the test would be recorded and later analyzed using AI-based software. Participants were asked to provide information including their name, age, occupation, duration of driving license ownership, type of license held and whether they had previously participated in a similar evaluation or used a driving simulator. The test procedure was explained in detail and participants were given the opportunity to ask questions. Participants were assured that there were no right or wrong answers and that the goal was to understand their perceptions and experiences.
2. *Driving simulator settings*: The participant was asked to sit in the driving simulator and make themselves comfortable, with the option to adjust the seat as needed. The functionality of the instrument panel, as well as the operation of the pedals, joystick, and steering wheel, was explained and the participant was given a brief opportunity to test the simulator to familiarize themselves with the driving environment.

3. *Performing test tasks*: Each participant was assigned three different tasks, which were performed two times: once using a digital touchscreen interface and once using physical buttons. The tasks were the following:
- Change radio channel
 - Connect the phone via Bluetooth
 - Decrease the temperature by 1.5 degrees Celsius

For each round, the facilitator introduced a brief scenario and instructed the participant to perform a specific action while driving. The order of tasks was randomized for each participant, but each task was performed in pairs of both approaches (digital and physical) before moving on to the next one. Figure 18 showcases a participant carrying out a task using the two different approaches.

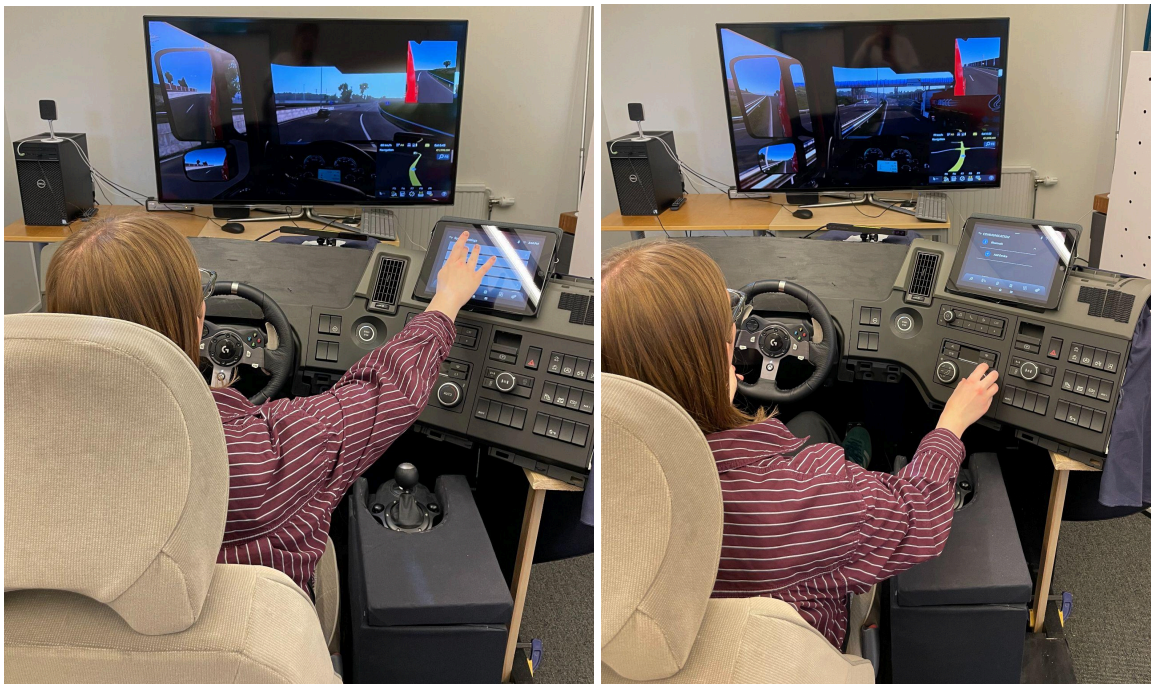


Figure 18. Performing tasks using both the digital and physical approach.

After each task, the game was paused, and the participant was handed an adapted version of the Geneva Emotion Wheel, adapted by the project team (Scherer, 2005; Scherer et al., 2013; see Appendix G). The emotions in the adapted Geneva Emotion Wheel were selected to represent the emotions in Noldus FaceReader along with some additional feelings predicted as relevant for the context. The participants had the option to select “none” if they did not feel any particular emotion and could add additional emotions if they felt something was missing in the wheel. After completing the self-assessment, the next task began. This process was repeated until all three tasks (six approaches) were completed, along with their corresponding self-assessments.

4. *Reflection*: After the completion of all tasks, a reflective discussion followed. The discussion touched on how the participants felt during each task with elaboration of the self-assessments, their perception of the instrument panel and the different ways of executing the tasks. Additionally, they were asked questions regarding opinions on the evaluation itself, including questions about how they felt about the process, their experience of being recorded and what changes they would suggest if they were to participate in the evaluation again. The workshop concluded with the distribution of thank-you gifts and fika.

6.5.5. Analysis of Outcomes from the Interaction Test

In the reliability test presenting video clips designed to evoke emotions, data from Noldus and participants' self-assessments were analyzed and compared to evaluate how well the AI-generated emotion data aligned with the participants' reported experiences. Similarly, in the context-based test presenting truck-related video clips, data from Noldus and the Geneva Emotion Wheel were analyzed and compared to assess how well the AI-generated emotion data aligned with participants' self-reported experiences. The results and the effectiveness of the technology were, in both cases, discussed within the project team.

The third and final test, which involved emotion interpretation technology during human-machine interactions in a driving simulator, was analyzed using a similar approach. First, video recordings from the Smart Eye camera were converted to MP4 format and imported into Noldus. Since the Smart Eye system incorporates emotion interpretation, it also allowed for observation of emotion interpretations. These interpretations were compared with those generated by Noldus and, subsequently, with the participants' self-assessments from the adapted Geneva Emotion Wheel. During the analysis, only the segments of the recordings corresponding to task execution were reviewed, based on the noted time frames from the testing sessions. Comments on behaviors or other factors that could influence either the software-generated or participant-reported interpretations were noted and discussed within the project team.

7

Phase Two - User Evaluation Testing: Results

This chapter presents the results from the second phase of the study, the user evaluation testing. The phase involved conducting various types of user evaluations to explore the insights that could be generated and assess the added value of emerging technologies. The results provide an overview of key findings from the tested evaluation approaches.

7.1. Listing of User Evaluation Use Cases

As part of the ideation process in Phase Two, a set of potential use cases for emerging technologies in user evaluation was developed. The term use case refers to a scenario intended for user evaluation. The outcome was a matrix outlining use cases in which each emerging technology could be applied to enhance user evaluations (see table 4). In this context, a *use case* refers to a conceptual scenario illustrating how the user evaluation might be conducted, including relevant procedures, technologies involved and intended insights.

Table 4. Use Cases for Emerging Technologies in User Evaluation.

Technology	Use Case Title	Description
XR: VR	<i>Hyper-realistic testing</i>	Test user interaction in a hyper-realistic virtual truck cabin, engaging multiple senses to simulate the real driving experience.
	<i>VR focus group (metaverse-inspired)</i>	Early-phase collaborative feedback in a shared interactive VR space. Enables real-time cross-department communication and more realistic evaluation of subjective aspects like semantics.
XR: MR	<i>Concept evaluation in MR</i>	Use MR to overlay and evaluate different instrument panel designs in a real truck. Enables realistic, iterative concept testing with co-design features like annotations and sketches.
	<i>Observation study with sensor integration</i>	Observe user reactions in an MR cabin with dynamic elements. Sensor data (e.g. heart rate) captures responses during free exploration.

EEG sensors	<i>Distraction measurement</i>	Use consumer-grade EEG (e.g., headbands) to detect cognitive overload during driving. Combine with eye tracking or emotion interpretation technology to triangulate and better understand user responses.
	<i>Objective UX evaluation</i>	Use EEG to gather objective data on user experience during human-machine interactions.
	<i>Attention to signals and warnings</i>	Evaluate driver attention and interest in response to different warnings and signals.
Emotion interpretation (using AI)	<i>Emotion-guided interface evaluation</i>	Combine with eye tracking to link emotional responses to specific UI elements. Complement with interviews and traditional methods. Use AI to adjust the interface in real time based on feedback. Enables quick iteration but requires a calm pace and clear voice guidance. Risk of misinterpretation by AI.
	<i>Experience evaluation</i>	Assess the emotional journey during product interaction and explore whether user preferences and overall experience can be objectively captured.
AI	<i>Adaptive user interfaces with AI</i>	Leverage AI to personalize and adapt the interface in real time based on user behavior and preferences. Enables rapid iterations, improved UX, and support for diverse user needs.
	<i>AI-driven interpretation and data analysis</i>	Utilize AI to interpret user expressions and analyze large datasets to summarize results effectively. This enables the extraction of meaningful insights from complex data, supporting informed decision-making in user evaluations.
	<i>AI-guided user evaluation</i>	AI autonomously conducts user evaluations, asking standardized questions and performing the entire test, ensuring consistent and efficient data collection.
	<i>AI-driven inspiration for designers</i>	AI generates diverse design concepts, inspiring creativity and expanding possibilities during the ideation phase.
	<i>Early interface prediction with AI</i>	AI analyzes user interaction patterns within an existing interface and predicts how users might behave with a new design. This approach allows early-stage testing and feedback, potentially saving time and resources by reducing the need for extensive user testing throughout the design process.

Eye tracking	<i>In-vehicle gaze tracking</i>	Use embedded cameras to monitor gaze during driving, enabling remote analysis of natural user behavior. Designers review recordings post-session.
	<i>Showroom attraction study</i>	Analyze user attention in a showroom setting using embedded cameras to track gaze and interaction. Identify which features attract interest (e.g., buttons, solutions). Follow up with user feedback and post-session analysis by designers. Can be combined with sensors or MR for deeper insight.

Building on the use cases presented above, additional ideation sessions were held, which led to the development of two distinct types of user evaluations for further testing. These developed user evaluations were selected for testing because they were considered the most promising in the study’s context, feasible for implementation at the stage of the study as well as offered significant potential value to Volvo Group. The chosen user evaluations include:

1. An XR workshop simulating remote collaboration during early-stage product development.
2. Using emotion interpretation in a driving simulator to observe emotional responses during various human-machine interactions.

The XR workshop specifically, was chosen with the hope of creating remote opportunities for designers by making the evaluation more accessible. If successful, it would enable online collaboration, increasing user (and designer) participation and allow for earlier user evaluations. The driving simulator, paired with emotion interpretation, was also chosen with the hope that it could capture more nuanced and reliable user insights early in the design process, potentially offering a deeper understanding of user experiences.

7.2. Insights from XR Workshop Testing

This user evaluation focused on using mixed reality (MR) within an XR environment to conduct a workshop with end-users during early-stage product development. The aim was to explore the type of insights that could be gathered through this approach. Both insights from the pilot test and the following main test is presented below.

7.2.1. Early Learnings from XR Workshop Pilot Test

During the pilot test of the XR workshop, participants shared many opinions and perceptions related to the instrument panel, including both low-tech and high-tech interactions, as well as feedback on the workshop itself.

Firstly, during the initial brainstorming session on what constitutes an optimal instrument panel, the participants discussed features typically appreciated or disliked in such panels based on personal experience. The discussion included types of buttons, placement of functions and the combination of haptic and digital elements. Secondly, when introduced to the specific instrument panel setup in the virtual environment, both participants instinctively brought forward a nearby chair (in the physical environment), originally intended for a later stage of the test, and positioned themselves in front of the steering wheel. Without prompting, they began exploring the setup, checking whether they could reach various buttons and screens while verbalizing thoughts throughout the process. Comments were made on the appearance of the panel, and specific design improvements were suggested.

