



Fachhochschule Nordwestschweiz
Hochschule für Angewandte Psychologie

Mensch und Künstliche Intelligenz im User Experience Research: Eine Fallstudie zur Evaluierung KI-gestützter Analyseprozesse in einem Schweizer KMU

MASTER-THESIS

Eingereicht per: 01/2026

Autor

Lohri Chasper

Betreuungsperson

Prof. Dr. Hauser Mirjam

Praxispartnerin

sinnhaft GmbH

Kontaktperson

Honegger Luca

Zeichenzahl: 191'166

Zusammenfassung

Diese explorative Master-Thesis untersucht den Einsatz Künstlicher Intelligenz (KI) im Bereich des User Experience Research (UXR) mit Fokus auf den Analyseprozess von Usability-Tests. Im Zentrum der Fallstudie stehen Herausforderungen kleiner UX-Agenturen, exemplarisch dargestellt an der Praxispartnerin sinnhaft GmbH. Vor dem Hintergrund begrenzter Ressourcen sowie der raschen Entwicklung von KI wächst das Interesse an der Integration KI-gestützter Methoden in UXR-Prozesse, um Arbeitsabläufe effizienter zu gestalten und Ressourcen gezielter einzusetzen. Bislang ist jedoch weitgehend ungeklärt, wie zuverlässig und aussagekräftig KI-gestützte Methoden insbesondere bei der Analyse von Usability-Tests sind.

Ziel der Arbeit ist, Potenziale und Herausforderungen der KI-Integration im UXR zu untersuchen, daraus resultierende Anforderungen an Anwender*innen und Forschende zu identifizieren sowie Hinweise auf die Zuverlässigkeit und Aussagekraft KI-generierter Analyseergebnisse im Vergleich zu menschlichen Referenzanalysen zu gewinnen. Methodisch folgt die Arbeit einem explorativen, qualitativen Forschungsdesign. Neben einer Literaturrecherche wurden $n = 4$ halbstrukturierte Experteninterviews durchgeführt. Ergänzend wurde ein KI-gestützter Analyseprozess zur Auswertung von Usability-Tests entwickelt und durch einen Mensch-KI-Vergleich mit einer menschlichen Referenzanalyse evaluiert.

Die Ergebnisse zeigen, dass KI Effizienzgewinne ermöglicht und verschiedene Phasen des UXR-Prozesses sinnvoll unterstützen kann. Dieser Mehrwert ist jedoch kontextabhängig und an die Qualität sowie Aufbereitung der Daten gebunden. Der Mensch-KI-Vergleich verdeutlicht, dass KI relevante Usability-Probleme zwar identifizieren, menschliche Analysen jedoch nicht vollständig ersetzen kann. Gleichzeitig werden Limitationen festgestellt, welche erhöhte Anforderungen an menschliche Supervision, Methodenkompetenz und einen reflektierten und kritischen Umgang mit KI-Ergebnissen mit sich bringen. Die Arbeit leistet einen Beitrag zur Bearbeitung einer Forschungslücke und liefert wissenschaftliche sowie praxisrelevante Implikationen für die Integration von KI in UXR-Prozesse.

Schlagnworte: Künstliche Intelligenz, User Experience Research, Usability-Test, Mensch-KI-Kollaboration, Qualitative Datenanalyse

Abstract

This exploratory master's thesis investigates the use of artificial intelligence (AI) in the field of User Experience Research (UXR), with a specific focus on the analysis of usability testing. The case study centres on challenges faced by small UX agencies, illustrated through the project partner sinnhaft GmbH. Given limited resources and the rapid advancement of AI technologies, interest in integrating AI-supported methods into UXR is growing to improve efficiency and optimize resource allocation. However, it remains unclear how reliable and meaningful AI-supported methods are, particularly for usability test analysis.

The aim of this thesis is to examine the potentials and challenges of AI integration into UXR, to outline resulting requirements for practitioners and researchers, and to assess the reliability and interpretability of AI-generated analysis results in comparison to human reference. The study adopts an exploratory qualitative research design, combining a literature review with four semi-structured expert interviews. Furthermore, an AI-supported analysis for usability tests was developed and evaluated through a human-AI comparison.

The findings indicate that AI can provide efficiency gains and support across multiple phases of the UXR process. However, its added value is context-dependent and linked to data quality and preparation. While AI can identify relevant usability issues, it cannot fully replace human analysis. Identified limitations underscore the need for human supervision, methodological expertise, and a critical and reflective use of AI-generated results. This thesis contributes to addressing a research gap and provides both scientific and practical implications for integrating AI into UXR.

Keywords: Artificial Intelligence, User Experience Research, Usability Testing, Human-AI Collaboration, Qualitative Data Analysis, AI-supported Analysis

Inhaltsverzeichnis

1	Einleitung	1
1.1	Vorstellung Praxispartnerin	2
1.2	Ausgangslage und Problemstellung	2
1.3	Forschungsinteresse und Zielsetzung der Arbeit	3
1.4	Fragestellung und Abgrenzung	3
1.5	Aufbau der Arbeit	4
2	Theoretische Grundlagen	5
2.1	User Experience Research und Usability-Tests	5
2.1.1	User Experience und Usability	5
2.1.2	Grundlagen und Ziele von UXR	8
2.1.3	Phasen und Methoden des UXR-Prozesses	9
2.1.4	Usability-Tests und Auswertung	11
2.2	Künstliche Intelligenz	14
2.2.1	Funktionsweise von LLMs	16
2.2.2	Potenziale des KI-Einsatzes	17
2.2.3	Risiken und Herausforderungen des KI-Einsatzes	18
2.2.4	Mensch und KI	19
2.3	Einsatzmöglichkeiten von KI im UXR	21
2.3.1	Planung und Datenerhebung	22
2.3.2	Datenaufbereitung und Transkription	22
2.3.3	Analyse und Interpretation	23
2.3.4	Generative KI und synthetische Personas	24
2.3.5	Grenzen des KI-Einsatzes	25
3	Methodik	26
3.1	Forschungsdesign	26
3.2	Datenerhebung	28
3.2.1	Stakeholderinterview	28
3.2.2	Literaturrecherche	28
3.2.3	Leitfadenerstellung	29
3.2.4	Durchführung Experteninterviews	30
3.2.5	Forschungstagebuch	31
3.3	Transkription und Auswertung Experteninterviews	31
3.3.1	Transkription	31

3.3.2	Qualitative Analyse.....	32
3.4	Fallstudie.....	33
3.4.1	Beschreibung Fallstudie	33
3.4.2	Beschreibung des Datensatzes.....	34
3.4.3	KI-Transkription Usability-Test	36
3.5	Mensch-KI-Vergleich.....	36
4	Ergebnisse	41
4.1	Experteninterviews.....	41
4.1.1	Aktueller Stand der Forschung zu KI-Einsatz.....	41
4.1.2	Mensch und KI im UXR	43
4.1.3	Limitationen und Herausforderungen des KI-Einsatzes	48
4.2	Mensch-KI-Vergleich.....	50
4.2.1	Vorbereitungen und Zeitaufwand	50
4.2.2	Zuverlässigkeit und Aussagekraft KI-Analyseergebnisse	52
4.2.3	Qualitative Einordnung zusätzlicher KI-identifizierter Usability-Probleme	54
5	Diskussion.....	56
5.1	Potenziale und Herausforderungen der KI-Integration im UXR.....	56
5.1.1	Potenziale der KI-Integration im UXR	56
5.1.2	Herausforderungen und Risiken der KI-Integration im UXR	58
5.2	Anforderungen an Anwender*innen bzw. Forschende	61
5.3	Einsatz von KI im UXR und Validität der KI-Ergebnisse	63
6	Fazit und Ausblick	68
7	Limitationen	70
	Literaturverzeichnis.....	72
	Abbildungsverzeichnis	79
	Tabellenverzeichnis.....	80
	Glossar	81
	Hilfsmittelverzeichnis mit Verwendungszweck	83
	Anhang	84

1 Einleitung

Mit Künstlicher Intelligenz (KI) befinden wir uns möglicherweise bereits inmitten in einer neuen technologischen Revolution, welche das Potenzial haben könnte, sowohl die Arbeitswelt als auch das alltägliche Leben grundlegend zu verändern (vgl. Brandao, 2025; Kunz, Sajtos & Flavián, 2025; Stockman, 2025). Bereits 1960 hatte der Psychologe und Computerpionier J. C. R. Licklider die Vision einer engen Zusammenarbeit zwischen Mensch und Maschine und hielt in seinem wegweisenden Aufsatz «Man-Computer Symbiosis» fest: «In not too many years, human brains and computing machines will be coupled together very tightly, and the resulting partnership will think as no human brain has ever thought» (Licklider, 1960, S. 4).

Um ein Produkt oder eine Dienstleistung erfolgreich am Markt etablieren zu können, reicht es nicht aus, ausschliesslich betriebswirtschaftliche Aspekte zu berücksichtigen. Es gilt auch, die Bedürfnisse und Anforderungen der Zielgruppe zu kennen, zu verstehen und das Produkt dementsprechend auszurichten (vgl. Fantapié Altobelli, 2023; Graner, 2015; Redler & Ullrich, 2021). Bei der Entwicklung und Gestaltung erfolgreicher (digitaler) Produkte und Dienstleistungen spielt daher User Experience Research (UXR) eine zentrale Rolle. Unternehmen führen fundierte UX-Analysen durch, um ihre Angebote gezielter auf die Bedürfnisse ihrer Zielgruppe bzw. ihren Nutzer*innen auszurichten und so die Kundenzufriedenheit sowie Wettbewerbsfähigkeit zu steigern (vgl. Aitim & Abdulla, 2024; Sarker, 2024; Unterkalmsteiner et al., 2016). Dabei verfügen Teams, die eng mit der Produktentwicklung zusammenarbeiten, häufig nur über begrenzte zeitliche Ressourcen und stehen nicht selten unter erheblichem Zeitdruck (Aitim & Abdulla, 2024; Flaherty & Dykes, 2024; Hokkanen & Väänänen-Vainio-Mattila, 2015; Norman, 2013; Redler & Ullrich, 2021).

Die rasante Entwicklung der KI scheint vielfältige und neue Möglichkeiten zur Bewältigung der erwähnten Aufgaben und Herausforderungen zu bieten (vgl. Fragiadakis, Diou, Kousiouris & Nikolaidou, 2024; Najana, Bhattacharya, Kewalramani & Pandiya, 2024; Sarker, 2024). Doch wie nützlich und aussagekräftig kann die Unterstützung durch eine KI tatsächlich sein? KI-gestützte Methoden könnten dazu beitragen, UXR oder allgemeiner eine Vielzahl interner Prozesse effizienter und kostengünstiger zu gestalten, indem sie bestimmte Arbeitsschritte automatisieren und beschleunigen (vgl. Brynjolfsson, Li & Raymond, 2025; Knearem, Khwaja, Gao, Bentley & Kliman-Silver, 2023). Gleichzeitig sind Unternehmen gefordert, sich mit den Potenzialen und Limitationen dieser neuen Technologie auseinanderzusetzen, um den Anschluss an aktuelle Entwicklungen und Innovationen nicht zu verlieren. Obgleich KI-gestützte Methoden als vielversprechende Ergänzung für UXR erkannt werden, besteht im professionellen Kontext Unsicherheit hinsichtlich ihrer Validität, Verlässlichkeit und praktischen Anwendbarkeit. Eine unkritische oder unreflektierte Verwendung von KI-Tools (dt. Werkzeug / Softwarelösungen) könnte zu Fehlinterpretationen oder einer Verzerrung der Ergebnisse führen, was weitere Auswirkungen auf Produktentscheidungen haben kann (Knearem et al., 2023).

1.1 Vorstellung Praxispartnerin

Die vorliegende Masterarbeit entsteht in Kooperation mit der Praxispartnerin *sinnhaft GmbH*, einer UX-Agentur mit Sitz in Zürich, Schweiz. Die *sinnhaft GmbH* unterstützt andere Unternehmen in einer Business-to-Business (B2B) Beziehung bei der Entwicklung und Optimierung digitaler Produkte und Dienstleistungen. Neben UX-Analysen und (Usability)Tests bietet die *sinnhaft GmbH* auch strategische Beratungen, Workshops, Schulungen sowie Coachings an. Die *sinnhaft GmbH* arbeitet bei der Gestaltung ihrer Produkte vorwiegend entlang dem Human Centered Design Process (HCD) (dt. menschenzentrierten Gestaltungsprozess).

Aktuell beschäftigt die Praxispartnerin sechs Personen und zählt somit zu den kleinen und mittleren Unternehmen (KMU). Wie bei anderen KMU in der Schweiz sind die begrenzten Ressourcen auch für die *sinnhaft GmbH* ein andauerndes Thema, sei es finanzieller oder personeller Natur.

1.2 Ausgangslage und Problemstellung

In der Schweiz gelten gemäss Angaben des Bundesamtes für Statistik (BFS, 2024) 99 % aller Unternehmen als KMU, also kleine und mittlere marktwirtschaftliche Unternehmen mit weniger als 250 Beschäftigten. Besonders für solche KMU kann UXR eine Herausforderung darstellen. Begrenzte personelle und finanzielle Ressourcen erschweren die Durchführung umfangreicher Nutzerstudien, wodurch wichtige Erkenntnisse für nutzerzentrierte Produktentscheidungen fehlen können und der Zugang zu grösseren sowie aufwändigeren Projekten eingeschränkt werden kann.

Folglich möchte sich die Praxispartnerin kontinuierlich verbessern, wettbewerbsfähig und technologisch auf dem neuesten Stand bleiben, um den Anschluss nicht zu verlieren. Sie interessiert sich daher für die Integration von KI in ihre Geschäftsprozesse. Trotz begrenzter Ressourcen strebt sie danach, ihren Kund*innen qualitativ hochwertige Einblicke in die Nutzererfahrung zu bieten und gleichzeitig wirtschaftlich wettbewerbsfähig zu bleiben. Zudem sieht sie Potenziale in der Anwendung von KI, um schlankere und effizientere Prozesse sowie kostengünstigere Produkte anbieten zu können. Diese sollen auch Kundschaft mit eingeschränktem Budget ansprechen, die je nach Projekt nicht unbedingt auf menschliche Auswertungen angewiesen sind. Die Praxispartnerin steht daher vor der Frage, wie zuverlässig und aussagekräftig die durch KI generierten Erkenntnisse aus UX-Analysen tatsächlich sind und in welchen Anwendungsfällen sich der Einsatz von KI besonders lohnen könnte.

Zusätzlich stellt sich die Frage, wie KI sinnvoll in bestehende UXR-Prozesse integriert werden kann. Während KI bestimmte Analyseaufgaben übernehmen kann, bleibt unklar, inwieweit sie menschliche UX-Researcher*innen ergänzen oder möglicherweise ersetzen kann. Durch die rasante Entwicklung der KI-Technologie fehlt es an wissenschaftlich fundierten Evaluierungen, welche die Potenziale und Limitationen von KI im UXR systematisch untersuchen. Besonders relevant ist dabei, inwieweit sich die Erkenntnisse aus KI-gestützten UXR-Methoden mit den Ergebnissen klassischer, von Menschen durchgeführter Analysen übereinstimmen oder sich von diesen unterscheiden.

1.3 Forschungsinteresse und Zielsetzung der Arbeit

Ziel der vorliegenden Masterarbeit ist es, die Anwendbarkeit KI-gestützter UXR-Methoden sowie die Zuverlässigkeit und Aussagekraft der KI-generierten Erkenntnisse anhand einer Fallstudie der sinnhaft GmbH explorativ zu untersuchen und mit einer klassischen, menschlichen Referenzanalyse zu vergleichen. Dabei sollen insbesondere die Potenziale und Limitationen von KI im UXR analysiert werden, um KMU fundierte Entscheidungsgrundlagen für den Einsatz dieser Technologie zu bieten. Dies soll ihnen helfen, den Anschluss an aktuelle Entwicklungen nicht zu verlieren und durch effizientere Prozesse wettbewerbsfähig bleiben zu können.

Darüber hinaus soll die Arbeit einen Beitrag zur aktuellen Forschung im bislang wenig untersuchten Feld von KI im UXR leisten. Ziel ist es, die Zuverlässigkeit und Aussagekraft der durch KI generierten Erkenntnisse zu analysieren und potenzielle Einsatzmöglichkeiten in der Praxis aufzuzeigen. Bislang fehlt weitgehend empirische Evidenz dazu, wie zuverlässig und aussagekräftig KI-gestützte Tools in UXR- bzw. Usability-Analysen tatsächlich sind (Liu, 2025). Diese Forschungslücke ist besonders relevant, da UX-Teams und Unternehmen zunehmend vor der Frage stehen, inwiefern KI sinnvoll in Analyseprozesse integriert werden kann. Die Arbeit soll damit nicht nur wissenschaftliche Erkenntnisse liefern, sondern auch Implikationen für den praktischen Einsatz bieten.

1.4 Fragestellung und Abgrenzung

Mit Blick auf die zuvor beschriebene Problemstellung und den aktuellen Stand der Forschung, wurde die Fragestellung wie folgt festgelegt:

Wie kann Künstliche Intelligenz (KI) genutzt werden, um den UX-Research-Prozess in kleinen Unternehmen (KMU) zu unterstützen und welche Unterschiede bestehen zwischen einer KI-gestützten und einer menschlichen Analyse?

Die folgenden zwei Leitfragen sollen dazu beitragen, die Hauptfragestellung umfassend zu beantworten:

- *Welche Potenziale und Herausforderungen ergeben sich bei der Integration von KI-gestützten Prozessen im UX-Research eines KMU?*
- *Welche Anforderungen stellt die Integration von KI in den UX-Research Prozess an die Anwender*innen bzw. Forschenden?*

Abgrenzung: Der Einsatz von KI im UXR bzw. im Kontext von Usability-Tests ist derzeit aufgrund der eingeschränkten Fähigkeit vieler KI-Systeme, multimodale Daten zu verarbeiten, technisch begrenzt. Vor diesem Hintergrund besteht im Rahmen der Konzeption des KI-gestützten Analyseprozesses kein Anspruch darauf, ein neues KI-Tool zu entwickeln oder eigene Modelle zu programmieren. Stattdessen werden bestehende, zum Zeitpunkt der Erstellung der Arbeit verfügbare

KI-Tools und -Modelle eingesetzt. Die Auswahl orientiert sich dabei primär an der praktischen Relevanz und Anwendbarkeit für die Fallstudie sowie für die Praxispartnerin. Die Beschreibung der KI-Technologie erfolgt bewusst nicht in technischer Tiefe, sondern beschränkt sich auf jene Aspekte, die für das Verständnis der eingesetzten Methoden und der Ergebnisse dieser Arbeit erforderlich sind. Weiter verfolgt die Arbeit nicht das Ziel, die Usability oder die UX des untersuchten Produkts der Fallstudie selbst zu analysieren und zu bewerten. Der Fokus liegt ausschliesslich auf dem Analyseprozess und dem Vergleich zwischen menschlicher und KI-gestützter Auswertung. Obwohl der übergeordnete Untersuchungsrahmen im Bereich des UXR angesiedelt ist, liegt der Fokus des Mensch-KI-Vergleichs, sowie der daraus abgeleiteten Ergebnisse auf dem Usability Bereich. Dies ist auf den verwendeten Datensatz zurückzuführen, der aus Usability-Testdaten besteht.

1.5 Aufbau der Arbeit

Auf die Einleitung folgt in Kapitel 2 die Erläuterung der theoretischen Grundlagen der vorliegenden Master-Thesis. Zunächst werden die zentralen Konzepte und Definitionen aus dem Bereich des UXR und der Usability eingeführt, mit einem besonderen Fokus auf das Analyseinstrument der Usability-Tests. Anschliessend werden grundlegende Konzepte von KI erläutert, einschliesslich ihrer Funktionsweise sowie zentraler Potenziale und Risiken. Den Abschluss des Kapitels bildet eine Darstellung der Einsatzmöglichkeiten von KI im UXR aus der Perspektive der wissenschaftlichen Literatur. Kapitel 3 widmet sich der Methodik. Nach der Beschreibung des zugrunde liegenden Forschungsdesigns werden die Datenerhebung sowie die Auswahl und Begründung der angewendeten Methoden erläutert. Darüber hinaus werden das Vorgehen der Datenaufbereitung und -auswertung sowie die zugrunde liegende Fallstudie vorgestellt. Abschliessend wird das Vorgehen des Mensch-KI-Vergleichs beschrieben. In Kapitel 4 werden die Ergebnisse der Experteninterviews sowie die Resultate des Mensch-KI-Vergleichs bei der Analyse von Usability-Tests aufbereitet und präsentiert. Darauf aufbauend erfolgt in Kapitel 5 die Diskussion der Ergebnisse und die Erkenntnisse der Arbeit in Bezug auf die Forschungsfragen sowie der bestehenden Literatur. Die Arbeit schliesst mit Kapitel 6, welches ein Fazit sowie einen Ausblick für zukünftige Forschung gibt, und Kapitel 7, in dem die Limitationen der vorliegenden Arbeit dargestellt werden.

2 Theoretische Grundlagen

Dieses Kapitel widmet sich den theoretischen Grundlagen der vorliegenden Arbeit und gibt einen Überblick über den aktuellen Stand der Forschung. Ziel ist es, wichtige Konzepte und Begriffe zu erläutern, welche für das Verständnis der Thematik sowie als Grundlage zur Beantwortung der Forschungsfrage relevant sind. Als Erstes werden die Grundlagen von UXR und Usability dargestellt, mit besonderem Fokus auf der Methode der Usability-Tests. Anschliessend werden zentrale Konzepte der KI sowie deren Funktionsweise, Potenziale und Risiken auf Basis der aktuellen wissenschaftlichen Literatur erläutert. Abschliessend werden die Einsatzmöglichkeiten und Grenzen von KI im UXR thematisiert.

2.1 User Experience Research und Usability-Tests

Die ersten Unterkapitel geben einen Überblick darüber, wie die Begriffe User Experience, Usability sowie UXR definiert und eingeordnet werden können und wie sich Usability-Tests in diesem Forschungsfeld verorten lassen. Diese stellen den Schwerpunkt des Mensch-KI-Vergleichs dar.

2.1.1 User Experience und Usability

Heutzutage übernimmt die User Experience (UX) (dt. Erfahrung von Nutzer*innen oder Benutzererlebnis) von Systemen, Produkten und Dienstleistungen zunehmend eine zentrale Rolle dafür ein, ob und in welchem Ausmass diese am Markt erfolgreich sind oder sein werden. Funktionale Merkmale sowie die Produktqualität werden von Kundinnen und Kunden zunehmend als selbstverständlich vorausgesetzt. Entscheidender wird darum vielmehr das Erleben der Nutzer*innen in der Interaktion mit dem Produkt. UX wird dadurch zu einem entscheidenden Erfolgsfaktor in der Produktentwicklung (Aitim & Abdulla, 2024; Hassenzahl, 2003; Weichert, Quint & Bartel, 2021). Auch aus Marketingsicht führen Redler und Ullrich (2021) aus, dass Produkte nur dann langfristig am Markt bestehen können, wenn sie konsequent auf die Bedürfnisse der Kund*innen ausgerichtet werden (vgl. Hokkanen & Väänänen-Vainio-Mattila, 2015; Unterkalmsteiner et al., 2016).

Zur konzeptionellen Einordnung von UX innerhalb der vorliegenden Arbeit wird zunächst der Begriff definiert. Wie der Name bereits andeutet, beschreibt dieser die Perspektive sowie die Erfahrungen der Nutzer*innen bei der Interaktion mit einem Produkt, System oder Service (Richter & Flückiger, 2016). Gemäss dem Deutschen Institut für Normung (DIN, 2010) und der Norm DIN EN ISO 9241-210 umfasst UX sämtliche Wahrnehmungen und Reaktionen einer Person, die *vor*, *während* oder *nach* der Nutzung eines Produktes, Systems oder Dienstes auftreten. Darunter fallen unter anderem Aspekte wie Emotionen, Erwartungen, Vorlieben, Erfahrungen sowie entsprechende Verhaltensweisen, die im Rahmen der Interaktion entstehen. Der UX-Ansatz geht damit über die traditionelle, eher rein funktional und aufgabenorientiert ausgerichtete Perspektive der Usability hinaus. UX stellt somit einen übergeordneten und umfassenden Ansatz zur Beschreibung und Analyse der subjektiven Nutzererfahrung in der Mensch-Technologie-Interaktion über alle Phasen

der Produktnutzung hinweg dar (Berni & Borgianni, 2021; Hassenzahl & Tractinsky, 2006; Lewis & Sauro, 2021; Martinelli, Lopes & Zaina, 2024; Vermeeren et al., 2010).

Dennoch sind die Begriffe UX und Usability eng miteinander verknüpft und werden in der Praxis teilweise sogar synonym verwendet (vgl. Berni & Borgianni, 2021; Kieffer, Rukonić, Kervyn de Meerendré & Vanderdonckt, 2019; Lewis & Sauro, 2021; Richter & Flückiger, 2016; Vermeeren et al., 2010). Beide Konzepte spielen eine bedeutende Rolle, wie Produkte oder Systeme mit Blick auf die Nutzung durch Menschen, entwickelt und anschliessend evaluiert werden (Lewis & Sauro, 2021). In der vorliegenden Arbeit werden die Begriffe UX und Usability jedoch bewusst getrennt betrachtet. Abbildung 1 verdeutlicht, welche Elemente zur UX gehören und zeigt, wie Usability als Teilbereich innerhalb der UX eingeordnet wird.