At a later stage in the session, suggestions on colors and the placement of various elements were iteratively adjusted together with the facilitator. The workshop proved effective for this purpose even though, for one participant, some color changes did not update correctly in Campfire, which made meaningful feedback more limited. Another issue arose when writing comments in Campfire. For both participants, the keyboard field occasionally disappeared, and the controller's pointer could not properly reach all the letters. It appeared as if the keyboard had "sunk" into the panel rather than being displayed on top of it, which caused interaction difficulties (see Figure 19 below). It was later discovered that the comments could be dragged out with the pointer to avoid this issue.

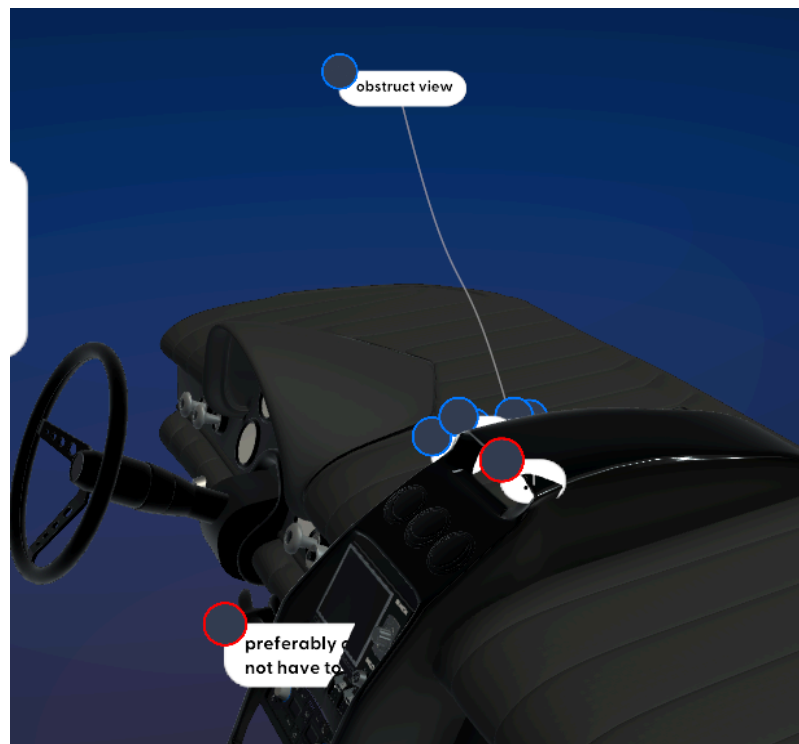


Figure 19. Displaying difficulties with commenting.

Another aspect that emerged during the pilot test was that one participant found the experience slightly uncomfortable. Although they could see their physical room (see Figure 20), it felt as though it was behind a screen, creating a sense of detachment. The experience was described as “feeling like being inside a video game” with a sense that “a monster might suddenly appear”.



Figure 20. MR view.

Based on findings from the pilot test, several adjustments were made to the procedure. One of the adjustments addressed the issue of commenting. Due to the comments taking longer to complete than anticipated, it was decided that participants could comment during the initial phase of the workshop, but then having the comment tool handled by the facilitator and technician. Other adjustments following this change was to inform the participants that their comments had been saved via a screenshot, before removing them, ensuring that their effort was still valuable even though the comments disappeared shortly after.

Prior to joining XR, both the technician and the facilitator helped participants set up in separate rooms. Despite the instrument panel being set in the correct position, it had been moved for one participant later in the evaluation. This issue was discussed within the project team and the conclusion was that the technician should assist both participants while the facilitator logged in, ensuring that the position issue was not related to the timing of the facilitator’s login as this person was considered the “owner of the project”.

7.2.2. Insights from the XR Workshop Test

The main test was conducted with two truck driver students and this test also generated many interesting perspectives. As in the pilot test, the initial brainstorming session on what constitutes an optimal instrument panel led to valuable insights from both participants,

although they were slightly less talkative than the participants in the pilot session. When shown the specific panel, both participants highlighted different preferences for the desired interface; one preferring more physical buttons and the other digital controls.

During the part of the workshop where participants interacted with the instrument panel, the experience was not as engaging as anticipated. One reason was that the music did not play as intended, due to the microphone failing to transmit audio, which may have reduced the realism of the scenario. However, the participants still provided valuable feedback during the part of interaction. When interacting with the instrument panel, it was mentioned that the screen and buttons felt too far away from the seated position. In response, the facilitator began adjusting the layout, trying different placements. This iterative approach proved helpful, as the participants expressed that it made them feel involved in the design process and engaged in providing feedback. For example, they used concrete input like “*there*” to indicate preferred screen placement, rather than more general comments such as “*I would have liked it higher*”. See the feedback provided during this part of the workshop in Figure 21 below. In some other instances, the participants provided relatively limited responses, initially offering simple comments such as that it “*looked good*”. However, as materials and colors were altered, their feedback became more detailed, with statements such as, “*It should not be glossy*”, or “*Oh, that actually looks better*”.

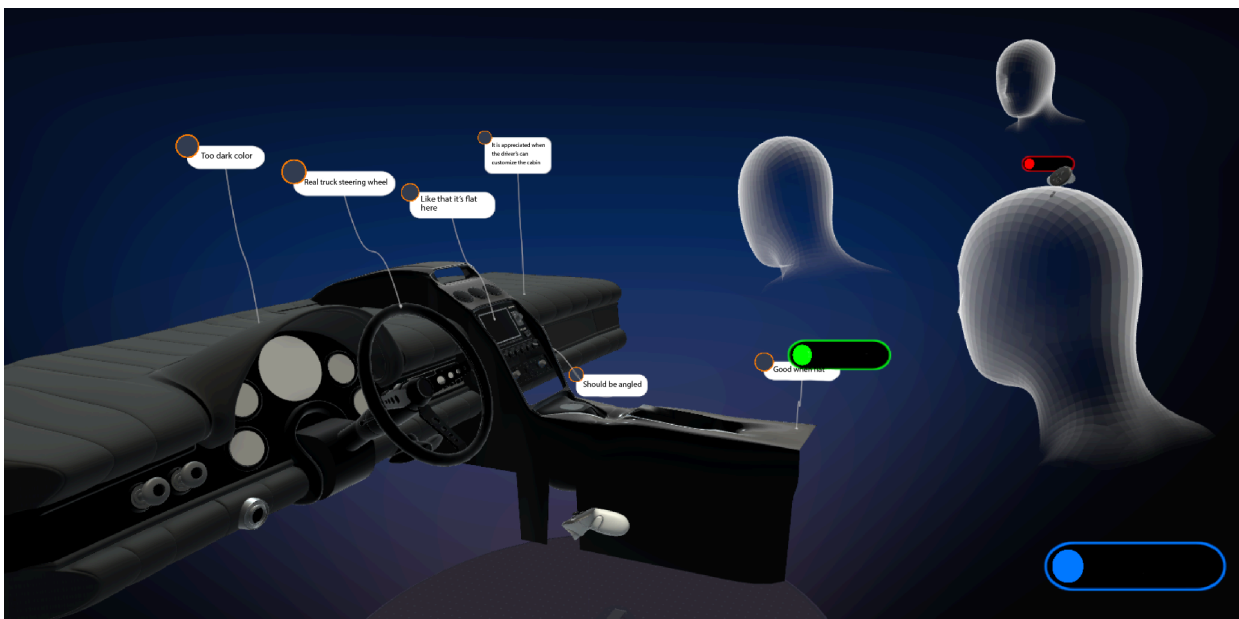


Figure 21. Feedback gathered from the workshop.

During the workshop, a challenge of recognition was revealed. Since the facilitator did not personally know the participants, it was sometimes difficult to recognize their voices. In the XR environment, the avatars did not indicate who was speaking, which made it harder to direct follow-up questions or engage participants who had been quiet for a while. This reliance on voice recognition, without visual cues, limited the facilitator’s ability to manage the conversation dynamically.

Another challenge was that one participant struggled to use the joystick to write and publish comments on the instrument panel. This revealed that it can be frustrating for the technician to assist in problems when they cannot see what the user sees in XR, making guidance difficult. Nevertheless, having a technician present was still valuable for resolving technical issues like this.

In the reflective session after the workshop, the participants mentioned that they could have continued the session for longer and probably generated more valuable ideas. Another point that emerged during the reflective session concerned the comment function. Both strengths and weaknesses regarding the tool were identified. It was frustrating when the system did not work as intended and for participants unfamiliar with it, the process became a barrier. However, the feature also proved valuable as, at one point, a participant noticed a comment made by the other and later revisited it, expressing a differing opinion.

It was also noted that integrating a real truck steering wheel with the instrument panel, or conducting the test in a virtual truck cabin, could have enhanced the sense of realism and improved the overall user experience.

7.2.3. Shared Observations from Both XR Workshops

In both tests, various valuable aspects and opinions were raised, providing key insights regarding the instrument panel and user interaction. All participants were highly impressed by the technology, particularly the ability to see where others were pointing and the physical room in combination with XR-elements. They found it exciting to point at objects and see the same things in real-time, appreciating the collaborative aspect. It also became evident that using XR for the workshop provided new insights, as the participants for example realized they needed to stretch to reach certain buttons. As one participant put it: *"It made a big difference in terms of distance, depth, material, gloss, etc., which would not have been as evident from just an image"*. As such, evaluating aspects such as placement, material, shapes, color and size worked very well in XR. However, when it came to finer details, such as icons and text, XR was less effective for evaluation due to issues with resolution.

Quick iterations were also possible in the XR workshop, allowing users to easily identify their preferences by testing different options. Moreover, the participants suggested that evaluating multiple concepts against each other in this setting would have been valuable, as even simple changes like adjusting color on the same model provided significant insights.

Both the pilot test and the main evaluation revealed that none of the participants experienced motion sickness, as there were no indications of nausea or fatigue. In the reflective session, they shared that while they typically feel dizzy with VR, they did not experience this issue with MR.

Regarding the workshop setup, it was unclear whether the context scenes added value, possibly due to the quality of the images used, which ended up looking rather flat. Additionally, it was challenging for the facilitator to manage conversation while

simultaneously scrolling, selecting materials or writing comments. In both workshops, the comments feature experienced issues, such as appearing in the middle of the product, making participants' feedback invisible. This was primarily a problem with Campfire and explaining the issue proved difficult in the "remote" XR environment. Some positive aspects regarding setup included ease of recording the test and switching between different scenes.

During the interaction, it was somewhat unclear for participants to touch a screen with no visible or haptic response. When changing the radio channel, both participants explained how they would proceed with the low-tech and high-tech versions, sharing their preferences and what they appreciated about each approach, an outcome that was anticipated.

7.3. Findings from Tests Using Emotion Interpretation Technology

Three separate tests were conducted using emotion interpretation technology. The first two involved presenting various types of video clips to evaluate the reliability and effectiveness of the method. The third test was carried out in a driving simulator while observing interactions. All three tests aimed to explore the types of insights that could be obtained through approaches with the technology. Findings from all tests are presented below.

7.3.1. Findings from Reliability Test

In the first test, which focused on assessing the reliability of Noldus FaceReader, the software generally demonstrated a good ability to detect the primary emotion in each video clip. The test included four participants, but due to technical issues where the camera in Noldus stopped functioning for one participant, usable data was only collected from three participants and the fourth participant was excluded from the analysis. For the remaining participants, Noldus successfully identified emotional responses during each segment in agreement with the participants own self-assessment. Among the emotions reported in the self-assessments that were also detected by Noldus, the intensity in Noldus varied from dominant to barely detectable.

However, the software often registered emotions that were not obvious through visual observation. The additional emotions varied among participants. In one case, it was happy, in another neutral and lastly scared for the third participant. During the self-assessments, participants did often not report these secondary emotions, highlighting potential limitations in the accuracy of AI-based emotion interpretation. Furthermore, discussions with the participants revealed that their emotional responses were often more complex than what Noldus was able to capture. For instance, during the clip showing food decay, Noldus registered emotions such as neutral, disgusted or happy. However, the participants described a broader range of feelings including fascination, boredom, fear, satisfaction and desire. Summaries of the emotion data from Noldus for each participant in the reliability test can be found below in Figure 22.