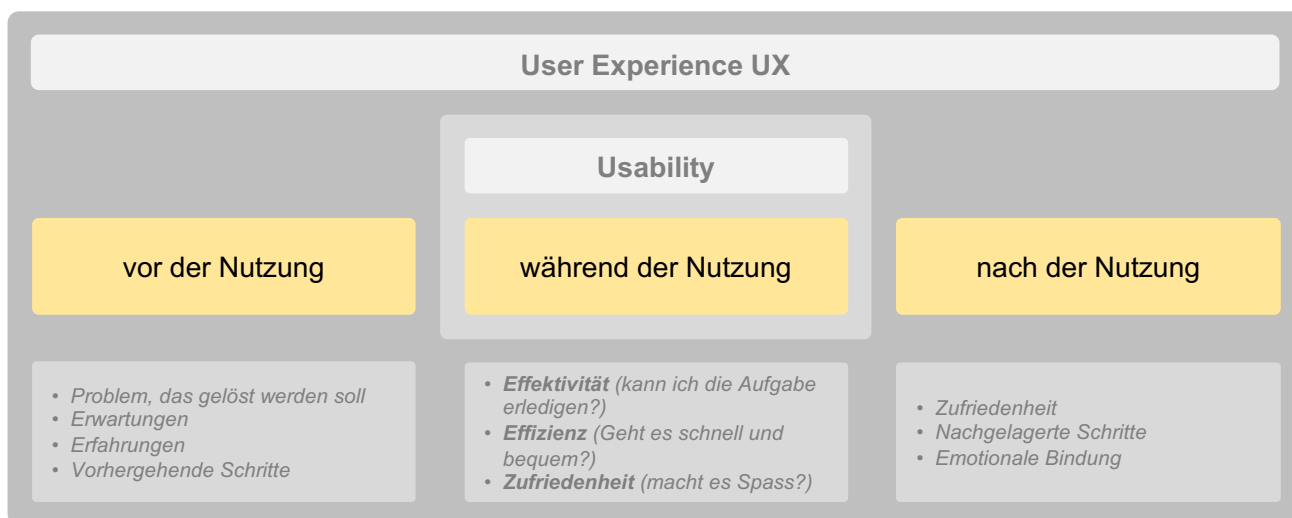


Abbildung 1. Übersicht über UX und Usability als Teilbereich (in Anlehnung an Jacobsen & Meyer, 2019, S.33, eigene Darstellung)

Usability, oder zu Deutsch *Gebrauchstauglichkeit* bzw. *Nutzerfreundlichkeit*, wird in der vorliegenden Arbeit, wie auch in der Praxis überwiegend, als englischer Begriff verwendet. Usability richtet die Perspektive vor allem auf die Aufgabenerfüllung bzw. Zielerreichung der Nutzer*innen *während* der Interaktion mit einem Produkt in einem spezifischen Nutzungskontext und stellt damit einen Teilbereich der umfassenderen UX dar. Das Usability-Konzept geht grundsätzlich davon aus, dass Nutzer*innen bei der Verwendung eines Produktes ein oder mehrere Ziele bzw. Aufgaben verfolgen (Geis & Tesch, 2023). Das Produkt, System oder der Dienst selbst wird dabei als Werkzeug bzw. Hilfsmittel verstanden, welches zur erfolgreichen Aufgabenerfüllung oder Zielerreichung benötigt bzw. verwendet wird. Darüber hinaus ist der jeweilige Nutzungskontext, in dem die Interaktion stattfindet, bei der Evaluation des Nutzungserlebnisses zu berücksichtigen (Richter & Flückiger, 2016). Das Zusammenspiel zwischen den einzelnen Komponenten von Usability wird in Abbildung 2 grafisch dargestellt und am Ende des Kapitels anhand eines Beispiels erläutert.

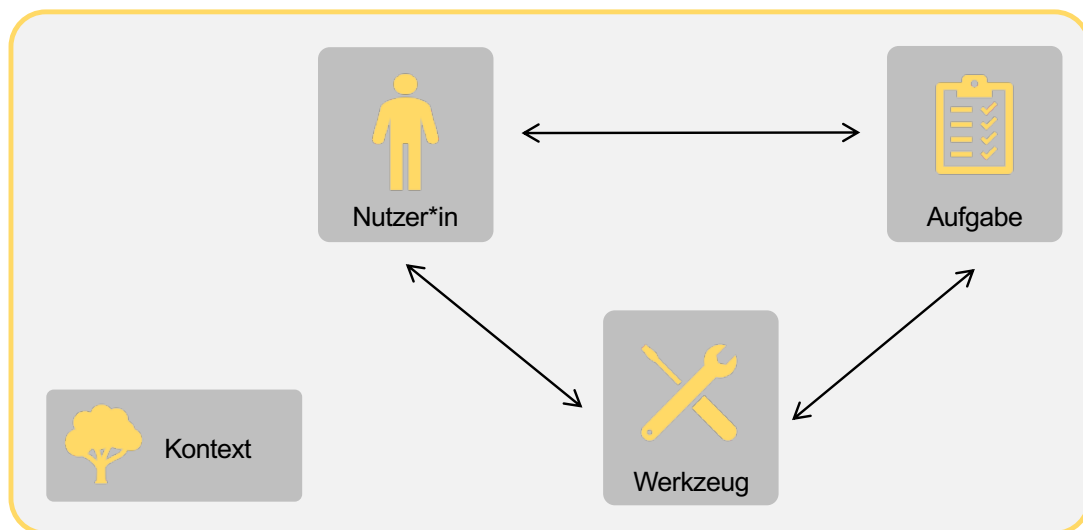


Abbildung 2. Die Komponenten von Usability: Nutzer*innen, Werkzeuge, Aufgaben und Kontext (in Anlehnung an Richter & Flückiger, 2016, S.11, eigene Darstellung)

Im Hinblick auf die Beurteilung der Aufgabenerfüllung lassen sich drei grundlegende Merkmale der Usability unterscheiden (Geis & Tesch, 2023; Jacobsen & Meyer, 2019):

- **Effektivität:** In welchem Ausmass ein angestrebtes Ziel tatsächlich erreicht wird.
- **Effizienz:** Welcher Aufwand bzw. welche Menge an Ressourcen erforderlich ist, um dieses Ziel zu erreichen.
- **Zufriedenheit:** Wie Nutzer*innen die Durchführung der Aufgabe wahrnehmen und beurteilen.

Neben dem Nutzungskontext, den Aufgaben bzw. Zielen und den Werkzeugen sind auch die Nutzer*innen selbst von zentraler Bedeutung. Als Nutzer*innen (engl. User) werden Personen verstanden, die direkt mit einem Produkt oder Service interagieren, um ein bestimmtes Ziel zu erreichen. Zwar existieren weitere Untergruppen von Nutzer*innen, diese sind für die vorliegende Arbeit jedoch nicht relevant und werden daher nicht weiter erörtert (Berni & Borgianni, 2021; Jacobsen & Meyer, 2019; Weichert et al., 2021).

Nicht zuletzt spielt auch der eigentliche Nutzen (engl. Utility) des Produktes eine entscheidende Rolle und darf als Grundvoraussetzung für die Nutzung eines Produktes, Systems oder Dienstes verstanden werden (Jacobsen & Meyer, 2019; Weichert et al., 2021).

Zur Veranschaulichung der zuvor erläuterten Konzepte wird im Folgenden ein vereinfachtes Beispiel herangezogen:

- **Ziel:** Einen neuen Job finden
- **Aufgabe:** Passende Stellen recherchieren und Bewerbungsunterlagen einreichen
- **Produkt / Werkzeug:** Online-Jobplattform
- **Kontext:** Nutzung über ein Smartphone
- **Nutzen:** Die Plattform unterstützt die Nutzer*innen effektiv, indem sie alle benötigten Funktionen bereitstellt, um das Ziel erfolgreich zu erreichen.

Usability: Bezieht sich hier auf die Nutzerfreundlichkeit der Jobplattform und die Effizienz sowie Effektivität bei der Bearbeitung der Aufgaben zur Zielerreichung.

User Experience: Umfasst alle weiteren Aspekte der Nutzung, wie etwa das subjektive Empfinden, die Motivation, Freude oder Frustration während der Interaktion, die wahrgenommene Ästhetik der Plattform sowie die emotionale Bindung an das Produkt.

Abschliessend sei erwähnt, dass die in der vorliegenden Arbeit verwendete Definition von UX gemäss der Norm DIN EN ISO 9241-210 in der wissenschaftlichen Literatur nicht unumstritten ist, jedoch trotzdem häufig angewendet und zitiert wird. Eine allgemein gültige, einheitliche Definition von UX scheint bisher nicht gefunden. Einige Forschende betonen, dass UX ein Konzept sei, welches sich nicht klar definieren lasse. Seine Grenzen seien unscharf definiert, da UX interdisziplinär verortet sei und in verschiedenen Disziplinen Anwendung finde. Darüber hinaus sei UX schwer zu messen und besitze bislang keine klar definierten methodischen Ansätze. Vielmehr werde auf Methoden und Verfahren aus vielen anderen Disziplinen zurückgegriffen (vgl. Berni & Borgianni, 2021; Hassenzahl & Tractinsky, 2006; Kieffer et al., 2019; Martinelli et al., 2024; Robinson, Lanius & Weber, 2018; Sauer, Sonderegger & Schmutz, 2020; Vermeeren et al., 2010).

2.1.2 Grundlagen und Ziele von UXR

Um auf Basis der zuvor dargestellten Konzepte ein erfolgreiches Produkt, System oder einen Dienst zu entwickeln und die UX fundiert zu untersuchen, müssen im ersten Schritt die Erwartungen, Bedürfnisse und Wünsche potenzieller Kundinnen und Kunden erhoben bzw. erforscht werden. Dazu kommen unterschiedliche Methoden wie Befragungen oder Tests zum Einsatz (Martinelli et al., 2024). Auch Prototypen- oder Produkttests spielen eine wichtige Rolle, um zu überprüfen, ob Nutzer*innen ihre Aufgaben und Ziele tatsächlich erreichen können und ob sich die reale UX eines Produktes von den, vom Design-Team geplanten Erfahrungen unterscheidet (Hassenzahl, 2003; Weichert et al., 2021).

Grundsätzlich kann bei der Erhebung von Nutzerdaten zwischen *User Research* und *User Experience Research (UXR)* unterschieden werden. In der Praxis wird jedoch häufig keine klare begriffliche Unterscheidung vorgenommen, weshalb beide Begriffe in der vorliegenden Arbeit weitgehend synonym verwendet werden. Ausserdem liegt der Fokus der vorliegenden Arbeit auf

UXR. Dennoch werden in späteren Kapiteln auch wissenschaftliche Quellen aus dem breiter gefassten Bereich des User Research herangezogen, insbesondere dort, wo die Behandlung von KI und deren Einsatz im Analyseprozess relevant ist. Dies ist sinnvoll, da sich beide Bereiche besonders im Hinblick auf verwendete Methoden und analytische Vorgehensweisen teilweise überschneiden.

- **User Research** bezieht sich allgemein auf die Erforschung von Bedürfnissen, Verhaltensweisen und Motivationen von Nutzer*innen. Der Fokus ist dabei breiter angelegt und lässt sich eher der Marktforschung oder den frühen Phasen der Produktentwicklung zuordnen.
- **User Experience Research (UXR)** konzentriert sich auf die systematische Erhebung ähnlicher Aspekte, jedoch spezifisch im Kontext der Interaktion von Nutzer*innen mit einem Produkt oder Service. Speziell im Fokus stehen dabei Wahrnehmungen, Emotionen, Erwartungen und Reaktionen der Nutzerinnen Ziel ist es, aus der Analyse dieser Daten relevante Erkenntnisse für die Produktentwicklung abzuleiten, um positive Nutzungserlebnisse zu schaffen.

(vgl. Fantapié Altobelli, 2023; Geis & Tesch, 2023; Graner, 2015; Kieffer et al., 2019; Martinelli et al., 2024)

Hassenzahl (2003) postuliert, dass bei der Entwicklung oder Evaluation von Produkten, die für Nutzer*innen ansprechend sind und gerne genutzt werden, nicht ausschliesslich pragmatische bzw. Usability-Aspekte im Zentrum stehen sollten. Vielmehr seien auch emotionale und hedonische Faktoren zu berücksichtigen. Entsprechend fließen sowohl Elemente der Usability als auch der UX in die Evaluation der Erfahrung bei der Interaktion mit einem Produkt oder Service mit ein. Eine gute Usability fällt den Nutzer*innen typischerweise nicht bewusst auf, während Defizite oder Probleme häufig deutlich wahrgenommen werden (Jacobsen & Meyer, 2019). Um diese Aspekte im Rahmen von Tests mit Nutzer*innen valide zu erfassen und einordnen zu können, sind geeignete Methoden sowie ein systematisches UXR-Vorgehen erforderlich (Kieffer et al., 2019; Martinelli et al., 2024; Vermeeren et al., 2010). Aufbauend darauf werden im nächsten Kapitel typische Phasen des UXR-Prozesses skizziert und gängige Methoden aus dem UXR-Bereich erörtert.

2.1.3 Phasen und Methoden des UXR-Prozesses

Bei der Entwicklung neuer Produkte oder Dienstleistungen orientieren sich viele Unternehmen am Human-Centered Design (HCD) beziehungsweise User-Centered Design (UCD). Im Rahmen dieser Arbeit werden die Begriffe zur Vereinfachung synonym verwendet, wobei durchgehend der Begriff HCD genutzt wird. Es handelt sich dabei um Varianten und Weiterentwicklungen, die sich bei genauer Betrachtung vor allem in der Zielsetzung, nicht jedoch in der grundsätzlichen Vorgehensweise unterscheiden (vgl. Richter & Flückiger, 2016; Weichert et al., 2021). Der HCD-Prozess richtet sich dabei nach der ISO-Norm ISO 9241-210. Weichert et al. (2021) definieren HCD

wie folgt: «Ein Vorgehen, das durch die direkte Einbeziehung der Nutzer, frühe Visualisierung in Form von Prototypen und ein iteratives Vorgehen sicherstellt, dass die Erwartungen der Nutzer erfüllt oder übertroffen werden und das Nutzungserlebnis positiv ausfällt» (S. 26).

Der HCD-Prozess wird in Abbildung 3 grafisch dargestellt. UXR kann dabei in allen Phasen des Prozesses eingesetzt werden und spielt eine zentrale Rolle, um Bedürfnisse, Erwartungen und Erfahrungen der Nutzer*innen systematisch zu erfassen und in die iterative Produktentwicklung einfließen zu lassen (Norman, 2013; Richter & Flückiger, 2016). Die vorliegende Arbeit fokussiert sich auf Grundlage der Fallstudie primär auf die Evaluationsphase des HCD-Prozesses.

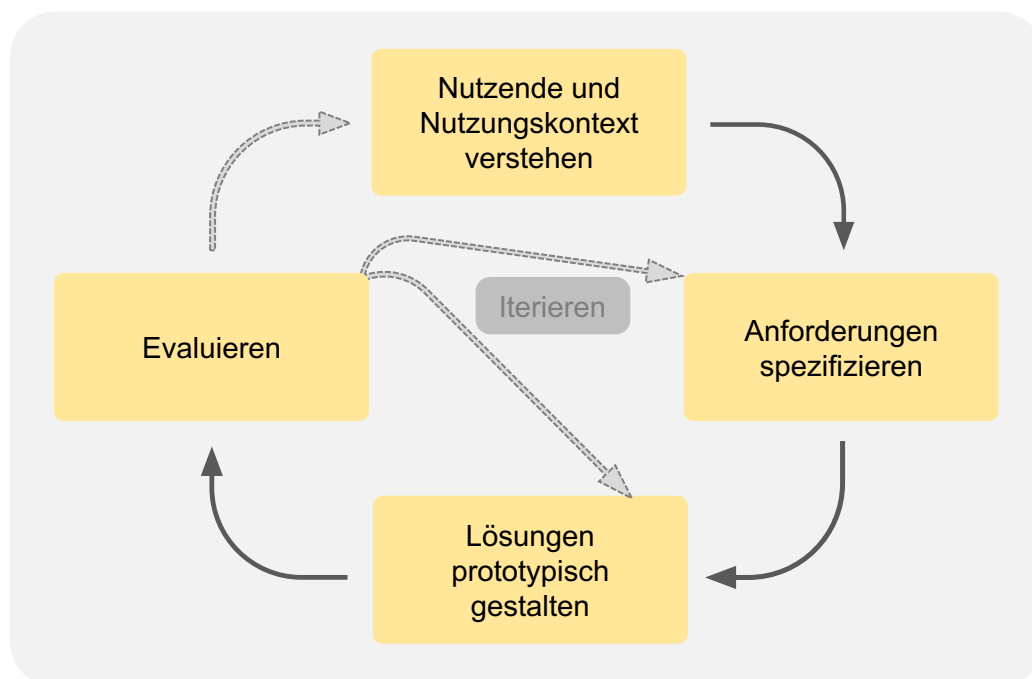


Abbildung 3. Human-Centered Design Prozess nach ISO 9241-210 (in Anlehnung an Weichert et al., 2021, S.25, eigene Darstellung)

Die Produktentwicklung entlang des HCD-Prozesses erfolgt iterativ und wird wesentlich durch Tests mit den Nutzer*innen beeinflusst. Erkenntnisse aus diesen Nutzertests können grundlegende Änderungen erforderlich machen, wenn sich bspw. zeigt, dass bestimmte Lösungen nicht wie vorgesehen funktionieren (Jacobsen & Meyer, 2019). Der Evaluationsprozess hat primär das Ziel zu prüfen, ob die entworfene Lösung die, in der UXR-Phase definierten UX-Ziele erfüllt. Hierbei wird die UX häufig anhand eines Prototyps analysiert und mit den festgelegten Anforderungen verglichen (Kieffer et al., 2019). Im UXR können je nach Phase des Entwicklungsprozesses unterschiedliche Methoden eingesetzt werden (Vermeeren et al., 2010).

Kieffer et al. (2019) weisen in ihrer Forschungsarbeit darauf hin, dass in der Literatur zu UX-Methoden unterschiedliche Perspektiven zwischen akademischer Forschung und der Praxis bestehen. Dies liegt unter anderem daran, dass die Meinungen über die Definition und die Reichweite von UX variieren, wie in Kapitel 2.1.1 bereits erläutert. Diese fehlende Einigkeit über die

Definition von User Experience führt zu Unsicherheiten bei der Messung und der Auswahl geeigneter Evaluationsmethoden. Zudem herrscht innerhalb der UX-Community Uneinigkeit darüber, ob qualitative oder quantitative Ansätze vorzuziehen sind (vgl. Kieffer et al., 2019; Vermeeren et al., 2010). Gleichzeitig bleibt UX als interdisziplinäres Feld methodisch unscharf abgegrenzt und greift auf Methoden unterschiedlichster anderer Forschungsdisziplinen zurück, ohne dass standardisierte Vorgehensweisen etabliert worden sind (Robinson et al., 2018). Kusuma, Jantan, Admodisaastro und Binti Mohd Norowi (2024) weisen darauf hin, dass in der Praxis der Fokus häufig auf der Phase *nach* der Produktinteraktion liegt. Da hier der Erfolg der UX unmittelbar messbar ist, können die Unterschiede zwischen praktischer und akademischer Perspektive teilweise erklärt werden. UX gilt zudem als schwer objektiv messbar, wobei Usability oft als leichter quantifizierbar betrachtet wird. Dies beeinflusst die methodischen Ansätze zur Messung und Evaluation von UX und Usability. Während die Usability durch ein breites Spektrum etablierter Methoden erhoben werden kann, stehen für die Messung von UX bislang deutlich weniger validierte Methoden und Werkzeuge zur Verfügung (Berni & Borgianni, 2021; Sauer et al., 2020). Der Fokus der vorliegenden Arbeit liegt einerseits deswegen und andererseits aufgrund der Fallstudie (Kapitel 3.4) auf den Methoden der Usability-Evaluation mit besonderem Schwerpunkt auf Usability-Tests. Besagte Methoden gelten aufgrund umfangreicher Forschung sowie der systematischen Erhebung von Daten als gut etabliert und methodisch ausgereift (Berni & Borgianni, 2021; Lewis & Sauro, 2021; Vermeeren et al., 2010).

2.1.4 Usability-Tests und Auswertung

In der Praxis werden häufig qualitative und beobachtungsbasierte Methoden eingesetzt, um nachvollziehen zu können, wie Nutzer*innen mit den Produkten oder Prototypen interagieren und welche Faktoren aus dem Kontext dabei eine Rolle spielen. Diese Methoden sind jedoch sehr zeitaufwendig, weshalb in der Praxis meist nur mit kleinen Stichproben gearbeitet wird (Norman, 2013). Gemäss Jacobsen und Meyer (2019) ist ein Usability-Test eine Evaluationsmethode, bei der Proband*innen aus der relevanten Zielgruppe dabei beobachtet und befragt werden, wie sie ein System in typischen Nutzungsszenarien bedienen und konkrete Aufgaben lösen. Dabei wird erfasst, wie Proband*innen mit dem Produkt oder Prototypen interagieren, um Usability-Probleme, Verständnisschwierigkeiten und Optimierungspotenziale aufzudecken (vgl. Lewis & Sauro, 2021; International Usability and UX Qualification Board, eingetragener Verein [UXQB], 2025). Usability-Tests werden als besonders wichtiges und flexibel einsetzbares Werkzeug innerhalb des Methodensets von Usability-Expert*innen angesehen (Jacobsen & Meyer, 2019).

Die Evaluation durch Usability-Tests kann sowohl in einem Labor als auch direkt im Feld oder auch Remote, also aus der Ferne durchgeführt werden. Zudem lässt sich zwischen moderierten und unmoderierten Usability-Tests differenzieren. Moderierte Tests ermöglichen tiefere Einblicke, da unklare Nutzungsverhalten unmittelbar hinterfragt werden können. Ferner bietet das moderierte Setting die Möglichkeit, offene Fragen oder Anschlussfragen zu stellen (Jacobsen & Meyer, 2019; UXQB, 2025). Bei moderierten Usability-Tests kann für das Nachfragen die Thinking-Aloud-Methode (dt. Lautes Denken) eingesetzt werden. Durch das kontinuierliche Verbalisieren ihrer Gedanken

werden unmittelbare Bewertungen, Annahmen und Fehlinterpretationen der Testperson erfasst, die in nachgelagerten Interviews häufig nicht mehr erinnerbar sind. Diese Methode stellt ein niedrigschwelliges Instrument dar, um authentische Einblicke in die subjektive Nutzererfahrung zu gewinnen. Eine Herausforderung besteht darin, die Testperson durch angemessene Moderation dazu anzuregen, das normalerweise ungewohnte Verbalisieren der eigenen Gedanken über die gesamte Testdauer hinweg aufrechtzuerhalten (Jacobsen & Meyer, 2019; UXQB, 2025).

Zur Interpretation und Analyse der Ergebnisse aus Usability-Tests lassen sich grundsätzlich zwei Formen unterscheiden. Die *formative Art* des Usability-Test verfolgt das Ziel, möglichst viele Informationen sowie potenzielle Usability-Probleme aufzudecken. Sie dient vornehmlich in den frühen Entwicklungsphasen der iterativen Weiterentwicklung und Verbesserung des untersuchten Produktes oder Systems. Demgegenüber steht der *summative Ansatz*, welcher typischerweise in späteren Entwicklungsphasen im Sinne einer Qualitätskontrolle eingesetzt wird. Dabei wird zusammenfassend die Gebrauchstauglichkeit eines weitgehend ausgereiften Produkts überprüft und beantwortet die Frage, ob Nutzer*innen das Produkt erfolgreich verwenden können (vgl. Kieffer et al., 2019; Lewis & Sauro, 2021; Richter & Flückiger, 2016; UXQB, 2025).

Es ist angezeigt, sich bei der Analyse und Interpretation von Usability-Daten auf Usability-Richtlinien und Leitlinien zu fokussieren (Richter & Flückiger, 2016). Exemplarisch entsprechend der Leitlinie zur Gestaltung von Benutzungsschnittstellen nach DIN EN ISO 9241 (Tabelle 1) sollten demnach nebst den klassischen Usability-Prinzipien folgende Kriterien erfüllt sein und bei Produktevaluationen während UXR beachtet werden (vgl. Jacobsen & Meyer, 2019):

Tabelle 1

Kriterien zur Beurteilung von Benutzungsschnittstellen nach ISO 9241 für eine nutzerfreundliche Anwendung (Jacobsen & Meyer, 2019)

Benutzungsschnittstellen	Erklärung
Aufgabe angemessen	Die Anwendung erfüllt die erwarteten Funktionen, unterstützt effizient bei der Zielerreichung und nutzt nur die für den Nutzungskontext relevante Technik.
Selbstbeschreibend	Die Anwendung macht jederzeit ersichtlich, wie Nutzer*innen ihr Ziel erreichen und was sie dafür tun sollen. Navigation und Hinweise sind klar und intuitiv.
Steuerbar	Nutzer*innen behalten die Kontrolle: Prozesse können abgebrochen, wiederholt oder angepasst werden (z. B. Lautstärke, Animationen, Zurück-Funktion).
Erwartungskonform	Die Anwendung entspricht gelernten Konventionen und bleibt konsistent, sodass keine überraschenden Interaktionen entstehen (z. B. eindeutig formulierte und unterstrichene Links).
Fehlertolerant	Das System geht verständlich mit Fehlern um, gibt klare Rückmeldungen und hält den Korrekturaufwand für Nutzer*innen gering.
Individualisierbar	Einstellungen lassen sich an Vorwissen oder Präferenzen anpassen; Personalisierung erleichtert wiederkehrende Nutzung.
Lernförderlich	Die Anwendung unterstützt Nutzer*innen beim schrittweisen Erlernen der Bedienung, etwa durch Hinweise oder sichtbare Tastaturkürzel.

2.2 Künstliche Intelligenz

Im Jahr 1997 schlug der Computer *Deep Blue* den Schachweltmeister *Garry Kasparov*. Damit gelang erstmals einem Computersystem der Sieg über einen Menschen in einer komplexen strategischen Denksportaufgabe wie dem Schach (Paass & Hecker, 2020). Der Grundstein für diesen Triumph wurde jedoch bereits Jahrzehnte zuvor gelegt. In den 1950er-Jahren entstand mit der Entwicklung der ersten programmierbaren Computer, bereits die Idee, dass diese Geräte eines Tages intelligentes Verhalten zeigen könnten. Der Begriff der *Artificial Intelligence* (dt. Künstliche Intelligenz) wurde 1956 von J. McCarthy und M. Minsky eingeführt. In der Folge entstanden die ersten Durchbrüche im Bereich neuronaler Netze, die trainiert wurden und einfache Muster erkennen konnten. Damit wurden bereits in den 1950er- und in den späteren 1980er-Jahren die konzeptionellen Grundlagen heutiger KI-Modelle gelegt. Zwar waren die möglichen Funktionsweisen und mathematischen Modelle definiert und bekannt, jedoch fehlte damals noch die notwendige Rechenleistung sowie entsprechend leistungsfähige Grafikkarten, um ihr Potenzial anwenden zu können (Paass & Hecker, 2020).

Humm, Buxmann und Schmidt (2022) definieren KI als «die Fähigkeit von Computersystemen, Aufgaben auszuführen, die normalerweise menschliche Intelligenz erfordern» (S. 15). In diesem Verständnis simulieren bzw. imitieren KI-Systeme menschliche Fähigkeiten wie Wahrnehmen, Lernen, Wissenserwerb, Denken, Kommunizieren und können Aktionen ausführen, um komplexe Probleme automatisiert zu lösen (Humm et al., 2022; Li, Wang, Zhang & Klein, 2024). Während einige Autorinnen und Autoren ebenfalls hervorheben, dass sich KI-Systeme ähnlich wie ein Mensch intelligent verhalten und selbstständig lernen sollen, wird zugleich betont, dass der Begriff *Intelligenz* schwer klar abzugrenzen ist (Paass & Hecker, 2020). KI-Systeme gelten daher nicht als wirklich intelligent, sondern vielmehr als Werkzeuge, die anspruchsvolle Aufgaben innerhalb ihrer technischen Grenzen effizient bearbeiten können und dabei menschliche Fähigkeiten imitieren (Humm et al., 2022). So verwendet KI-Software beispielsweise häufig natürliche Sprache als Eingabe und erzeugt wiederum natürliche Sprache für die Ausgabe, sodass sie in der Lage ist, mit Nutzer*innen in einen interaktiven Austausch zu treten und Aufgaben zu lösen (Zheng, Tang, Liu, Liu & Huang, 2022).