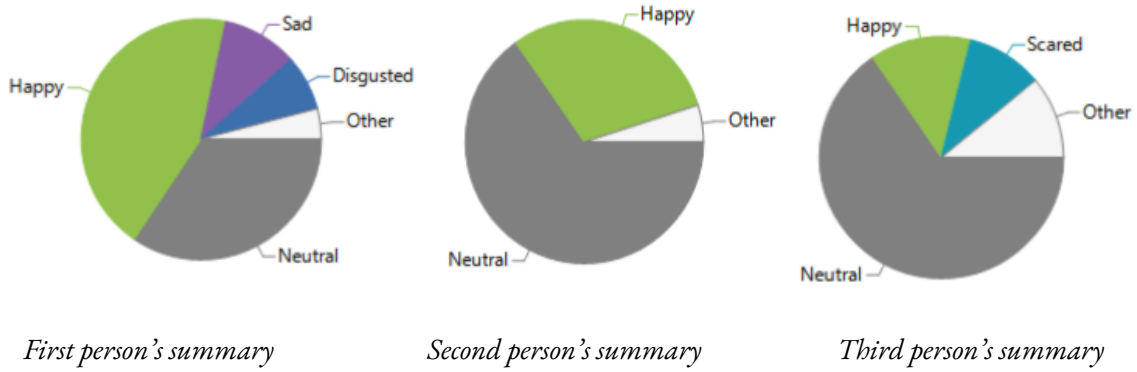


Figure 22. Summary of Noldus emotion interpretation from reliability test.

7.3.2. Findings from Context-Based Test

In the second test, which evaluated its performance in an automotive setting, Noldus primarily detected a neutral emotional state in both participants. Similar to the first test, a secondary emotion also appeared, reinforcing the earlier observation that Noldus often identifies additional emotions beyond the dominant one, even in an automotive context. See the emotion summaries from Noldus below in Figure 23. For the first participant, the secondary emotion detected was surprised, which the participant reported experiencing briefly during the second clip. For the second participant, the software identified sadness, although the participant did not report feeling this emotion at all.

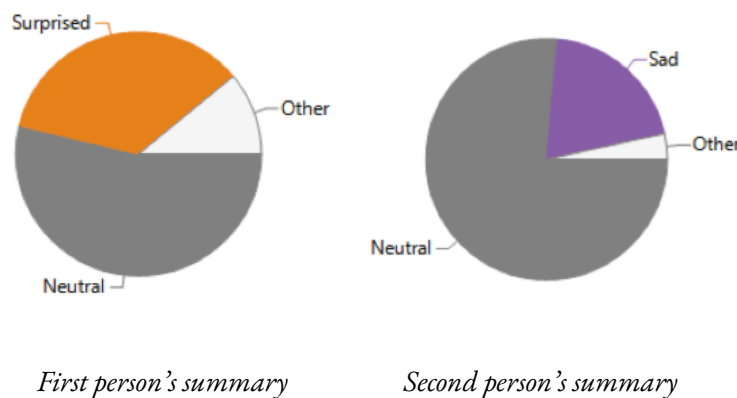


Figure 23. Participants summary in context-based test.

The test also highlighted that Noldus FaceReader can be highly sensitive to head positioning. For the first participant, certain segments of the video could not be analyzed because their face was angled slightly upward or to the side in relation to the recording camera. Although the angle changes were minimal, they were enough to prevent the software from capturing facial data. This sensitivity could pose challenges in real-world applications.

Overall, the results of this test suggest that while the reliability of AI-based emotion interpretation is acceptable in controlled settings, its usefulness in automotive contexts may be limited.

7.3.3. Findings from Interaction Test

An interaction test in a driving simulator was conducted to assess whether being situated in a simulated vehicle environment would evoke stronger emotional responses than simply watching truck-driving videos.

The video recording using the Smart Eye camera worked well with the Noldus FaceReader software and the analysis proceeded without any major interruptions, unlike in previous tests. Only a few brief disruptions occurred when a participant leaned forward significantly or covered their face. The recording worked well in cases when a participant was wearing glasses or a hat, as well as across different ethnic backgrounds.

The built-in emotion interpretation in the Smart Eye system predominantly registered the emotion *neutral*, with only occasional exceptions such as *joy* or *surprise*. The main emotion identified by Noldus FaceReader during the driving simulator test was *neutral*. This was also the most commonly aligned emotion between participants' self-assessments and the software's analysis. Occasionally, there was also agreement between the analysis from Noldus and the self-assessments on emotions such as *scared*, *surprised*, and *happy*. However, discrepancies were just as frequent for these emotions, highlighting inconsistencies between the software and the self-assessments.

It appears that Noldus assigned one or two dominant baseline emotions in addition to neutral to each participant, which remained fairly consistent throughout the evaluation. These emotions were seldom identified by participants in their self-assessments. As illustrated below in Figure 24, the dominant emotions vary between individuals (task order is: *change radio station digital*, *connect bluetooth digital*, *change temperature physical*):

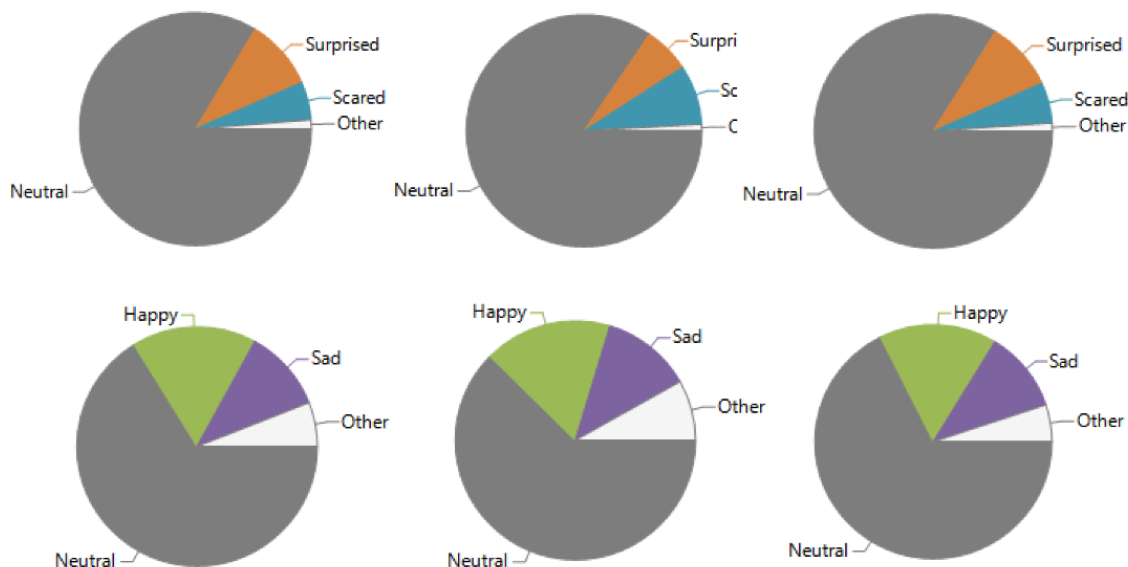


Figure 24. Selected summaries from three tasks for participant one and two.

All participants at some points reported experiencing emotions found, or added, in the Geneva Emotion Wheel that were not available in the FaceReader software. Nonetheless, the majority of self-reported emotions were available within the set of emotions detectable by the software.

In several cases, when a participant spoke during a task, Noldus registered enhancements in certain emotions. These emotional spikes varied between individuals where for example, one participant showed increased levels of fear or surprise during speech. Additionally, head movements such as turning the head, sometimes triggered shifts in the detected emotions, even when the participant's facial expression seemingly remained unchanged. Moreover, participants who laughed as a result of stress, fear and nervousness - as stated by participants - were by Noldus registered as happy. In such instances, it became clear that it was not possible to access underlying causes behind the emotional expression through the software alone. Such valuable insights emerged during qualitative follow-up interviews, which provided a deeper understanding of the internal emotional processes behind the participants' reactions.

Regarding being aware of the tracking camera, the majority of participants reported that they did not actively think about it during the recording. One participant even mentioned forgetting about it completely. On the contrary, another participant said that they occasionally considered the filming, particularly when speaking. When asked about their feelings toward the use of AI technology to analyze the recording, participants expressed mixed but generally positive reactions. One participant found the idea "cool" and "exciting". Another participant also felt positively about the technology, but also acknowledged potential concerns about privacy. There were also expressed uncertainty about AI's ability to detect subtle emotional responses, believing it would more likely identify extreme reactions but might struggle with more nuanced emotional states and the reasoning behind the expressed emotions.

For practical concerns of the evaluation itself, all participants appreciated the opportunity to conduct self-assessment and further elaborate on their emotions. However, several difficulties were noted, including the difficulty to fill out the Geneva Emotion Wheel without being able to reference previous answers for comparison. Additionally, some emotions, such as "uncomfortable" and "annoyed," were said to be missing and certain scales (e.g., "interested" and "happy") were seen as too similar. Lastly, one participant emphasized the need for more practice in the driving simulator before beginning the tests, as both the simulator and tasks were new to them.

8

Synthesized Results

Based on the results from phase one and phase two, each explored technology demonstrates distinct strengths, offering a wide range of potential applications. The technologies also present specific challenges and limitations that must be addressed to enable broader or novel use.

XR technology is rapidly evolving, with improvements such as better wireless headsets and tactile gloves being some of the most prominent improvements. However, challenges for use in practice as well as in a corporate setting remain. In testing, an XR workshop together with end-users proved valuable, where the participants engaged in and enjoyed real-time collaboration and interaction. It proved especially valuable for assessing layouts, placements and materials. While XR technology showed potential for iterative design testing in practice, limitations highlighted in phase one, such as resolution issues and the lack of haptic feedback, surfaced and made it harder to evaluate interaction with finer details.

Emotion interpretation technology analyzes facial expressions to uncover subconscious reactions and interpret emotions. Testing of the technology revealed that it often assigned neutral emotions to participants but missed other emotions that were mentioned in self-assessment. In some cases it also identified secondary emotions that participants did not report. While the technology has the potential to provide valuable insights and prove useful in combination with other methods, it faces challenges in accuracy due to individual differences in emotional expression and difficulties of capturing the complexity of emotions in real-world settings.

Regarding other technologies, many potential use cases for user evaluation were identified; however, these were not tested in practice during this study. EEG sensor technology is advancing but remains complex, with challenges such as the need for high-quality signal acquisition and lengthy setup times. Wireless devices offer convenience and the potential to work in specific industry contexts, but may lack accuracy. AI is gaining immense traction across various fields. For user evaluations in particular, AI has the potential to revolutionize efficiency by supporting data analysis, generating insights and guiding decision-making. However, concerns about data transparency, ethical implications and over-reliance on AI persist. Eye tracking technology, while becoming more affordable and evolving, has the ability to reveal where users focus but not precisely why. To gain a deeper understanding of underlying causes for eye-gaze, eye tracking should be combined with other data sources, such as emotion interpretation or interviews.

There are several needs for user evaluation in the automotive context. They include in-depth understanding of users, insight into driver interaction with other road users, high realism in lab simulations and efficient data collection and analysis. While the explored technologies present opportunities to meet these needs and improve user evaluations, their limitations must be addressed for practical use in industry settings.

9

Discussion

The following chapter presents a comprehensive discussion of the study. It reflects on the overall implications by addressing the study's aim and research questions, evaluating the results, examining the process and execution, and developing recommendations.

9.1. Aim & Research Question

The aim of the thesis was to investigate how emerging technologies can enhance the evaluation of human-machine interactions in the design- and product development process. It also sought to identify the most promising technologies for future user evaluation and understand the types of insights these technologies can offer Volvo Group. This aim has successfully been fulfilled and the three research questions have been addressed through a two-phase approach. The first phase focused on exploring relevant technologies and gathering information, while the second phase involved practical testing of various types of user evaluation.

Research Question 1: *How can emerging technology be used in future user evaluation methods?* This question was addressed through a technology scouting process that included extensive literature research and expert interviews. The focus was on identifying relevant technologies and exploring how they could be applied within user evaluation contexts. Several potential applications and use cases were identified and two particularly promising evaluation methods incorporating emerging technology were developed in greater detail.

A limitation of the part of the study answering Research Question 1, is the restricted amount of emerging technologies that were explored. Some were deemed less relevant for short-term user evaluations focused on HMI, while others were excluded due to limited feasibility for testing within the scope of this project.

Research Question 2: *What types of insights and results can be obtained from user evaluation methods that incorporate emerging technologies?* This question was primarily addressed through the testing in practice of two types of user evaluations developed during the study, alongside continued technology scouting. The activities offered insights into the potential outcomes if Volvo Group were to adopt similar methods in future user evaluation.

However, the tests remained exploratory, comparative and focused on a broad overview rather than detailed analysis. They were conducted on a small scale with few iterations, which constitutes a limitation of the study.

Research Question 3: *How can Volvo Group leverage these opportunities?* This question was addressed by developing recommendations on how emerging technologies could be integrated into Volvo Group's user evaluation processes. The study examined current evaluation practices within the automotive context and at Volvo Group to identify needs and gaps, and explored how emerging technologies could add value by addressing these areas.

It is, however, difficult to state with certainty how the technologies should be used and managed. Since this study remained exploratory, the recommendations are not highly specific for each technology, as the study focused on a broad range of technologies, rather than in-depth analysis of each.

9.2. Results

The aim of the study's practical tests was to assess whether XR and emotion interpretation technology could capture more nuanced and deeper insights into user experiences. While it is challenging to definitively evaluate the success of this goal, the results suggest that users felt highly engaged and involved in the XR experience. Participants for example adjusted their answers after experiencing changes in XR, indicating that they might have stuck with their initial response or perceived the change less clearly without the XR context. This implies a potential increase in response engagement and evaluation participation, though it could also lead to overly positive feedback, introducing a bias. Without repeated or a comparative test, this cannot be fully confirmed. Furthermore, it became clear during the workshop that both the designers (pilot test) and truck drivers (main workshop) contributed valuable insights, despite sometimes approaching the evaluation differently. This suggests that the method is effective across participants with different expertise, personalities or goals - an important quality given the study's aim of addressing multiple user groups.

Regarding the evaluations involving emotion interpretation, the technology provided additional insights to the self-assessment, but the insights were not deemed reliable enough to draw definitive conclusions. The emotion interpretation software identified neutral as the main emotion across all participants and all tests, which also turned out to be the most commonly aligned emotion between the software's analysis and the participants' self-assessments. This alignment may indicate the softwares' ability to successfully reflect the participants' own perceptions. However, it could also suggest that the emotion interpretation software has limited value within the context of this study. As presented in the results, a secondary emotion was often detected by the emotion interpretation software, which did not coincide with participants' self-assessment. One possible explanation for this could be that individual differences in baseline facial expressions, commonly referred to as "resting" faces, may influence how the Noldus software interprets subtle cues. For instance, one participant appeared calm and neutral while watching the video, yet the software interpreted their expression as fear. It is, however, noteworthy that users' self-assessments of emotions might not always be entirely accurate due to exaggeration or memory issues. As noted in the results, participants frequently reported more nuanced emotions in their self-assessments than the software was able to capture, underscoring the limitations of AI-based emotion

interpretation in fully representing the complexity of human emotion. Additionally, the context-based test revealed that the software was particularly sensitive to changes in facial angle when using an integrated or external web camera, which could pose challenges in real-world applications. This sensitivity was significantly reduced in the interaction test, underscoring the importance of using cameras aimed at tracking facial expressions such as the eye tracking camera in this test.

9.2.1. Reliability of Results

The reliability of the study's results can be regarded as relatively high, as the findings are intended to serve as an initial indication within the area explored. This is further supported by the inclusion of diverse stakeholder perspectives through expert meetings and successful testing of methods despite limited resources, suggesting that implementation would be even more feasible and effective with increased resources. However, the degree of reliability varies across different technology areas due to the variation in number of expert meetings conducted. For XR technology, multiple meetings were conducted which provided a strong foundation and enhanced credibility of results and recommendations. Although only two meetings were conducted regarding emotion interpretation, more extensive practical testing was carried out to support the validity of the conclusions made. Regarding AI and eye tracking, there was only a single meeting conducted for each area, resulting in less credible results and somewhat more vague recommendations.

Given the scale and complexity of Volvo Group (and Volvo GTT), a comprehensive mapping of their current practices regarding the different technologies and user evaluation practices were not possible in this study. It is therefore likely that certain practices, tools and methods exist that were not identified or considered. As such, some of the study's recommendations might not be fully applicable or relevant to take into consideration for implementation.

9.2.2. Feasibility of Implementing the Results

For the integration of emerging technologies in user evaluations to be feasible, there must be sufficient expertise and necessary resources for implementation. Additionally, the field of emerging technology demands a forward-looking vision as well as the ability to assess how close certain technologies are to maturity, how likely they are to deliver on their promises etc. Another key factor influencing whether emerging technologies realistically can be integrated into user evaluation processes is how the user perceives and accepts the new technology. Their comfort plays a crucial role and it might take some time for users to adapt to unfamiliar methods or tools. To Volvo Group's advantage, the company has internal experts in emerging technologies, technology research, user evaluations and various specialized areas. This internal expertise would likely increase the possibility in implementing the study's findings.

9.2.3. Future Considerations of Emerging Technologies in User Evaluation

One question that remains after this study is whether emerging technologies simply generate more user data without necessarily providing more value. While the answer will vary between technologies and execution, it is important to critically consider whether collecting additional data contributes meaningful insights or merely increases the analytical burden, especially when time and resources for analysis are limited. However, a common challenge in large automotive companies is the inability to test with as many users as desired. As a result, a smaller group of participants can be expected to represent an entire global target audience. In such cases, maximizing the amount, depth and quality of data collected from each participant may be crucial for capturing diverse user needs.

Looking ahead, it is therefore important to consider that future advancements in user evaluations may not center around the new ways of measuring but rather the methods used to analyze data in more efficient and transformative ways. This also includes ways of gaining deeper understanding of the underlying causes or implications revealed by the data.

Moreover, there remains a risk in user evaluation of assuming that subjective experiences can be captured by purely objective measurements. Conversely, users may not always fully convey their actual experiences, whether due to memory lapses, a tendency to exaggerate, lack of awareness or other factors. As such, relying solely on techniques gathering either objective data or self-reported input in future user evaluation could lead to an incomplete understanding of the user experience.

9.3. Process & Execution

This study adopted an exploratory approach, focusing on investigating potential emerging technologies in user evaluation and the insights they provide, rather than aiming to prove a specific hypothesis. The study's research questions were formulated accordingly and explored what insights can be gained from the technology itself, rather than focusing solely on how to extract specific information. This flexible methodology allowed for a broad understanding of the technologies' capabilities, limitations and opportunities which might not have been as apparent in a more constrained study.

The chosen methods, technology scouting and practical evaluation testing, proved well-suited for the study's exploratory approach and provided insights collected in several ways. The iterative, collaborative approach between expert meetings, technology scouting and testing in practice was highly effective, continuously yielding insights during all parts of the study. The conducted expert meetings were not always set out to be a formal interview, which made the meeting setup and what insights would be gathered somewhat unclear beforehand. As a result, the interviews were flexible and adapted to the context of each meeting, yet based on a set interview format which allowed for easy comparison between different technology areas. Many of the expert meetings additionally included facility tours and hands-on testing, which proved valuable for providing inspiration.

A potential limitation of the study is the selection of technologies tested in practice, which was based on brainstorming and discussions within the project team. Had different technologies been chosen, the outcomes and recommendations could have varied. An example is EEG technology, which was initially planned to be tested in practice, but was not included due to challenges in acquiring the necessary equipment and expertise. Testing this technology might have provided additional insights in this area, potentially shifting the focus from other technologies.

Additionally, the participant selection and limited sample size in this study inevitably affected the reliability of the results. The tests predominantly involved a small group of participants, most of whom were young and had limited experience as professional truck drivers. Several of the participants were Industrial Design Engineering students with prior user testing experience. This background may have made them more aware of the test coordinator's observations, influencing their behavior. At the same time, this familiarity likely made them more comfortable reflecting on and articulating their emotional responses, which could have led to more nuanced feedback compared to less experienced participants. However, the absence of truck drivers and internal Volvo designers in the testing phase represents a limitation and may have reduced the overall value of the results. The decision was made to prioritize involving truck drivers and design students in the XR workshop, while using non-truck drivers in the simulator-based interaction test. This was due to the focus of that test being on evaluating the functionality of the technology itself in relation to self-assessment, rather than assessing user experience specific to truck driving.

9.3.1. Sources of Variability in Emotion Interpretation Tests

Regarding the user evaluation utilizing emotion interpretation there are several possible sources of variability. One relates to participants' awareness of being recorded. Whether or not to inform the participants of this aspect were discussed, as this awareness could potentially influence their behavior. Ultimately, it was decided that participants should be informed in advance to avoid discomfort or a sense of having been misled. Nonetheless, knowing they were recorded may have led participants to alter their facial expressions, either consciously or unconsciously, and as such the results from the AI-based software. Another limitation is that the camera setup remained unchanged between participants, potentially affecting the interpretation of some emotional responses due to differences in participants' heights and thus camera angles.

Additionally, the use of the PrEmo instrument and Geneva Emotion Wheel for self-assessment might have introduced some limitations. Participants may have felt constrained by the predefined set of emotions, even though they were encouraged to mention any feelings they found missing. Furthermore, the descriptions provided for each emotion in PrEmo may have influenced how participants interpreted and selected them during the self-assessment. If the test were to be repeated, an improvement would be to ask participants to provide detailed explanations for the emotions they reported in the self-assessment as well as reflect on the emotions interpreted by the software.

The emotional responses observed in the automotive context were predominantly neutral. This could suggest that the emotional reactions evoked in such scenarios may not be sufficiently expressive for the technology to offer substantial value in user evaluations. However, the neutral response could also be due to camera setup, the test lacking in realism or the choice of software.

9.3.2. Sources of Variability in XR Workshop

Regarding the user evaluation using XR, the initial plan was to begin with a physical truck cabin model to create a realistic setting, followed by an interaction with a Volvo instrument panel, replacing the generic car panel used in the final setup. However, due to challenges accessing the necessary 3D-files and addressing confidentiality concerns, this approach was revised and resulted in an external 3D-model of a car being used. Additionally, the original plan was to include an interactive element in the XR-environment but technical difficulties and time constraints prevented this from being implemented as intended. The workshop indicated that a realistic touch screen interface in XR could have enhanced the interaction and provided a deeper sense of engagement and more detailed user insights. Despite this, valuable insights were still gathered from the test.

Moreover, the use of a generic car instrument panel and its resolution could have influenced the recommendation of primarily focusing on evaluating less-detailed models in an XR-environment. With a more accurate 3D-model and functioning interaction, the outcome might have been different.

9.5. Recommendations Regarding Future User Evaluations

Based on the results of the study, the third research question: *How can Volvo Group leverage these opportunities?* could be addressed through the recommendations presented in this section.