Einige Autorinnen und Autoren weisen darauf hin, dass der Begriff KI aktuell häufig in sehr allgemeinen Formulierungen bzw. Definitionen diskutiert wird und sich zunehmend, als Marketing-Schlagwort etabliert hat. Dabei wird KI eher als Sammelbegriff genutzt, dem unterschiedliche Bedeutungen zugeschrieben werden und der häufig deutlich vereinfacht dargestellt wird. Tatsächlich umfasst KI eine Vielzahl unterschiedlicher Technologien, die für verschiedenste Zwecke eingesetzt werden können (vgl. Fischer, 2023; Paass & Hecker, 2020; Zhuhadar & Lytras, 2023). In Abbildung 4 werden die verschiedenen Ebenen dargestellt, die im Zusammenhang mit KI relevant sind und im Folgenden kurz erläutert werden. Der Begriff KI wurde im vorherigen Abschnitt bereits näher ausgeführt.

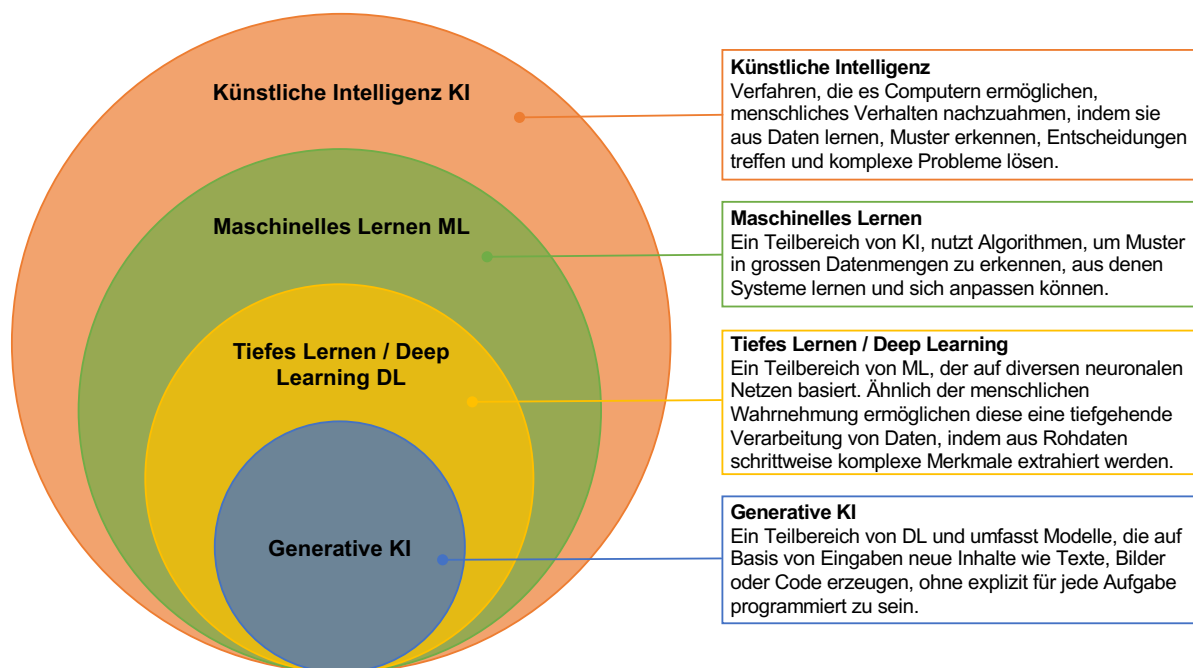


Abbildung 4. Künstliche Intelligenz und verwandte Teilbereiche (in Anlehnung an Zhuhadar & Lytras, 2023, S.4, eigene Darstellung)

Machine Learning (ML) (dt. Maschinelles Lernen) bezeichnet ein Teilgebiet von KI, welches sich auf Algorithmen konzentriert, die es Computern ermöglichen, aus Datensets zu lernen und Muster zu erkennen, ohne explizit dafür programmiert zu werden. ML wird häufig irrtümlicherweise mit KI gleichgesetzt, stellt jedoch lediglich einen spezifischen Ansatz innerhalb der KI-Forschung dar (Brandao, 2025; Humm et al., 2022; Zhuhadar & Lytras, 2023).

Deep Learning (DL) (dt. tiefes Lernen) ist wiederum ein Teilgebiet des maschinellen Lernens. Vereinfacht beschrieben basiert DL auf künstlichen neuronalen Netzen, die sich an biologischen neuronalen Strukturen orientieren und es Systemen ermöglichen, komplexe Muster zu erkennen. Dazu nutzt dieses Verfahren grosse Datenmengen sowie statistische Wahrscheinlichkeitsmodelle und lernt aus Erfahrung, insbesondere aus unstrukturierten Trainingsdaten wie Bildern, Audio oder Text. Dadurch ist DL in einer Vielzahl von Anwendungsbereichen einsetzbar, etwa in der Bild- und Spracherkennung sowie in der Verarbeitung natürlicher Sprache (Brandao, 2025; Neha, Bhati, Shukla & Amiruzzaman, 2024; Zhuhadar & Lytras, 2023).

Generative KI ist eine Untergruppe von DL-Algorithmen, baut auf den vorherigen Ebenen auf und umfasst Modelle, die in der Lage sind, neue Inhalte (z. B. Texte, Bilder oder Audio) sowie andere Medienformen zu erzeugen. Während generative KI eine breite Palette von Modelltypen einschliesst, konzentrieren sich *Large Language Models* (LLMs) (dt. grosse Sprachmodelle) spezifisch auf das Verstehen und Generieren menschlicher Sprache (Fischer, 2023). Ziel generativer KI ist es dabei, menschliche Interaktion nachzubilden. LLMs wie das KI-Modell ChatGPT basieren auf einer Kombination aus ML- und DL-Algorithmen und sind in der Lage, autonom neue Inhalte zu erzeugen. Sie zeichnen sich dadurch aus, kontextbezogene Sprache zu verarbeiten und zu

generieren, sodass sprachliche Ausgaben entstehen, die menschlicher Kommunikation ähneln (Sarker, 2024; Zhuhadar & Lytras, 2023). Im Rahmen dieser Arbeit wird die Analyse von Usability-Tests primär mithilfe von Software-Tools untersucht, die auf LLMs basieren.

2.2.1 Funktionsweise von LLMs

LLMs verarbeiten und generieren natürliche Sprache und sind dem Forschungs- und Anwendungsbereich des Natural Language Processing (NLP) (dt. Verarbeitung natürlicher Sprache) zuzuordnen. Das bedeutet beispielsweise, dass für die Interaktion mit solchen Systemen grundsätzlich keine speziellen technischen oder entsprechende Programmierfähigkeiten seitens der Nutzer*innen erforderlich sind. Stattdessen ermöglichen LLMs eine direkte und dialogorientierte Interaktion in menschlicher Sprache (Liu, 2025; Zheng et al., 2022). Seit der Veröffentlichung von ChatGPT durch das Unternehmen OpenAI, als webbasiertem Chatbot haben LLMs besondere Aufmerksamkeit erfahren (vgl. OpenAI, 2025b; Scott & Zuccon, 2024).

Der Begriff *GPT* steht dabei für *Generative Pre-trained Transformer* und verweist auf die zugrunde liegende Modellarchitektur. Es handelt sich dabei um eine Form tiefer neuronaler Netze, die darauf ausgelegt sind, Verbindungen und Abhängigkeiten zwischen Elementen einer Sequenz (bspw. Wörtern in einem Satz) zu erfassen. Die Texteingaben werden vor der Verarbeitung in sogenannte *Tokens* zerlegt, wobei ein Token im Durchschnitt etwa vier Zeichen umfasst. LLMs basieren auf probabilistischen Modellen, die auf Basis der erlernten Verbindungen und Abhängigkeiten zwischen den einzelnen Elementen versuchen, das jeweils nächste Token bei der Ausgabe und Interpretation von Sprache vorherzusagen. Dieser Mechanismus ermöglicht es LLMs, relevante sprachliche Zusammenhänge zu identifizieren und unwesentliche Informationen zu vernachlässigen. Dazu werden die Modelle gezielt mit Daten trainiert, die Milliarden, bis Billionen von Wörtern und Texten umfassen können. Auf deren Grundlage werden dann die Modellparameter optimiert, welche genutzt werden, um die sprachlichen Muster und Abhängigkeiten statistisch vorherzusagen und daraus Inhalte zu generieren (Bender, Gebru, McMillan-Major & Shmitchell, 2021; Brynjolfsson, Li & Raymond, 2025; Sarker, 2024; Scott & Zuccon, 2024).

Über das einfache Verstehen und Generieren von Sprache hinaus werden LLMs heute in einer Vielzahl von Anwendungsfeldern eingesetzt (vgl. Brandao, 2025; Li et al., 2024; Rane, Choudhary & Rane, 2024; Sarker, 2024; Scott & Zuccon, 2024; Zhuhadar & Lytras, 2023). Durch ihr zunehmendes Verständnis von Sprachstruktur, Semantik und Kontext erweitern LLMs kontinuierlich die Grenzen maschineller Sprachverarbeitung (Sarker, 2024). Dazu zählen unter anderem Textzusammenfassungen, Übersetzungen sowie weitere Aufgaben der Sprachverarbeitung wie bspw. die Transkription, wobei diese Aufzählung nicht abschliessend ist (Sarker, 2024; Zhuhadar & Lytras, 2023).

2.2.2 Potenziale des KI-Einsatzes

Die Integration von KI, insbesondere von LLMs, in unterschiedlichste Arbeitsprozesse wird in der Literatur mit diversen Potenzialen in Verbindung gebracht (vgl. Brynjolfsson et al., 2025; Najana et al., 2024; Zhuhadar & Lytras, 2023). Beschrieben wird ein breites Spektrum an Vorteilen und Einsatzmöglichkeiten, das von einer operativen Ebene bis hin zu strategischen Innovationsprozessen und Wettbewerbsvorteilen reicht (Brandao, 2025).

Ein in der Literatur häufig thematisierter Aspekt betrifft dabei die potenzielle Steigerung der Produktivität und Effizienz. So können bspw. durch die Automatisierung repetitiver Tätigkeiten Arbeitsprozesse beschleunigt werden (Brandao, 2025; Humm et al., 2022; Marti, Liu, Kour, Bilgihan & Xu, 2024; Najana et al., 2024; Sarker, 2024). Insgesamt kann der KI-Einsatz Organisationen dabei unterstützen, in einem zunehmend digitalen und dynamischen Umfeld wettbewerbsfähig zu bleiben, indem durch die KI-Integration Prozesse effizienter, schneller und adaptiver gestaltet werden (Najana et al., 2024). Empirische Hinweise darauf liefert unter anderem eine Untersuchung von Brynjolfsson et al. (2025), welche am Beispiel der KI-gestützten Arbeit im Kundenservice von einer Produktivitätssteigerung von rund 15 % berichten. Neben der erhöhten Produktivität kann der Einsatz von KI zur Sicherung der Ergebnisqualität beitragen. KI-Systeme können aufgrund ihrer Fähigkeit, grosse Datenmengen konsistent und präzise zu verarbeiten, dazu beitragen, dass Forschungsfragen unter bestimmten Bedingungen detaillierter und präziser analysiert werden können. Zudem lassen sich dadurch Muster und Ergebnisse erkennen, die ohne den Einsatz von KI möglicherweise unentdeckt geblieben wären (Bano, Zowghi & Whittle, 2024; Brandao, 2025; Humm et al., 2022).

Im Bereich Design und Entwicklung wird KI darüber hinaus als ein Treiber für Innovationsprozesse beschrieben. Trotz weiter bestehender Unterschiede zur menschlichen Kreativität können insbesondere generative KI-Systeme neue Perspektiven eröffnen und kreative Denkanstösse liefern (Zhuhadar & Lytras, 2023). Gemäss Brynjolfsson et al. (2025) könnte der Einsatz von KI durch die Reduktion von Stress zu einer Entlastung der Mitarbeitenden beitragen, was wiederum positive Effekte auf die Arbeitszufriedenheit haben dürfte. Auf organisationaler Ebene ergeben sich daraus nicht nur potenzielle Wettbewerbsvorteile, sondern auch Möglichkeiten zur Reduktion der Arbeitsbelastung von Mitarbeitenden (vgl. Najana et al., 2024).

Trotz der genannten Vorteile treten diese nicht zwangsläufig durch den blossen Einsatz von KI-Technologien ein. Vielmehr hängen sie massgeblich von den jeweiligen Rahmenbedingungen ab, etwa von der Menge und Qualität der Daten (Bender et al., 2021; Humm et al., 2022; Paass & Hecker, 2020). Darüber hinaus sind die sorgfältige Auswahl geeigneter KI-Tools sowie die Berücksichtigung von Unterschieden zwischen verschiedenen KI-Modellen und den spezifischen Anforderungen einzelner Aufgaben entscheidend (vgl. Aitim & Abdulla, 2024; Bender et al., 2021; Brown, 2024; Brynjolfsson et al., 2025; Fischer, 2023; Li et al., 2024; Paass & Hecker, 2020).