Regarding user evaluation in the automotive industry, there were several identified needs and opportunities that offer the potential of deeper insights into user behavior. It is therefore recommended that Volvo Group explore ways of further enhancing current practices by considering the following:

- Perform user testing early in the process
- Perform user testing on a large number of users
- Ensure reliable and trustworthy results
- Perform user testing in a life-like environment
- Make user testing accessible for designers and developers
- Make user testing seamless (time, budget, planning...)
- Enable remote user testing

Furthermore, it is recommended to realize the opportunities offered by emerging technologies in this context, while ensuring their integration is guided by clear goals rather than the appeal of novelty. New methods for user evaluation should ideally complement traditional approaches or enhance existing processes to add value, rather than replace them outright.

For the specific emerging technologies investigated in this study, it is recommended for Volvo Group to actively monitor and explore ongoing developments in relation to user evaluations, to ensure early and effective utilization. Examples of aspects to consider are technological advancements (e.g. XR haptic feedback), accessibility within the automotive context (e.g. broader adoption of emotion interpretation) and opportunities for combining technologies (e.g. EEG and XR). Recommendations for each technology area are outlined below.

Extended reality

- Build on the internal work and value-adding that is currently in place regarding XR.
- Involve end-users more actively in XR-based product evaluations.
- Conduct early-stage product exploration in XR together with end-users for assessing overall shapes, colors, and component placement. Highly detailed models may not be necessary to obtain meaningful user feedback.
- Aim for high fidelity between XR models and the evaluated product to enhance immersion and ensure relevant user insights.
- Keep technical aspects in mind when conducting user evaluations utilizing XR technology (e.g. familiarity with tools, software selection, level of immersion and professionalism of visuals).
- Develop an internal platform supporting collaboration and design evaluation with end-users. Alternatively, use standardized, non-sensitive models to address confidentiality concerns in external programs.

Emotion interpretation

- Do not rely solely on emotion interpretation software for user emotion assessment.
- Conduct further testing of emotion interpretation technology with a more detailed set-up and different emotion interpretation softwares
- Explore the possibilities of using the results from emotion interpretation as a basis for discussion in post-test interviews.

EEG sensors

- Investigate the potential of consumer-grade EEG devices and data interpretation tools (software).
- Be aware of the limitations of EEG technology in user evaluation as it requires careful and precise test execution. It is furthermore, important to not overinterpret the abilities of EEG technology as its signals and their implications are inherently complex and sensitive to a range of factors.

Eye tracking

- Explore the opportunity to combine eye tracking technology with other tools or emerging technologies for user evaluation.

Artificial intelligence

- Explore the opportunities of integrating AI into user evaluation for data analysis and enhanced efficiency.
- Develop a clear understanding of the AI tools' intended purpose, limitations and the data used to train the algorithms before implementing them in user evaluations on a large-scale.

10

Conclusion

The purpose of this thesis was to explore how emerging technologies can support future user evaluation in an industrial automotive context and how they might be applied within Volvo Group. To achieve this, the study consisted of a two-phase method, combining technology scouting through literature research and expert meetings with practical testing.

Several technologies were identified as relevant for user evaluation, including XR, emotion interpretation, EEG sensors, eye tracking and AI. Key findings indicate that such technologies can uncover deep insights into user behavior. However, new methods for user evaluation should ideally complement traditional approaches or improve existing ones to add value, rather than replace them outright. Adoption decisions must furthermore be driven by clear evaluation objectives rather than by novelty alone.

This thesis provides several recommendations for Volvo Group on how to approach each of the explored technologies in relation to user evaluation. A key overarching recommendation is to actively monitor developments related to user evaluation in each technology area to enable timely and effective implementation. It was also shown that XR is well suited for early user evaluation of spatial and visual aspects and that emotion interpretation technology may offer complementary insights but requires further validation. Furthermore, EEG sensors can provide an expanded understanding of users' mental states through various types of devices, but caution should be taken to avoid overinterpreting their capabilities. Advancements in eye-tracking technology also contribute to its continued relevance, particularly when combined with other tools or emerging technologies in user evaluation. Additionally, AI has the potential to significantly support data analysis within user evaluation when applied transparently and with critical consideration.

In conclusion, the identified technologies offer valuable opportunities to address user evaluation needs within the automotive industry, but they also come with considerable limitations. While this study provides early insights into the potential of the identified technologies, further research and testing are necessary. The promise of more immersive, comprehensive and dynamic user insights is significant, provided that implementation is guided by critical reflection and a clear focus on practical application.

References

- Agency for Healthcare Research and Quality. (n.d.). *NASA Task Load Index (NASA-TLX)*. U.S. Department of Health and Human Services. Retrieved May 9, 2025, from <https://digital.ahrq.gov/health-it-tools-and-resources/evaluation-resources/workflow-assessment-health-it-toolkit/all-workflow-tools/nasa-task-load-index>
- Anyverse. (2025, February 20). *Eye-tracking and AI: The role of high-quality data in automotive safety*. Retrieved May 9, 2025, from <https://anyverse.ai/eye-tracking-and-ai-the-role-of-high-quality-data-in-automotive-safety/>
- Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.
- Bellamy III, W. (2020, December 18). Can real-time pilot monitoring help reduce unexplained physiological events? *Avionics International*. Retrieved May 9, 2025, from <https://www.aviationtoday.com/2020/12/18/can-real-time-pilot-monitoring-help-reduce-military-pilot-unexplained-physiological-events/>
- Berlin, C., & Babapour Chafi, M. (2024, December 1). *Virtuellt kontorsarbete (VirKA) – Fysiska, kognitiva och sociala effekter av kunskapsarbete i VR-miljöer*. Chalmers Tekniska Högskola & Institutet för stressmedicin (ISM). https://research.chalmers.se/publication/545495/file/545495_Fulltext.pdf
- Bertini, E., Catarci, T., Dix, A., Gabrielli, S., Kimani, S., & Santucci, G. (2008). Appropriating heuristic evaluation methods for mobile computing. In J. Lumsden (Ed.), *Handbook of research on user interface design and evaluation for mobile technology* (pp. 1107-1123). IGI Global. <https://doi.org/10.4018/978-1-59904-871-0.ch046>
- Brave, S., & Nass, C. (2008). Emotion in human-computer interaction. In A. Sears & J. A. Jacko (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (2nd ed., pp. 77-92). Taylor & Francis Group. <https://doi.org/10.1201/9781410615862>
- Brey, P. (2017). Ethics of emerging technologies. In S. O. Hansson (Ed.), *Ethics of emerging technologies* (pp. 69-86). Rowman & Littlefield International. https://books.google.se/books?hl=sv&lr=&id=YeLaDwAAQBAJ&oi=fnd&pg=PA175&dq=emerging+technology&ots=M_MGBo3I1z&sig=pv4WeSWpIBOupidgdeiwAg-9AD4&redir_esc=y#v=onepage&q=emerging%20technology&f=false

Burmaoglu, S., Trajkovik, V., Loncar Tutukalo, T., Yalcin, H., & Caulfield, B. (2018). Chapter 14 - Evolution map of wearable technology patents for the healthcare field. In A. K. Y. Tong (Ed.), *Wearable technology in medicine and health care* (pp. 303-319). Academic Press. <https://doi.org/10.1016/B978-0-12-811810-8.00014-2>

Bødker, S., & Buur, J. (2002). The design collaboratorium: A place for usability design. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 9(2), 152–169. <https://doi.org/10.1145/514216.514217>

Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A. J. (2014). The use of triangulation in qualitative research. *Oncology Nursing Forum*, 41(5), 545–547. <https://doi.org/10.1188/14.ONF.545-547>

Center for Law, Brain & Behavior. (2020, February 25). *Do not rely on facial expressions for how people are feeling*. <https://clbb.mgh.harvard.edu/do-not-rely-on-facial-expressions-for-how-people-are-feeling/>

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., Medaglia, R., & Williams, M. D. (2021, April). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57, 101994. <https://doi.org/10.1016/j.ijinfomgt.2020.101994>

Eason, K. D. (1995). User-centred design: For users or by users? *Ergonomics*, 38(8), 1667–1673. <https://doi.org/10.1080/00140139508925120>

Ehrlich, K., & Rohn, J. A. (1994). Cost justification of usability engineering: A vendor's perspective. In R. G. Bias & D. J. Mayhew (Eds.), *Cost-justifying usability* (pp. 73–110). Academic Press.

European Commission. (2022, July 6). *New rules to improve road safety and enable fully driverless vehicles in the EU*. Retrieved May 9, 2025, from https://ec.europa.eu/commission/presscorner/detail/en/ip_22_4312

European Innovation Council. (2024, December 18). 2024 EIC tech report identifies watch list of 34 emerging technologies that could change the future of Europe. *European Innovation Council*. Retrieved May 9, 2025, from https://eic.ec.europa.eu/news/2024-eic-tech-report-identifies-watch-list-34-emerging-technologies-could-change-future-european-2024-12-18_en

Experience Haus. (2024, August 21). *Revolutionising user experiences: The future of UX design and research*. Retrieved April 4, 2025, from <https://experiencehaus.com/blog/revolutionising-user-experiences-the-future-of-ux-design-and-research/>

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., *et al.* (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4). <https://doi.org/10.1007/s11023-018-9482-5>

Gartner. (n.d.). *Gartner Hype Cycle*. Retrieved Feb 25, 2025, from <https://www.gartner.com/en/research/methodologies/gartner-hype-cycle>

Gaskill, M. L. (2024, November 13). Wearable tech for space station research. *NASA*. Retrieved April 4, 2025, from <https://www.nasa.gov/humans-in-space/wearable-tech-for-space-station-research/>

Glaessgen, E., & Stargel, D. (2012). *The digital twin paradigm for future NASA and U.S. Air Force vehicles*. AIAA. <https://ntrs.nasa.gov/api/citations/20120008178/downloads/20120008178.pdf>

Gong, Z. (2024, December 11). The history of automotive human-machine interaction: Seven directions of evolution. In D. Harris, W. C. Li, & H. Krömker (Eds.), *HCI International 2024 – Late breaking papers* (Vol. 15381, pp. 47–56). Springer. https://doi.org/10.1007/978-3-031-76824-8_4

Greenwald, W. (2025, April 14). *The best VR headsets for 2025*. PCMag. Retrieved April 4, 2025, from <https://uk.pcmag.com/virtual-reality/75926/the-best-vr-headsets>

Güiza Caicedo, D., & van Beuzekom, M. (2006). *"How do you feel?" An assessment of existing tools for the measurement of emotions and their application in consumer products research* (Report). Delft University of Technology, Department of Industrial Design. <https://www.bluehair.co/corner/wp-content/uploads/2008/10/final-report-assessment-of-existing-tools-for-the-measurement-of-emotions.pdf>

Halaweh, M. (2013). Emerging technology: What is it? *Journal of Technology Management & Innovation*, 8(7). https://www.scielo.cl/scielo.php?pid=S0718-27242013000400010&script=sci_arttext

Henninger, S., Elbaum, S., & Rothermel, G. (2005). Domain specific user centered design for HCMDSS. *Position paper for HCMDSS 2005*. HCMDSS.

IBM. (2021, August 5). *What is a digital twin?* Retrieved May 9, 2025, from <https://www.ibm.com/think/topics/what-is-a-digital-twin>

Interaction Design Foundation (IxDF). (2016, June 6). *What is usability evaluation?* Retrieved April 4, 2025, from <https://www.interaction-design.org/literature/topics/usability-evaluation>

Interaction Design Foundation (IxDF). (2022, January 24). *What is extended reality (XR)?* Retrieved April 4, 2025, from <https://www.interaction-design.org/literature/topics/extended-reality-xr>

Johns Hopkins Medicine. (n.d.). *Electroencephalogram (EEG)*. Retrieved April 4, 2025, from <https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/electroencephalogram-eeeg>

Khare, S. K., Blanes-Vidal, V., Nadimi, E. S., & Acharya, U. R. (2024, February). Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations. *Information Fusion*, 102, 102019. <https://doi.org/10.1016/j.inffus.2023.102019>

Khasanah, I., & Gunawan, R. (2023). Method use enhanced cognitive walkthrough (ECW) in testing the interface design of student interface rating system assessment program. *International Journal of Progressive Sciences and Technologies (IJPSAT)*, 40(1), 263–268.