2.2.3 Risiken und Herausforderungen des KI-Einsatzes

Neben den erläuterten Potenzialen müssen auch die systemimmanenten Restriktionen von LLMs berücksichtigt werden, deren Einsatz mit gewissen Risiken mit einhergehen kann. Da LLMs, vereinfacht ausgedrückt, auf einem wahrscheinlichkeitsbasierten Prinzip beruhen, generieren sie Texte unter anderem auf Grundlage statistischer Muster aus den Trainingsdaten. Aufgrund des fehlenden tatsächlichen inhaltlichen Verständnisses kann daher die sachliche Richtigkeit der erzeugten Ergebnisse jedoch nicht durch das System selbst gewährleistet werden (Ayyamperumal & Ge, 2024; Bano et al., 2024; Gethmann, 2022; Korteling, van de Boer-Visschedijk, Blankendaal, Boonekamp & Eikelboom, 2021).

Ein wesentliches Fehlerisiko stellt das Phänomen der sogenannten *Halluzinationen* dar. Darunter wird verstanden, dass LLMs Inhalte generieren, die zwar überzeugend und plausibel erscheinen, jedoch nicht auf den Trainingsdaten oder gesicherten faktischen Informationen beruhen (Bano et al., 2024; Farquhar, Kossen, Kuhn & Gal, 2024; Fischer, 2023; Ledger & Mancinni, 2024). In diesem Zusammenhang wird in der Praxis häufig auch von *sycophantic behavior* (dt. übermäßig zustimmendes Verhalten) bzw. vom Phänomen des *Pleasings* gesprochen. Dies beschreibt die Tendenz von KI-Systemen, im Bestreben, hilfreich, unterstützend und respektvoll zu sein, eher übertrieben zustimmende oder schmeichelnde Antworten zu geben (OpenAI, 2025c). In der Folge können Fehlinformationen entstehen, die Nutzer*innen potenziell fehlleiten oder zu Missverständnissen führen. Die Berücksichtigung dieser Limitation gilt daher als Voraussetzung für einen verantwortungsvollen und effektiven Einsatz solcher Systeme (Farquhar et al., 2024; Fischer, 2023; Ledger & Mancinni, 2024). Weiter ist die Qualität der KI-Ergebnisse unmittelbar von der zugrunde liegenden Datenbasis bzw. den Trainingsdaten abhängig (Bender et al., 2021; Brynjolfsson et al., 2025; Humm et al., 2022; Paass & Hecker, 2020). Da Trainingsdaten auf menschlicher Autorenschaft basieren, spiegeln sie deren ethische Einstellungen, Vorurteile sowie Qualitätsunterschiede im Datenmaterial wider. Dies führt dazu, dass LLMs bestehende gesellschaftliche und kulturelle Verzerrungen (engl. Biases) nicht nur reproduzieren, sondern sie auch potenziell verstärken können (Bano et al., 2024; Bender et al., 2021; Stockman, 2025).

Eine weitere zentrale Einschränkung ist die mangelnde Interpretierbarkeit bzw. Transparenz der KI-Modelle, die in der Literatur oft als Blackbox-Problematik bezeichnet wird (Ayyamperumal & Ge, 2024; Komera & Manche, 2023; Sarker, 2024). Aufgrund der intransparenten internen Prozesse von KI-Systemen, ist die Grundlage ihrer Entscheidungen für Nutzer*innen häufig nicht direkt nachvollziehbar (Brandao, 2025). Wenn Nutzer*innen die Entscheidungsbasis von KI-Systemen nicht verstehen oder nachvollziehen können, verlieren sie die Möglichkeit, diese kritisch zu hinterfragen. Dadurch leidet die Transparenz, und das Vertrauen in die KI-Ergebnisse kann beeinträchtigt werden (Brandao, 2025; Sarker, 2024). Hinzu kommt, dass die stochastische Natur von LLMs zu einer begrenzten Reproduzierbarkeit der Ergebnisse führt, da identische Eingaben unterschiedliche Ausgaben bzw. Ergebnisse erzeugen können. Dadurch wird bspw. die Evaluation der Leistungsfähigkeit von LLMs erschwert (vgl. Ayyamperumal & Ge, 2024). Zusätzlich zu diesen technischen Risiken des KI-Einsatzes rücken auch ethische Fragen und datenschutzrechtliche Bedenken in den Fokus (Ayyamperumal & Ge, 2024; Bano et al., 2024; Sarker, 2024).

Ein aktueller Entwicklungssprung zeigt sich im Übergang von passiven KI-Modellen hin zur agentischer KI (engl. Agentic AI). Im Gegensatz zu klassischen LLMs agieren agentische Systeme eigenständig, indem sie selbstständig verschiedene Prozessschritte ausführen und dabei selbst Entscheidungen treffen (Ayyamperumal & Ge, 2024). Diese Systeme erfordern nur noch minimale menschliche Eingriffe, um hochkomplexe Aufgabenstellungen autonom zu bewältigen (Goethals, Luther & Matz, 2025). Mit der zunehmenden Autonomie von KI-Systemen steigen auch die Anforderungen an die menschliche Aufsicht. Es besteht die Gefahr, dass menschliche Kompetenzen schrittweise an KI-Agenten abgegeben werden und deren Empfehlungen und Ergebnisse zunehmend unkritisch übernommen werden (Brandao, 2025; Brynjolfsson et al., 2025; Gethmann, 2022).

Auch ökologische Aspekte rücken zunehmend in den Fokus. Der Betrieb und das Training grosser KI-Modelle erfordern leistungsfähige Serverinfrastrukturen, die mit einem erheblichen Energieverbrauch verbunden sind (Bender et al., 2021). In diesem Zusammenhang werden KI-Systeme sowohl als Verursacher als auch als Nebenprodukt ökologischer Belastungen beschrieben (Brandao, 2025). Kritisch diskutiert wird dabei insbesondere die Diskrepanz zwischen den Nachhaltigkeitsversprechen der Technologiebranche und dem tatsächlichen Ressourcenverbrauch. Nach Brandao (2025) bestehen weiterhin unklare und teilweise widersprüchliche Angaben zur langfristigen Nachhaltigkeit sowie zur realen ökologischen Bilanz einer grossflächigen Implementierung von KI-Technologien.

2.2.4 Mensch und KI

Der rasante Fortschritt von KI lässt die Grenzen zwischen menschlichen und maschinellen Rollen zunehmend verschwimmen (Fragiadakis et al., 2024). Infolgedessen wird in diversen Fachdisziplinen die Frage nach der potenziellen Ersetzbarkeit des Menschen, bzw. seiner zukünftigen Rolle im Arbeitsprozess diskutiert (Brandao, 2025; Gethmann, 2022). Der Fokus richtet sich dabei weniger auf eine vollständige Substitution menschlicher Arbeit, sondern vielmehr auf die Entwicklung neuer Formen der Interaktion zwischen Mensch und KI (Brandao, 2025; Fragiadakis et al., 2024; Kondapaka, Khanra, Malik, Kagzi & Hemachandran, 2023).

Natarajan, Mathur, Sidheekh, Stammer und Kersting (2025) unterscheiden zwischen reiner Automatisierung und kollaborativen Interaktionsformen. Im Konzept des *Human-in-the-Loop* (dt. Einbindung des Menschen) nimmt der Mensch innerhalb eines automatisierten Prozesses die Rolle einer überwachenden Instanz bspw. eines Supervisors ein, indem er KI-generierte Ergebnisse prüft, korrigiert und validiert. Demgegenüber beschreibt *AI-in-the-Loop* (Einbindung der KI) eine partnerschaftliche Interaktionsform, bei der die KI den Menschen unterstützend ergänzt, die Verantwortung jedoch beim Menschen verbleibt. Die Wahl einer kollaborativen Strategie anstelle einer vollständigen Automatisierung erweist sich insbesondere dann als entscheidend, wenn der Kontext der Aufgaben nur unzureichend definiert ist oder sich nicht eindeutig festlegen lässt und daher menschliche Urteilskraft erforderlich ist (Natarajan et al., 2025; Stockman, 2025). Vor diesem Hintergrund bewerten einige Forschende eine vollständige Automatisierung kritisch, da KI-Systeme

zwar zur Beschleunigung von Arbeitsprozessen beitragen können, für die Interpretation und abschliessende Bewertung der Ergebnisse jedoch weiterhin auf den Menschen angewiesen sind, der dabei zugleich als Absicherung gegenüber potenziellen Fehlern dienen kann (Kondapaka et al., 2023; Li, 2024).

Als ein Beispiel für diese kollaborative Ausrichtung kann der Human-AI-Collaboration (HAIC) Ansatz herangezogen werden. Dieser markiert einen Wandel hin zu einer integrierten und gegenseitig vorteilhaften Zusammenarbeit, die über die blossе Dominanz des Menschen über die Maschine hinausgeht. HAIC wird dabei als eine kooperative Partnerschaft definiert, in der Menschen und KI-Systeme aktiv zusammenarbeiten, um gemeinsame Ziele zu erreichen, die sowohl individueller als auch kollektiver Natur sein können. Das Spektrum der möglichen Tätigkeiten ist dabei breit gefächert und reicht von komplexen Entscheidungsprozessen über wissensbasierte Arbeit bis hin zu kreativen Aufgaben. Ein entscheidendes Merkmal erfolgreicher HAIC ist eine dynamische Aufgabenverteilung, bei der Aufgaben auf Grundlage der jeweiligen Stärken und Schwächen von Mensch und KI zugewiesen und bei Bedarf angepasst werden können. Der Fokus liegt dabei auf der Interaktion, da der Erfolg der HAIC von einer klaren Kommunikation und regelmässigen Feedbackschleifen zwischen Mensch und Maschine abhängt. Ein kontinuierlicher Austausch ist zentral, um potenzielle Missverständnisse frühzeitig zu vermeiden und die jeweiligen komplementären Stärken beider Akteure synergetisch zu nutzen (Fragiadakis et al., 2024; Song, Zhu & Luo, 2024).

Aus wissenschaftstheoretischer Perspektive begründet Gethmann (2022) die Unverzichtbarkeit des Menschen mit dem Konzept der epistemischen Verantwortung. Demnach ist wissenschaftliche Erkenntnis untrennbar an ein handelndes Subjekt gebunden, welches Verantwortung für methodische Korrektheit und Wahrheitsansprüche übernimmt. Während KI-Systeme in der Lage sind, Korrelationen in grossen Datenmengen zu identifizieren, fehlt ihnen die Fähigkeit zur eigenständigen logischen Begründung sowie zur Einsicht in Sinn- und Bedeutungszusammenhänge, wodurch sie nicht selbst als Subjekte fungieren können. Menschliche Fähigkeiten wie Erfahrung, Kreativität, situatives Urteilsvermögen sowie Kommunikationsfähigkeit bleiben daher für den Forschungsprozess zentral und sind nicht durch algorithmische Systeme substituierbar (Gethmann, 2022; Stockman, 2025). Ergänzend weisen Korteling et al. (2021) darauf hin, dass KI-Systeme primär als funktionale Werkzeuge mit grundlegend anderen kognitiven Eigenschaften als biologische Akteure zu verstehen sind. Auch sie postulieren, dass eine erfolgreiche Zusammenarbeit erfordert, KI als *Intelligenter Assistent* oder *Partner* zu sehen (Korteling et al., 2021). Entsprechend soll die Verantwortung für die Bewertung und Nutzung der Ergebnisse beim Menschen bzw. bei den Nutzer*innen verbleiben (Brynjolfsson et al., 2025; Stockman, 2025).

Damit diese Form der Zusammenarbeit erfolgreich ist, müssen Nutzer*innen von KI über die reine Übernahme von Verantwortung hinaus fachliche und methodische Kompetenzen entwickeln:

1. **KI-Kompetenz und Medienkompetenz:** Ein grundlegendes Verständnis der Funktionsweise, Fähigkeiten und Limitationen von KI-Modellen ist nötig. Nutzer*innen müssen Aufgabenkomplexitäten aus der KI-Perspektive bewerten und sich Wechselwirkungen zwischen menschlichen, kognitiven und algorithmischen Fehlern bewusst sein. Dies ermöglicht eine reflektierte Haltung einzunehmen und weder in *blinde Akzeptanz* noch in *pauschale Ablehnung* zu verfallen (Bender et al. 2021; Korteling et al., 2021; Najana et al., 2024; Scott & Zuccon, 2024).
2. **Prompting:** Dies beschreibt, die vom Menschen gesteuerte Kommunikation mit dem Computer, bei der gezielte Aufforderungen so formuliert werden, dass die KI relevante und nützliche Ergebnisse generiert. Die Qualität der KI-Ergebnisse hängt massgeblich davon ab, präzise Instruktionen und Kontexte zu formulieren (Brown, 2024; Stockman, 2025).
3. **Kritisches und analytisches Denken:** Nutzer*innen müssen in der Lage sein, generierte KI-Inhalte kritisch zu hinterfragen und zu verstehen, um Falschinformationen zu erkennen und vorzubeugen (Brandao, 2025; Fischer, 2023). Das Verstehen und Interpretieren der von der KI generierten Ergebnisse kann sich als anspruchsvoll erweisen (Lauer, Storey & Soley, 2024).