KMS Healthcare. (2024, June 19). *Wearable tech in healthcare: Revolutionizing patient care*. Retrieved May 9, 2025, from <https://kms-healthcare.com/blog/wearable-technology-in-healthcare/>

Koniakowsky, I., Forster, Y., Wiedemann, K., Naujoks, F., Krems, J. F., & Keinath, A. (2025, February). The effectiveness of driver monitoring systems in mitigating visual distraction depends on secondary task complexity and experience – A driving simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour*, 109, 1-10. <https://doi.org/10.1016/j.trf.2024.12.008>

Larsen, L. B., Øvad, T., Nielsen, L., & Larusdottir, M. (2021). Remote user testing. In C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, & K. Inkpen (Eds.), *Human-computer interaction – INTERACT 2021* (Vol. 12936, pp. 579–583). Springer. https://doi.org/10.1007/978-3-030-85607-6_81

Laurans, G., & Desmet, P. M. A. (2017). Developing 14 animated characters for non-verbal self-report of categorical emotions. *Journal of Design Research*, 15(3/4), 214–233. <https://doi.org/10.1504/JDR.2017.089903>

Lawrence Livermore National Laboratory. (2021, November 5). *The birth of artificial intelligence (AI) research*. Retrieved May 9, 2025, from <https://st.llnl.gov/news/look-back/birth-artificial-intelligence-ai-research>

Leggett, D. (2010, January 19). A brief history of eye-tracking. *UX Booth*. Retrieved April 4, 2025, from <https://uxbooth.com/articles/a-brief-history-of-eye-tracking/>

Lyrheden, F. (2022, June 17). *Fit to drive? 5 key insights into the state of intoxicated driving research*. Smart Eye. Retrieved April 4, 2025, from <https://www.smarteye.se/blog/fit-to-drive-5-key-insights-into-the-state-of-intoxicated-driving-research/>

Lyrheden, F. (2024, July 8). *DDAW and ADDW systems were just made mandatory in new cars: Here's what you should know*. Smart Eye. Retrieved April 4, 2025, from <https://www.smarteye.se/blog/what-to-know-about-ddaw-and-addw-systems/>

Marr, B. (n.d.). *The fascinating history and evolution of extended reality (XR) – Covering AR, VR and MR*. Retrieved April 19, 2025, from <https://bernardmarr.com/the-fascinating-history-and-evolution-of-extended-reality-xr-covering-ar-vr-and-mr/>

McKinsey Technology Council. (2024). *The top trends in tech*. McKinsey & Company. Retrieved April 4, 2025, from [https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-top-trends-in-tech#ech-trends-2024](https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-top-trends-in-tech#tech-trends-2024)

McKinsey & Company. (2024, August 26). *What is digital-twin technology?* Retrieved April 4, 2025, from <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-digital-twin-technology>

Merkle. (2019). *Emotion tracking*. Retrieved April 21, 2025, from https://www.merkle.com/dach/en/system/files?file=media/protected_file/2019-07/Namics-Emotion-Tracking-Whitepaper-English-version.pdf

Moran, K. (2019, December 1). *Usability (user) testing 101*. Nielsen Norman Group. <https://www.nngroup.com/articles/usability-testing-101/>

Noldus Information Technology. (n.d.). *Noldus FaceReader*. Retrieved April 21, 2025, from <https://noldus.com/facereader>

Novák, J. Š., Masner, J., Benda, P., Šimek, P., & Merunka, V. (2023, June 18). Eye tracking, usability, and user experience: A systematic review. *International Journal of Human-Computer Interaction*, 40(17). <https://doi.org/10.1080/10447318.2023.2221600>

Noyes, J. M., Starr, A. F., & Frankish, C. R. (1996). User involvement in the early stages of the development of an aircraft warning system. *Behaviour & Information Technology*, 15(2), 67–75.

- Patel, J. (2024, November 9). *The future of UX research: Emerging trends and technologies*. Cubed. <https://blog.cubed.run/the-future-of-ux-research-c412ce1fcb9c>
- Paul, S., & Rosala, M. (2021, March 14). *Wizard of Oz prototyping*. Nielsen Norman Group. <https://www.nngroup.com/articles/wizard-of-oz/>
- Program-Ace. (2024, December 11). *Top 5 virtual reality trends of 2025: The future of VR*. Retrieved April 21, 2025, from <https://program-ace.com/blog/virtual-reality-trends/>
- RISE Research Institutes of Sweden. (n.d.). *Digital twins can make communities and businesses better equipped for the future*. Retrieved April 21, 2025, from <https://www.ri.se/en/digital-twins-can-make-communities-and-businesses-better-equipped-for-the-future>
- Rotolo, D., Hicks, D., & Martin, B. R. (2015). What is an emerging technology? *Research Policy*, 44(10), 1827–1843. <https://doi.org/10.1016/j.respol.2015.06.004>
- Sabio, J., Williams, N. S., McArthur, G. M., & Badcock, N. A. (2024, March 6). A scoping review on the use of consumer-grade EEG devices for research. *PLOS ONE*, 19(3), e0291186. <https://doi.org/10.1371/journal.pone.0291186>
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695–729. <https://doi.org/10.1177/0539018405058216>
- Scherer, K. R., Shuman, V., Fontaine, J. J. R., & Soriano, C. (2013). The GRID meets the Wheel: Assessing emotional feeling via self-report. In J. J. R. Fontaine, K. R. Scherer, & C. Soriano (Eds.), *Components of emotional meaning: A sourcebook* (pp. 281–298). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199592746.003.0019>
- Sheikh, H., Prins, C., & Schrijvers, E. (2023, January). Artificial intelligence: Definition and background. In H. Sheikh, C. Prins, & E. Schrijvers (Eds.), *Mission AI: Research for policy* (pp. 15–41). Springer International Publishing. https://doi.org/10.1007/978-3-031-21448-6_2
- Shen, H., Liu, T., Qin, D., Bo, X., Wang, L., Wang, F., Yuan, Q., Wagberg, T., Hu, G., & Zhou, M. (2017). Chapter 7 - Wearable carbon nanotube devices for sensing. In H. Peng, Q. Li, & T. Chen (Eds.), *Industrial applications of carbon nanotubes* (pp. 151-168). Elsevier. <https://doi.org/10.1016/B978-0-323-41481-4.00007-1>
- Skarredghost. (2021, August 12). *HoloLens 2 review: The best in class for AR, but...* Retrieved March 11, 2025, from <https://skarredghost.com/2021/08/12/hololens-2-review/>
- Smart Eye. (2025, April 7). *5 future eye tracking trends: Multiple sensors, health monitoring, and more*. Retrieved April 4, 2025, from

<https://www.smarteye.se/blog/5-future-eye-tracking-trends-multiple-sensors-health-monitoring-and-more/>

Soufineyestani, M., Dowling, D., & Khan, A. (2020, October 23). Electroencephalography (EEG) technology applications and available devices. *Applied Sciences*, 10(21), 7453. <https://doi.org/10.3390/app10217453>

Stige, Å., Zamani, E. D., Mikalef, P., & Zhu, Y. (2023, August 29). Artificial intelligence (AI) for user experience (UX) design: A systematic literature review and future research agenda. *Information Technology & People*, 37(6). <https://doi.org/10.1108/ITP-07-2022-0519>

Urveshk2623. (2022). *Dashboard* [CAD model]. Sketchfab. <https://sketchfab.com/3d-models/dashboard-4cc3d5062d6d45598acbc3ddd0a978a1>

UserBit. (2025, March 5). *End users in UX – Meaning, example, and importance*. <https://userbit.com/content/blog/endusers-ux-terms>

Varjo Technologies. (2025). *Varjo products*. Retrieved May 20, 2025, from <https://varjo.com/products/>

Vinnova. (2024). *Strategiska teknikområden för forskning och innovation 2024*. https://www.vinnova.se/globalassets/publikationer/2024/rapport-ru-strategiska-teknikomraden_ver-01.0-final-1.pdf?cb=20241030174857

Volvo Trucks. (2025). *Electric trucks*. <https://www.volvotrucks.se/sv-se/trucks/electric.html>

Volvo Trucks. (2020, September 8). *Volvo Trucks beginnt den Verkauf der neuen Lkw in Deutschland und stellt dabei den Fahrer in den Fokus*. <https://www.volvotrucks.de/de-de/news/press-releases/2020/sep/volvo-trucks-beginnt-verkauf-der-neuen-lkw.html>

Walsh, L. (2023, July 27). *What is the metaverse? And will it help us or harm us?* University of Cambridge. <https://www.cam.ac.uk/stories/metaverse>

Wikberg Nilsson, Å., Ericson, Å., & Törlind, P. (2015). *Design: Process och metod* (1st ed.). Studentlitteratur.

Willige, A. (2022, May 24). *Digital twins: What are they and why do they matter?* World Economic Forum.

<https://www.weforum.org/stories/2022/05/digital-twin-technology-virtual-model-tech-for-good/>

YORD Studio. (2024, November 11). *Best XR headsets of 2025*. Retrieved March 11, 2025, from <https://yordstudio.com/best-xr-headsets-of-2025/>

Appendices

Appendix A - Example (XR) of Expert Interview Template

Briefly tell them about us, what we study as well as our thesis project and what we are going to investigate. We are currently exploring different emerging technologies and how they could be used in user evaluations. We are interested in learning more about how you, and others at Volvo Group, with such technologies, what obstacles there might be, what opportunities you see for the future etc.

Questions

- Could you tell us about your (and the department's) work here?
- In what way do you use XR technology? When in the process?
- What type of XR technology do you use? What are its capabilities?
- Do you work with user evaluations? If so, how? With XR?
- XR has been around for a while, but is still developing. What do you see as emerging with the technology?
- What potentials for the future do you see with the technology? New use-cases? What future possibilities with XR do you see within user evaluations ****if relevant****?
- What do you think are the biggest obstacles to a broader and more advanced use of XR (for you and in general)?
- Do you know any other department that works with XR at Volvo? Recommendations on external actors to contact?

Appendix B - Script XR Workshop

Hej! Tack för att ni ville ställa upp och hjälpa oss idag! Innan vi börjar så får du gärna skriva på 2 exemplar av den här GDPR-blanketten. Det är något som måste göras i sådana här studier med personuppgifter, vilket i detta fall är ålder, kön samt ljud och videoupptagning.

Skriver på GDPR

Presentation av oss

Toppen! Då ska vi börja med att förklara lite mer om varför ni är här. Jag heter Jonna och detta är Lovisa, och vi skriver just nu vårt examensarbete på Chalmers och Volvo Group Trucks Technology. Vårt arbete handlar om hur man kan utvärdera produkter ihop med dem som ska använda dem, så mer specifikt för Volvo Group Trucks Technology hur man kan utvärdera en lastbil med lastbilsförare. Och vi kollar på hur man kan göra det med olika typer av framtida teknik, exempelvis XR.

XR är ett samlingsord för bland annat VR, som ni kanske hört om, men även något som kallas MR och AR, vilket är liknande tekniker. Så idag ska vi helt enkelt testa hur det är att genomföra en utvärdering i XR ihop med er: lastbilschaufförer.

Presentation av dem

Ni kanske redan känner varandra lite sen innan, men vi har ju inte träffats förut så vi tänker att vi kör en snabb presentationsrunda där ni gärna får berätta vad ni heter, ålder, hur länge ni har studerat till lastbilschaufför och om ni använt VR/XR eller liknande innan.

Mer om workshopen

Då tänker vi att vi ska förklara syftet med workshopen lite mer. Största syftet är att förstå hur den bästa instrumentpanelen bör se ut utifrån lastbilsförarens perspektiv, vilket så klart är viktigt eftersom det faktiskt är ni som ska använda lastbilen i slutändan. Och vi vill som sagt testa en sådan här typ av workshop med hjälp av XR, för att se om man får fram ny information, hur det är att vara på olika platser och samarbeta osv.

Vi kommer använda oss mest av MR - mixed reality - vilket är när man är när man ser både sin vanliga omgivning och virtuella element. Och det finns inget rätt eller fel här idag eller att ni ska känna att ni behöver prestera något. Vi är intresserade av att höra var ni tycker och tänker om en kontrollpanel och tekniken!

Set-up

Det vi kommer göra nu är att dela upp oss så att alla kommer vara i varsitt rum och så tar vi på oss XR-headseten och loggar in. Vi ska gå igenom headseten lite kort här och sen följer vi med och hjälper er ta på dem och ser till att allt funkar som det ska. Har ni några frågor?

Förklara headseten, kontrollerna, passform

Gå in i separata rum med varsitt headset, rätt rum med rätt deltagare. Förklara headseten igen och säg att de ska vänta på oss inne i MR. Följ checklista

Workshop

Scen 1 - introduktion

Hej igen alla. Kan ni höra mig? Kan alla testa och prata så att vi dubbelkollar att alla hörs
om inte, berätta hur de kan sätta på mikrofon i Campfire

Känns det okej för alla? Ser alla bra? Om det krånglar med tekniken, om ni känner er obekväma eller behöver en paus så är det bara att säga till så kommer vi och hjälper till. Lovisa är tech-support.

Så, nu ser vi ju lite annorlunda ut så vi kör en presentation igen.

Alla presenterar sig

Som vi gick igenom innan så är det knappen A som ni använder för att få upp en meny med några verktyg. De två knapparna där bak som ni känner med fingrarna styr vissa av verktygen. Testa att trycka upp er meny via knappen A, funkar det? Härifrån kan ni skriva en kommentar, rita något, peka på grejer osv.

Då drar vi igång med workshopen. Först ska vi sätta oss in i situationen med att köra en lastbil så nu kommer jag byta miljö här och omgivningen kommer ändras.

Byt till kontext lastbil

Scen 2 - kontext lastbil

Här ser vi produkten som ska utvärderas, alltså en lastbil från Volvo. Har ni kört någon lastbil från Volvo innan? Får ni prova att köra många olika lastbilar i utbildningen?

Scen 3 - kontext köra

Här är vi istället inne i lastbilen och får en närmre bild av instrumentpanelen och vyn man har som förare. Känner ni igen er från modeller som ni har kört tidigare? Är ni vana vid en skärm på instrumentpanelen? Är ni bekanta med knapparna som syns i bilden?

Scen 4 - instrumentpanel

Jag tänker att vi kan börja prata om vad som är viktigt med en instrumentpanel när man är ute och kör? Det kan vara vad som helst ni tänker på. Hur ser den ideala instrumentpanelen ut? Något ni gillat extra mycket från olika ni kört?

Då ska vi gå in i den sista omgivningen och det är där vi ska vara resten av workshopen. Ni kommer se en instrumentpanel här alldeles strax. Den tillhör en bil egentligen, men försök ha med er kontexten som vi har här framför oss och som ni är vana med från utbildningen.

Byt till instrumentpanelen

Ser alla instrumentpanelen? Toppen! Om vi går in mer specifikt och pratar om den instrumentpanelen vi har framför oss, vad tänker ni om den? Vad gillar ni? Vad gillar ni inte?

Interaktion 1

Nu tänker jag att vi ska testa att sitta vid ratten och interagera med instrumentpanelen. Vi kommer fokusera på knapparna och skärmen här som styr musiken. Det ska finnas en stol i varje rum så XXX vill du börja och sätta dig framför ratten?

Den första deltagaren sätter sig framför ratten

Nu hör du kanske lite musik (spelas via tech-ansvarig)? Situationen är att du är ute och kör, men inser att du inte gillar låten som spelas och du vill byta. Så nu kan du testa att byta till nästa låt via den här touch-skärmen.

Då testar vi igen med samma situation att du är ute och kör och musik spelas men du vill byta till nästa låt. Den här gången har din lastbil inte en touchskärm utan du måste trycka på rätt knapp här under skärmen och sedan vrida till nästa låt. Så nu kan du testa att byta till nästa låt på det sättet.

Diskutera testet

Hur kändes det? Är det något i den här interaktionen som du hade velat ändra? Förbättra? Hur hade ni velat göra det här om ni var ute och körde lastbil? Varför?

Interaktion 2

Då kan vi byta så att XXX sätter sig framför ratten istället.

Den andra deltagaren sätter sig framför ratten

Nu ska vi prata om placering på saker istället. Vi kan börja med skärmen här.

Flytta runt skärm

Flytta runt ratt

Flytta runt mittkonsol

Hur kändes det? Vilka funktioner tycker ni är viktiga att de finns nära på instrumentpanelen när ni kör? Vilka är mindre viktiga? Hur hade ni velat ha det här om ni var ute och körde lastbil? Varför?

Färg och material

Om vi nu istället går tillbaka och kollar på hela instrumentpanelen igen, vad tycker ni om färgsättningen på de olika delarna? Materialval?

Testa att ändra till vad de säger, snabba iterationer

Hur känns det med de nya ändringarna? Är det något mer ni hade velat testa att ändra? Placering, material, färg?

Avslutning

Då är det snart dags för oss att avsluta. Jättebra jobbat allihopa! Om jag försöker summera det som sagts idag så har vi kommit fram till att XXX... Håller ni med om det? Är det något mer ni vill lägga till? Då avslutar vi här i MR och samlas igen i första rummet där vi sågs i början. Ni kan ta av er headsetet och vänta i rummet så kommer vi och möter er!

Vi hämtar headseten och visar dom tillbaka till stora rummet

Nu ska vi ha en gemensam reflektion kring hur ni tyckte att workshopen var.

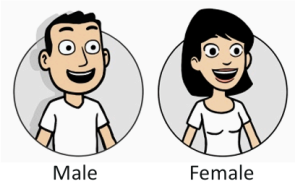
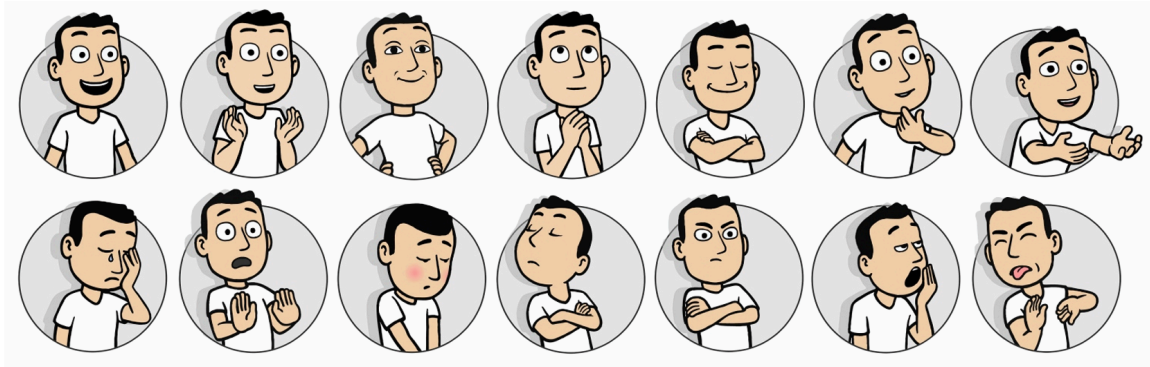
- Hur tyckte ni att workshopen kändes?
- Hur tyckte ni det var att ha en workshop i XR? Hur fungerade tekniken? Vad var bra/dåligt? Var det något ni önskade att ni kunde göra i VR som ni inte kunde nu?
- Vad tycker ni om resultatet vi kom fram till i workshopen?
- Hur tror ni resultatet hade blivit om workshopen var fysisk istället (utanför VR)? *Om det inte hade gått att köra lastbilen, men se bilder/videos på det. Gav det mer eller mindre än om man hade gjort det i verkligheten?*
- Om ni hade gjort workshopen igen, vilka förändringar hade ni velat se?

Stort tack för att ni tog er tid att komma hit och hjälpa oss med detta!

Dela ut goodiebags

Appendix C - PrEmo Instrument for Reliability Test

The self-assessment tool used by participants during the reliability emotion interpretation test, which was a PrEmo Emotion Measurement Instrument, is shown below.



Pride
The feeling when you possess (or have accomplished) something that exceeds your own expectations, or that others find praiseworthy.



Shame
The feeling when someone knows something bad about you. You believe this to be true, so you also feel bad about yourself.



Admiration
The feeling when you look up to someone who has excellent abilities or impressive accomplishments.



Contempt
The feeling when someone is inferior or unworthy in your eyes; they have done something blameworthy.



Joy
The feeling when something good happens to you. You fulfil a need, achieve something, or make progress towards achieving a goal.



Sadness
The feeling when you have lost something that was important to you, and believe it cannot be undone.



Hope
The feeling when you believe (but are not certain) that something good or desirable may happen in the future.



Fear
The feeling when something can harm you or someone you care about, and you are not sure whether you can do anything about it.



Satisfaction
The feeling when something meets or exceeds your expectations.



Anger
The feeling when someone does something bad that harms or offends you, and you blame that other person.



Desire
The feeling when you anticipate that something will be beneficial once acquired or consumed.



Disgust
The feeling when you find something repulsive and want to avoid having contact with it in any way.



Fascination
The feeling when you encounter something new and interesting that you do not immediately understand.



Boredom
The feeling when there is nothing interesting or engaging to your experience.



PrEmo Instrument (Laurans & Desmet, 2017). CCBY-NC-ND4.0

Appendix D - Self-Assessment for Reliability Test

How did you feel when you watched the videos? Select as many PrEmo-avatars as you like to express how you felt when watching the videos. If none of the avatars represent your emotions, you don't have to choose any. Please list the numbers of the avatars you selected and describe your feelings in more detail through text.

1. First Video: *Teacher*

Number on chosen avatar(s):

2. Second Video: *Babies*

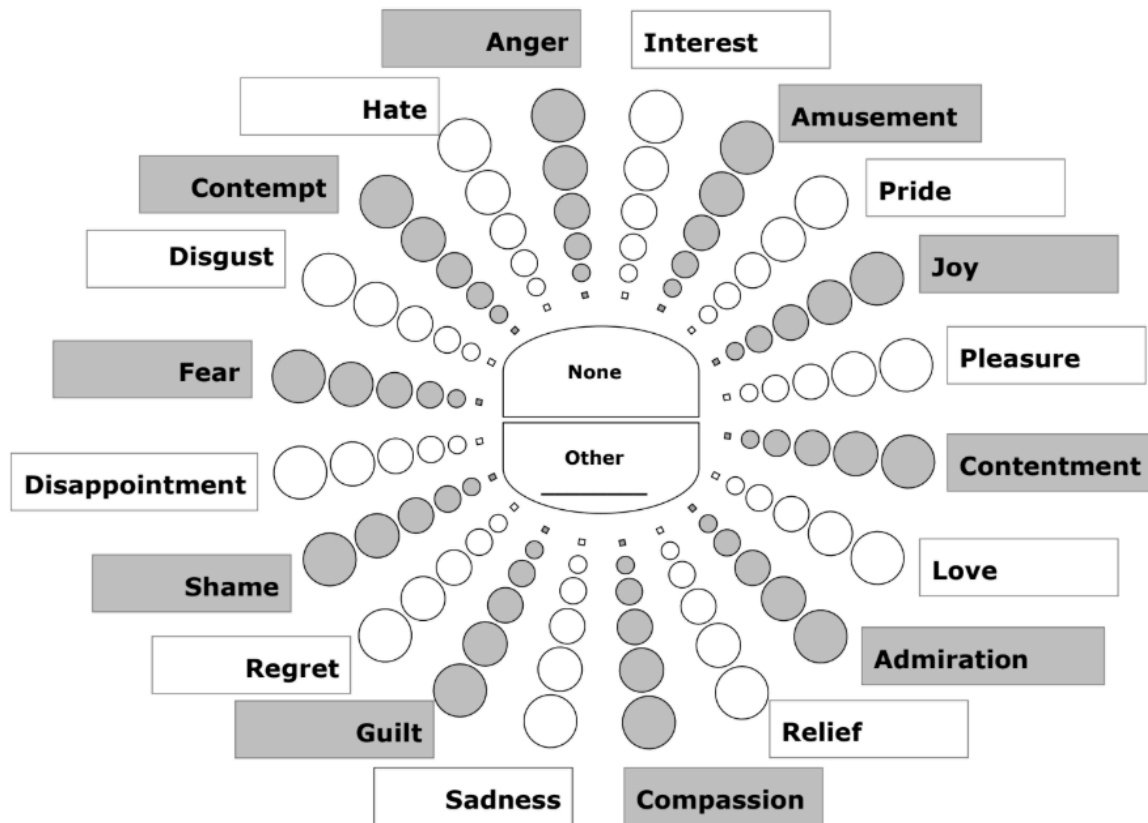
Number on chosen avatar(s):

3. Third Video: *Food*

Number on chosen avatar(s):

Appendix E - Self-Assessment for Context-Based Test

The self-assessment tool used by participants during the context-based emotion interpretation test, which was a Geneva Emotion Wheel, is shown below.