Neben der Notwendigkeit des Erwerbs neuer Kompetenzen für den Umgang mit KI besteht das Risiko einer schleichenden Erosion der methodischen Expertise. Forschende gehen davon aus, dass eine zu starke Abhängigkeit von KI-Assistenz dazu führen könnte, dass menschliche analytische Fähigkeiten langfristig verkümmern, da komplexe kognitive Aufgaben zunehmend an die KI-Technologie delegiert werden (Brynjolfsson et al., 2025; Li et al., 2024). Anstatt menschliche Expertise zu ersetzen, sollte KI als unterstützendes Werkzeug verstanden werden, das bestehende menschliche Kompetenzen erweitert und ergänzt (Li et al., 2024).

2.3 Einsatzmöglichkeiten von KI im UXR

UX ist heutzutage ein wesentlicher Faktor für den Erfolg von Produkten und Dienstleistungen. Mit der zunehmenden Menge und Komplexität digitaler Daten stellt insbesondere die Bearbeitung und Analyse grosser Mengen unstrukturierter Daten, wie bspw. Texte, Bilder, Audio- und Videomaterial, eine wesentliche Herausforderung im UXR dar. Die Analysephase kann im UXR-Prozess zu einem Engpass werden, da es anspruchsvoll sein kann, aus umfangreichen Datensätzen zeitnah handlungsrelevante Erkenntnisse abzuleiten, welche für die kurzen Entwicklungszyklen der Produktentwicklung möglichst schnell verfügbar sein müssen (Aitim & Abdulla, 2024; Graner, 2015; Li, 2024; Liu, 2025; Manning, Zhu & Horton, 2024). Vor diesem Hintergrund gewinnt der Einsatz von KI im UXR zunehmend an Bedeutung. Die Integration KI-gestützter Methoden markiert einen Paradigmenwechsel in der Usability- und UX-Evaluation, da vormals manuelle Prozesse zunehmend automatisiert, skalierbar und effizienter gestaltet werden. Dies ist besonders in schnellen, iterativen Entwicklungsumgebungen von Relevanz (Liu, 2025; Najana et al., 2024). Fortschritte in den Bereichen NLP und DL haben diese Entwicklung zusätzlich begünstigt, indem sie neue

Möglichkeiten zur Erkennung und Analyse inhaltlich relevanter Muster und Informationen aus komplexen unstrukturierten Datensätzen ermöglichen (Aitim & Abdulla, 2024). Aktuelle Studien zeigen, dass KI entlang nahezu aller Phasen des UXR-Prozesses eingesetzt werden kann. Von der Vorbereitung und Datenerhebung über die Analyse bis hin zur Simulation von Anwendungsszenarien oder der Unterstützung des Design- und Prototyping-Prozesses (Aitim & Abdulla, 2024; Knearem et al., 2023; Li et al., 2024; Liu, 2025; Soares et al., 2025). Bisherige Forschung legte den Schwerpunkt hauptsächlich auf den Einsatz von KI in Designprozessen und der Erstellung von Inhalten, wohingegen ihr Nutzen für automatisierte Analysemethoden wie etwa Usability-Tests bislang nur unzureichend untersucht wurde (vgl. Knearem et al., 2023; Liu, 2025).

2.3.1 Planung und Datenerhebung

Bereits in frühen Phasen des UXR kann der Einsatz von KI zur Unterstützung der Planung und ersten Datenerhebungen beitragen. KI-gestützte Verfahren ermöglichen es, automatisiert erste Hinweise auf Nutzungsmuster, Bedürfnisse und Problembereiche aus bestehenden Datenquellen zu gewinnen, etwa aus Online-Reviews, App-Bewertungen oder Social-Media-Inhalten. Dabei kann KI auch als Instrument zur explorativen Marktforschung eingesetzt werden, um ein erstes Verständnis für relevante Themenfelder und Markttrends zu entwickeln (Aitim & Abdulla, 2024; Najana et al., 2024; Rane et al., 2024). Dafür können bspw. NLP-Algorithmen eingesetzt werden, um grosse Mengen textbasierter Daten systematisch auszuwerten (Li, 2024; Rane et al., 2024). Darüber hinaus kann KI etwa bei der Erstellung von Interviewleitfäden, Beobachtungsbögen oder Tagebuchstudien unterstützend verwendet werden. Die KI-Systeme können dazu Vorlagen bereitstellen, bei der Strukturierung der Forschungsinstrumente helfen und Vorschläge für Umfrageitems generieren. Zudem lassen sich simulierte Antworten für Pretests nutzen, um Fragebögen vor der Durchführung zu prüfen (Li, 2024).

2.3.2 Datenaufbereitung und Transkription

In der Phase der Datenaufbereitung kann KI in erster Linie für die Transkription und Strukturierung von grossen Datenmengen genutzt werden. Automatisierte Transkriptionswerkzeuge ermöglichen die Umwandlung von Audio- und Videodaten in eine Textform, identifizieren verschiedene Sprechende und können dadurch den manuellen Transkriptionsaufwand erheblich reduzieren (Aitim & Abdulla, 2024; Li, 2024; Li et al., 2024). Neben der Transkription können KI-Systeme in dieser Phase Aufgaben wie die Datenbereinigung, Kategorisierung und erste Strukturierung der Inhalte, etwa durch Zusammenfassungen oder die Hervorhebung relevanter Textstellen übernehmen (Penkert, 2025). NLP und ML-Verfahren können gezielt dabei unterstützen Muster in den Daten zu erkennen und relevante Themen zu identifizieren. Damit können bspw. erste Erkenntnisse für die Persona-Entwicklung oder Feedback zur Prototyperstellung generiert werden (Aitim & Abdulla, 2024; Li, 2024; Liu, 2025). Ein praktisches Beispiel hierfür liefert ein UX-Forschungsprojekt im Bereich Gesundheits-Apps. Dabei wurden Interviews mit Google Speech-to-

Text transkribiert und anschliessend mittels NLP-Modellen analysiert. Die automatisierte Analyse ermöglichte die schnelle Identifikation von wiederkehrenden Themen und Usability-Problemen, wie Navigationsschwierigkeiten oder Unklarheiten in der App, und reduzierte den manuellen Aufwand signifikant (Aitim & Abdulla, 2024).

2.3.3 Analyse und Interpretation

Einen Schritt weiter, in der Analysephase von UXR werden die zuvor aufbereiteten und transkribierten Daten inhaltlich ausgewertet. KI-gestützte Werkzeuge bieten dafür vielfältige Unterstützungsmöglichkeiten, welche die Effizienz und Aussagekraft der Analyse erhöhen können (Aitim & Abdulla, 2024; Li, 2024; Norman, 2013). Sprachbasierte KI-Modelle auf Grundlage von NLP, einschliesslich LLMs, ermöglichen die automatisierte Analyse qualitativer UXR-Daten wie etwa von Thinking-Aloud-Transkripten oder Interviewdaten. Dabei können Transkripte zusammengefasst, wiederkehrende Themen und Muster identifiziert und Inhalte nach Stimmungen oder spezifischen Problembereiche kategorisiert werden (Penkert, 2025; Soares et al., 2025; Liu, 2025).

Viele KI-Tools konzentrieren sich bislang vor allem auf die Auswertung textbasierter Daten (Soares et al., 2025). Neben Text können jedoch auch weitere Datenformen mithilfe von KI analysiert werden. Digitale Inhalte bestehen heutzutage häufig aus einer Kombination von Text, Audio, Bildern und Videos und sind damit multimodal (Paass & Hecker, 2020). Insbesondere Algorithmen und Methoden aus dem Forschungsbereich der Computer Vision, der ebenfalls dem Feld der KI zuzuordnen ist, erlauben es, Informationen aus Bildern und Videos automatisch zu erfassen und auszuwerten. Dazu gehören nonverbale Signale wie Blickbewegungen, Gesichtsausdrücke, emotionale Reaktionen oder Interaktionen während der Nutzung eines Produktes (Liu, 2025; Humm et al., 2022). Bildbasierte KI-Systeme kommen bereits in anderen Anwendungsfeldern zum Einsatz, etwa in der Medizin zur automatischen Erkennung von Krankheiten oder zur Gesichtserkennung in Kameras bspw. an Flughäfen (Humm et al., 2022). Der Vorteil einer Kombination unterschiedlicher Datenquellen liegt darin, eine umfassendere und kontextsensitivere UX-Analyse zu ermöglichen, die bislang mit einem hohen manuellen Aufwand durch die UX-Forscher*innen verbunden war (Soares et al., 2025; Zaghdoudi & Glomann, 2021).

Im Forschungskontext existieren bislang nur wenige konkrete Beispiele für KI-Tools, die tatsächlich multimodale Daten im UXR-Kontext systematisch auswerten können. Zwar kommen in der UX-Praxis visuelle und physiologische Analysemethoden wie Eye-Tracking (dt. Blickbewegungsanalyse) oder Hautleitwiderstandsmessung zur Aktivierungsmessung zum Einsatz, nach aktuellem Kenntnisstand sind jedoch keine etablierten KI-Tools verfügbar, welche die Kombination unterschiedlicher Datenquellen wie etwa Videodaten aus Usability-Tests inkl. der Audiospur umfassend und automatisiert analysieren können (Jacobsen & Meyer, 2019).

Obschon die grundsätzliche Machbarkeit multimodaler Ansätze in der Literatur beschrieben wird, beschränkt sich diese jedoch weitgehend auf konzeptionelle Arbeiten und experimentelle Prototypen (Li, 2024; Liu, 2025; OpenAI, 2025a; Zaghdoudi & Glomann, 2021; Zhang et al., 2024).

Kommerzielle, ausgereifte Produkte zur multimodalen Analyse von Usability-Tests, etwa durch die kombinierte Auswertung von Video-, Audio- und Textdaten, sind gegenwärtig kaum verfügbar. Ein exemplarischer Forschungsprototyp ist das System *DesignWatch* aus China. Diese Software kann unterstützt bei der Analyse von Interaktionen in mobilen Anwendungen verwendet werden und basiert Bildschirmaufzeichnungen des Tests. Diese Software ermöglicht eine visuelle Analyse von Nutzungspfaden und nutzt multimodale LLMs, um mögliche Gedanken der Nutzer*innen während einzelner Interaktionen zu simulieren (Zhang et al., 2024). Vergleichbare Systeme verdeutlichen das Potenzial KI-gestützter Methoden zur Analyse von Nutzerinteraktionen und Nutzungserfahrung (vgl. Gardey, Grigera, Rodriguez & Garrido, 2024; Liu, 2025).

2.3.4 Generative KI und synthetische Personas

Durch die Integration von KI in den UXR-Prozess können sich auch Formen der Datenerhebung verändern. Teile des Prozesses können dabei durch künstliche, auf LLMs basierende Systeme unterstützt oder voll automatisiert und somit vollständig simuliert werden (Lauer et al., 2024; Liu, 2025). Manning et al. (2024) vertreten die Auffassung, dass solche Ansätze unter bestimmten Bedingungen als synthetische Nutzer*innen in künstlichen Experimenten eingesetzt werden können. Dies setzt jedoch unter anderem voraus, dass sich die daraus gewonnenen Erkenntnisse auf reale Nutzer*innen übertragen lassen. Daraus entstehen neue Möglichkeiten, bspw. bei der Erstellung von Personas oder bei Interviews und Befragungen mit synthetischen Nutzerprofilen (vgl. Goethals et al., 2025). Personas stellen im UX-Design ein wichtiges Instrument dar, um Nutzergruppen mit spezifischen Bedürfnissen, Zielen und Verhaltensweisen prototypisch abzubilden (Jacobsen & Meyer, 2019). Ihre Qualität hängt dabei massgebend von der zugrunde liegenden Datengrundlage ab. Lauer et al. (2024) weisen darauf hin, dass Personas, welche auf unzureichenden Daten basieren, eher Probleme verursachen können, als dass sie Nutzen stiften. Durch die Kombination und Analyse unterschiedlicher sowie grosser Datenquellen, kann generative KI den Entwicklungsprozess von Personas dennoch erheblich beschleunigen. Da die KI über kein umfassendes Verständnis der realen Nutzungskontexte verfügt ist die Zuverlässigkeit der Ergebnisse jedoch eingeschränkt (Lauer et al., 2024).

Penkert (2025) geht davon aus, dass sich klassische UX-Forschungsmethoden wie Umfragen und Interviews zukünftig verändern werden, da sie bisher mit einem hohen manuellen und zeitlichen Aufwand in der Datenerhebung und -auswertung verbunden waren. KI-moderierte Interviews stellen in diesem Zusammenhang eine vielversprechende Lösung dar, um mit minimalem Aufwand präzise und strukturierte Erkenntnisse über Nutzer*innen zu gewinnen. Solche KI-Tools ermöglichen es, die erhobenen Daten direkt zu transkribieren, zusammenzufassen und zu analysieren. In der Literatur wird zudem der Einsatz sogenannter agentischer KI-Systeme im UXR-Kontext diskutiert. Solche Systeme können durch die Analyse grosser Datenmengen Vorhersagen über Bedürfnisse und Nutzungsverhalten treffen, ohne dass ein direkter Kontakt zu realen Nutzer*innen erforderlich ist. Auf dieser Grundlage lassen sich Interviews mit synthetischen Nutzer*innen vollständig automatisiert durchführen und auswerten. Unternehmen können dadurch die Bedürfnisse und Nutzungsmuster

verschiedener Nutzergruppen künstlich simulieren, bspw. in Situationen, in denen der Zugang zu echten Nutzer*innen eingeschränkt ist (Ayyamperumal & Ge, 2024; Goethals et al., 2025; Liu, 2025).