Geneva Emotion Wheel (GEW; see Scherer, 2005; Scherer et al., 2013).

Appendix F - Script for Interaction-Based Test

Hej! Tack för att du ville ställa upp och komma hit idag för att hjälpa oss! Innan vi börjar så får du gärna skriva på 2 exemplar av den här GDPR-blanketten. Det är något som måste göras i sådana här studier med personuppgifter, vilket i detta fall är ålder, kön samt ljud och videoupptagning via ett AI-baserat datorprogram.

Skriver på GDPR

Presentation av oss

Toppen! Då ska vi börja med att förklara lite mer om varför ni är här. Jag heter Jonna och detta är Lovisa, och vi skriver just nu vårt examensarbete på Chalmers och Volvo Group Trucks Technology. Vårt arbete handlar om hur man kan utvärdera produkter ihop med dem som ska använda dem, så mer specifikt för Volvo Group Trucks Technology hur man kan utvärdera en lastbil med lastbilsförare. Och vi kollar på hur man kan göra det med olika typer av framtida teknik.

Presentation av dem

Du får gärna berätta lite om dig, vad du gör/arbetar med, hur gammal du är, hur länge du haft körkort, vilka typer av körkort du har, om du gjort någon sån här utvärdering innan och om du testat en körsimulator.

Mer om studien

Då tänker vi att vi ska förklara syftet med utvärderingen lite mer. Vi kollar som sagt på hur utvärderingar kan genomföras med framtida teknik, för att man ska kunna få mer eller djupare insikter om användaren. I just det här testet ska vi spela in med hjälp av den här kameran och sen använda en teknik som analyserar den filmen med hjälp av AI och se vad den analysen kommer fram till jämfört med vad du själv säger att du upplevde när du använde instrumentpanelen.

Det finns alltså inget rätt eller fel här idag eller att du behöver prestera något utan vi är intresserade av att höra var du tycker och tänker om instrumentpanelen!

Set-up

Under den här utvärderingen ska du få testa att köra en lastbil i den här körsimulatore. Du kommer köra en rutt och ska försöka köra som du gjort i vanliga fall. Under tiden kommer vi be dig utföra en uppgift på instrumentpanelen och sen fylla i ett sånt här papper med en självskattningsskala om hur du upplevde uppgiften. Totalt är det 3 uppgifter, men du kommer få göra varje uppgift på 2 olika sätt så totalt 6 omgångar. När alla uppgifter är genomförda så kör vi en generell diskussion om uppgifterna och om instrumentpanelen överlag.

Vi ska börja med att du får testa hur simulatore fungerar så du får gärna slå dig ner här i sätet och sätt dig bekvämt. Du kan justera stolen närmare eller längre från ratten och pedalerna med hjälp av spaken under sätet och du kan justera hur sätet är med hjälp av vreden på vänstra sidan där nere. Sitter du bekvämt? **Om inte, hjälp dem**

Sen ser du ratten här framför dig. Du behöver inte använda några av knapparna där, utan endast vrider på den som vanligt och använda de här spakarna för att blinka. Man kan använda korset för att vrida vyn, för det kan vara lite svårt att se på sidan. Men det är frivilligt. Här är växelspaken. Detta är en automatbil, så du behöver inte växla utan det är endast bak för att backa och fram med spaken för att köra. Behöver inte heller använda kopplingen (vänstra pedalen). Gas längst till höger, broms i mitten Du ska få köra en liten stund för att komma in i känslan hur simulatorn fungerar, och sen sätter vi igång.

Starta testbanan och låt användaren kör runt i några minuter

Gå igenom instrumentpanelen och ungefär hur uppgifterna kommer gå till

Gå igenom hur självskattningen kommer gå till, visa Geneva

Utvärdering med emotion interpretation

Starta testbanan - användaren börjar köra (vänta en liten stund innan uppgift 1)

Pausa banan mellan uppgifterna

Uppgift 1: Byt radiokanal (vi startar igång musik)

a) Fysiska knappar

Du är ute och kör och som du hör är musiken igång. Du inser att du inte gillar denna låt och vill byta radiokanal. Gör detta nu via fysiska knappar.

De byter låt och vi ändrar framen i Figma och vi startar nästa låt

Genomför kort självskattning (ingen diskussion)

b) Digital skärm

Du är ute och kör och som du hör är musiken igång. Du inser att du inte gillar denna låt och vill byta radiokanal. Gör detta nu via skärmen.

De byter låt och framen ändras i Figma och vi startar nästa låt

Genomför kort självskattning (ingen diskussion)

Uppgift 2: Koppla upp telefon på Bluetooth

a) Fysiska knappar

Du har precis startat din arbetsdag och börjar köra. 5 minuter in så märker du att du inte har kopplat upp din mobil på Bluetooth så du bestämmer dig för att göra det samtidigt som du fortsätter köra. Gör detta nu via fysiska knappar.

De kopplar upp sig och vi ändrar framen i Figma

Genomför kort självskattning (ingen diskussion)

b) Digital skärm

Du har precis startat din arbetsdag och börjar köra. 5 minuter in så märker du att du inte har kopplat upp din mobil på bluetooth så du bestämmer dig för att göra det samtidigt som du fortsätter köra. Gör detta nu via skärmen.

De kopplar upp sig och framen ändras i Figma

Genomför kort självskattning (ingen diskussion)

Uppgift 3: Sänk temperaturen

c) Fysiska knappar

Det är eftermiddag och du har varit ute och kört i några timmar nu. Solen ligger på och det börjar bli varmt i hytten. Du vill därför sänka temperaturen med 1.5 grader. Gör detta nu via fysiska knappar.

De sänker temperaturen och vi ändrar framen i Figma

Genomför kort självskattning (ingen diskussion)

d) Digital skärm

Det är eftermiddag och du har varit ute och kört i några timmar nu. Solen ligger på och det börjar bli varmt i hytten. Du vill därför sänka temperaturen med 1.5 grader. Gör detta nu via skärmen.

De sänker temperaturen och framen ändras i Figma

Genomför kort självskattning (ingen diskussion)

Efterföljande och jämförande intervju

Nu ska vi ha en generell diskussion om de olika uppgifterna, formulären du fyllde i, instrumentpanelen och själva utvärderingen. Vi kan börja med de olika uppgifterna och instrumentpanelen. Ta fika!

Gå igenom a) och b) på varje uppgift parallellt

Be dom beskriva varför de valt de olika känslorna

- Hur gick det (kändes) att utföra uppgifterna?
- Hur upplevde du de olika sätten att utföra uppgifterna på? Vad gillade du? Vad gillade du inte? Hur hade du velat göra om du var ute och körde? Varför?
- Vad tyckte du om den här instrumentpanelen?
 - Placeringen av knapparna?
 - Storlek på skärm och knappar?
 - Enkelhet att använda och förstå?

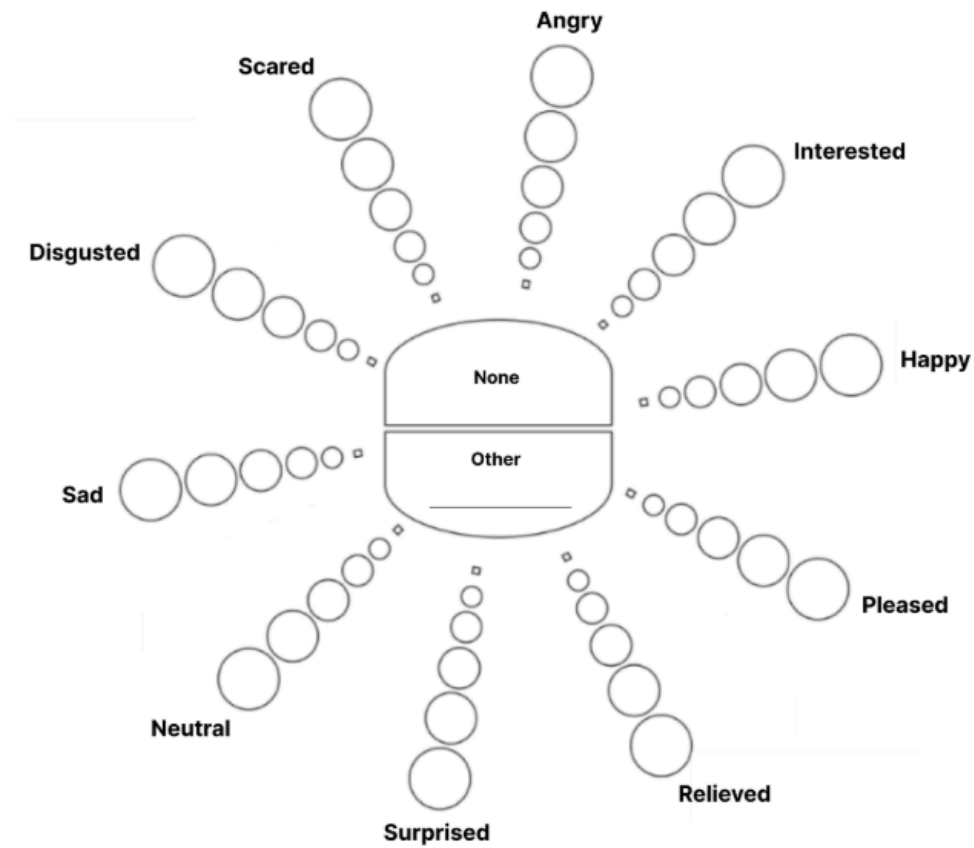
Då ska vi prata mer generellt om själva utvärderingen och testet.

- Hur tyckte du att utvärderingen kändes?
- Hur tyckte du det var att bli filmad under testet?
- Hur tyckte du det var att veta att teknik ska analysera filmen med hjälp av AI?
- Om du hade gjort utvärderingen igen, vilka förändringar hade ni velat se?

Dela ut goodiebags

Appendix G - Adapted Geneva Emotion Wheel

The self-assessment tool used by participants during the interaction emotion interpretation test, which was an adapted version of the Geneva Emotion Wheel, made by the project team, is shown below.



Adapted Geneva Emotion Wheel (GEW; see Scherer, 2005; Scherer et al., 2013).

DEPARTMENT OF INDUSTRIAL AND MATERIALS SCIENCE

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2025

www.chalmers.se



CHALMERS
UNIVERSITY OF TECHNOLOGY