2.3.5 Grenzen des KI-Einsatzes

Die Implementierung von KI-Systemen kann für KMU mit verschiedenen Hürden verbunden sein. Neben begrenzten finanziellen und personellen Ressourcen zählen insbesondere hohe Lernkurven infolge notwendiger Weiterbildungen der Mitarbeitenden sowie der erforderliche zeitliche Aufwand für die Datenaufbereitung zu den zentralen Herausforderungen bei der Integration von KI (Li et al., 2024; Wallraff et al., 2023). Auf soziotechnischer Ebene können darüber hinaus interne Widerstände innerhalb der Belegschaft sowie Bedenken der Kundschaft in Bezug auf Datenschutz und digitale Ethik die Integration von KI-Technologien beeinträchtigen (Li et al., 2024; Wallraff et al., 2023).

Trotz der beschriebenen technologischen Fortschritte bleibt die Erhebung realer Nutzerdaten für die Validität der UX-Forschung zentral und unverzichtbar. Eine ausschliessliche Orientierung an synthetischen Nutzerprofilen wird in der Literatur kritisch betrachtet, da daraus das Risiko entsteht, dass simulierte Erkenntnisse die tatsächlichen menschlichen Erfahrungen verzerren oder ungenügend abbilden (Lauer et al., 2024; Li, 2024). Eine weitere Hürde für kleinere UX-Agenturen können die hohen technischen und rechnerischen Anforderungen grosser KI-Modelle darstellen, die erhebliche Ressourcen und spezialisierte Hardware erfordern (Liu, 2025). Dies kann gerade dann relevant werden, wenn KI-Systeme nicht über externe Dienste, sondern lokal betrieben werden sollen. Für die Zukunft wird eine Entwicklung hin zu spezialisierten, auf spezifisch ausgewählten Trainingsdaten basierenden LLMs erwartet, um methodische Verzerrungen und Fehlinformationen von grossen allgemeinen Modellen weiter zu minimieren und weniger Ressourcen benötigen dürften (Liu, 2025; Scott & Zuccon, 2024). Generativer KI fehlen zudem echte Empathie sowie die Fähigkeit, implizite emotionale Reaktionen und authentische Nutzungserfahrungen vollständig abzubilden und tatsächlich zu verstehen (Li, 2024).

Die Integration von KI in den UXR-Prozess setzt ein Bewusstsein für die Stärken und Schwächen der KI-Modelle sowie deren gezielte Verknüpfung mit menschlichen Kompetenzen voraus. Gleichzeitig bleiben menschliches Urteilsvermögen und Kontrolle besonders bei Aufgaben, die Empathie und kontextuelles Verständnis erfordern, unverzichtbar (Brynjolfsson et al., 2025; Gethmann, 2022; Lauer et al., 2024; Li et al., 2024).

3 Methodik

Im folgenden Kapitel wird die Methodik der vorliegenden Arbeit begründet und beschrieben. Zunächst werden das zugrunde liegende Forschungsdesign vorgestellt sowie die Grundlagen der Datenerhebung durch Experteninterviews und Literaturrecherche dargelegt. Darüber hinaus wird der Einsatz des Forschungstagebuchs als ergänzendes Instrument zur Reflexion und Dokumentation des Forschungsprozesses beschrieben. Anschliessend werden die Transkription und Auswertung der erhobenen Daten sowie die in dieser Arbeit behandelte Fallstudie näher erläutert. Darauf aufbauend wird abschliessend das Vorgehen des Mensch-KI-Vergleichs dargestellt und erklärt.

3.1 Forschungsdesign

Bisher gibt es nur wenige umfassende und differenzierte Forschungsarbeiten zum Einsatz von KI im Bereich des UXR. Um neue Einsichten und ein vertieftes Verständnis dieses noch jungen und dynamischen Forschungsfeldes zu gewinnen, ist eine möglichst grosse Offenheit gegenüber dem Untersuchungsgegenstand erforderlich. Aus diesem Grund und in Anlehnung an die Forschungsfrage wurde ein exploratives, qualitatives Vorgehen gewählt. Dieses zeichnet sich durch Offenheit und Flexibilität im Umgang mit dem Forschungsfeld aus, was entscheidend ist, um unvoreingenommen neue Perspektiven und Erkenntnisse zu entwickeln (vgl. Flick, 2019). Zudem ermöglicht diese offene Herangehensweise, die Integration und Auswirkungen von KI im Analyseprozess in ihrer gesamten Komplexität zu erfassen sowie bislang möglicherweise unbeachtete Aspekte sichtbar zu machen.

Zur Datenerhebung wurden im Rahmen der Arbeit mehrere methodische Vorgehensweisen kombiniert. Der Ablauf des Forschungsprozesses wird im Folgenden in drei Phasen beschrieben. Diese sind in Abbildung 5 schematisch dargestellt und werden ergänzend im Text erläutert. Die Begründung der Auswahl und Anwendung der einzelnen Methoden erfolgt detailliert in den nachfolgenden Kapiteln zu den jeweiligen Methoden.

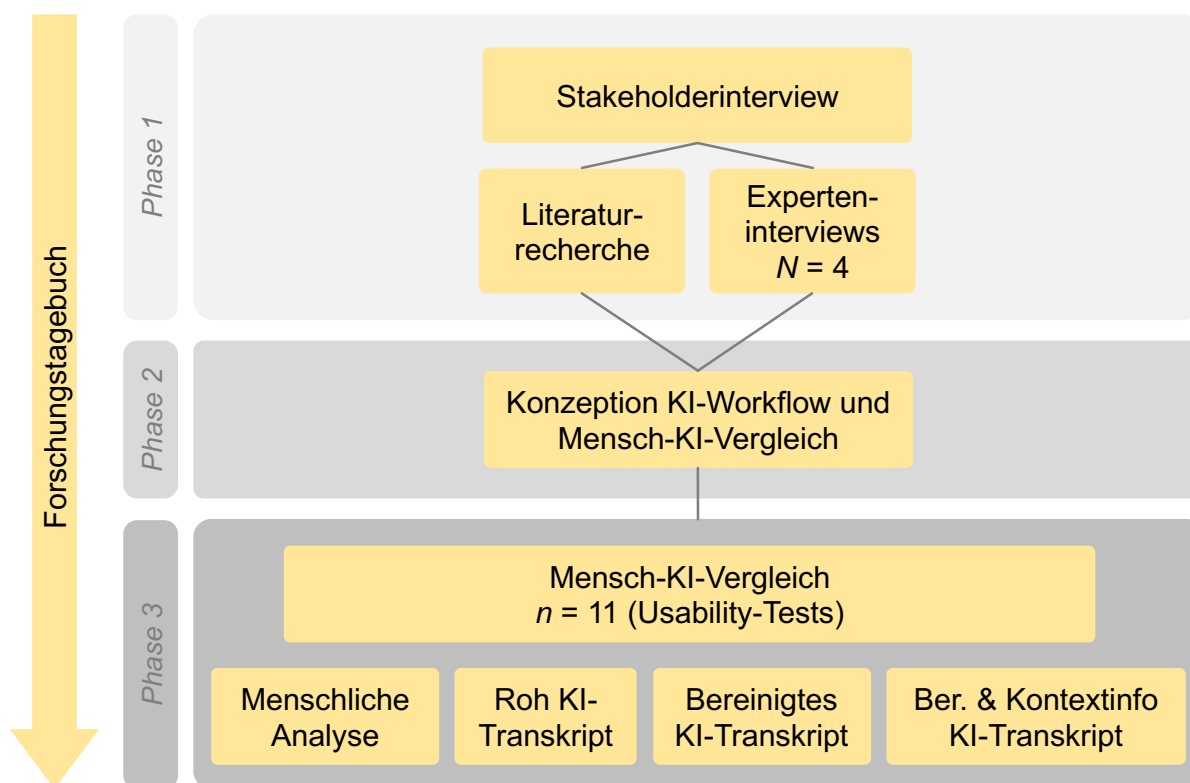


Abbildung 5. Forschungsdesign: Mensch und KI im UXR (eigene Darstellung)

Phase 1: Datenerhebung

Zu Beginn wurde eine Literaturrecherche durchgeführt, um den aktuellen Forschungsstand sowie mögliche Umsetzungen des KI-gestützten Analyseprozesses zu erfassen. Die gewonnenen Erkenntnisse dienen sowohl der thematischen Einarbeitung ins Forschungsfeld als auch als Grundlage für die Erstellung der Leitfäden für die Experteninterviews. Ausserdem wurde ein Stakeholderinterview mit dem Geschäftsführer der Praxispartnerin geführt, um Prozesse, Arbeitsweisen sowie die Motivation und Erwartungen im Zusammenhang mit der Implementierung von KI im UXR des Unternehmens zu erfassen. Zusätzlich zu diesem ersten Schritt wurden halbstrukturierte Experteninterviews mit Fachpersonen aus den Bereichen UXR und KI bzw. Informatik geführt. Ziel war es, Einblicke in das Forschungsfeld zu gewinnen, Einsatzmöglichkeiten von KI im UXR zu identifizieren sowie Potenziale, Grenzen und potenzielle KI-Tools für den KI-gestützten Analyseprozess zu erfassen. Zudem sollten relevante Kriterien und Ansätze zur Evaluation des Vergleichs zwischen menschlicher und KI-gestützter Analyse ermittelt werden.

Phase 2: Auswertung und Konzeption

Basierend auf den Ergebnissen der Literaturrecherche und der Experteninterviews wurde ein auf die Fallstudie der sinnhaft GmbH angepasster KI-gestützter Analyseprozess entwickelt (Fokus auf Transkription / Analyse / Auswertung von Usability-Tests). Zur Sicherstellung der Nachvollziehbarkeit und zur späteren Auswertung wurden wichtige Erkenntnisse, Beobachtungen

sowie persönliche Reflexionen zu den einzelnen Schritten, insbesondere zum Prompting sowie auftretenden Herausforderungen, fortlaufend im Forschungstagebuch dokumentiert.

Phase 3: Anwendung & Validierung

Im dritten Schritt wurden bestehende Usability-Test-Daten aus einem vorherigen Projekt der Praxispartnerin herangezogen und mithilfe des entwickelten KI-gestützten Analyseprozesses ausgewertet. Die durch die KI gewonnenen Ergebnisse wurden anschliessend im Rahmen eines Mensch-KI-Vergleichs mit den menschlichen Auswertungen verglichen.

3.2 Datenerhebung

Im Kapitel Datenerhebung werden die eingesetzten Methoden der Datenerhebung beschrieben, darunter Stakeholderinterviews, die Entwicklung der Leitfäden für die Experteninterviews, deren Durchführung sowie die Beschreibung der Stichprobe. Darüber hinaus wird das Forschungstagebuch vorgestellt, welches zur Dokumentation des Forschungsprozesses diente.

3.2.1 Stakeholderinterview

Der Begriff des Stakeholderinterviews stammt ursprünglich aus dem Stakeholder- und Projektmanagement und wird in dieser Arbeit als Form des Experteninterviews eingesetzt, um organisationsspezifisches Wissen der Praxispartnerin zu erschliessen und die Fallstudie kontextuell einzuordnen (vgl. Flick, 2019; Meuser & Nagel, 2009). Ziel des Interviews war es, Kriterien für die Nutzbarkeit sowie Möglichkeiten für die Integration des KI-gestützten Analyseprozesses zu identifizieren. Zu diesem Zweck wurde ein Stakeholderinterview mit dem Geschäftsführer der Praxispartnerin sinnhaft GmbH durchgeführt. Im Interview wurden interne Prozesse, Arbeits- und Vorgehensweisen der Praxispartnerin thematisiert sowie Motive, Erwartungen und potenzielle Herausforderungen im Zusammenhang mit der Implementierung von KI in bestehende oder neu gestaltete UXR-Prozesse erhoben.

3.2.2 Literaturrecherche

Zur Erarbeitung der theoretischen Grundlagen sowie zur Vorbereitung der Experteninterviews und des Mensch-KI-Vergleichs wurde zwischen Mitte März und Ende Dezember 2025 eine Literaturrecherche durchgeführt. Ziel war es, einen Überblick über den aktuellen Forschungsstand zu KI im UXR, zu möglichen Einsatzgebieten innerhalb von UX-Analyseprozessen sowie zu damit verbundenen Vorteilen und Herausforderungen zu gewinnen. Ausserdem sollten bestehende Forschungslücken identifiziert werden. Die Ergebnisse der Literaturrecherche dienten der theoretischen Einbettung der Arbeit sowie als Grundlage für die Konzeption der Leitfäden der Experteninterviews und die Entwicklung der KI-gestützten Analyse. Die Literaturrecherche erfolgte in

wissenschaftlichen Datenbanken und Fachportalen, darunter Google Scholar, ACM Digital Library, IEEE Xplore, SpringerLink, Swisscovery und Elsevier (ScienceDirect). Ergänzend wurde arXiv genutzt, um im dynamischen und sich rasch entwickelnden Forschungsfeld auch aktuelle Fachbeiträge und Preprints bzw. Vorabveröffentlichungen im Bereich KI zu berücksichtigen. Für die Recherche wurden unter anderem folgende Suchbegriffe in deutscher und englischer Sprache verwendet: User Experience Research, Usability-Methoden, Künstliche Intelligenz, Large Language Models, KI-gestützte Analyse, qualitative Analyse und KI sowie Mensch-KI-Zusammenarbeit. Die Suchbegriffe wurden jeweils auch mit entsprechenden englischsprachigen Begriffen zur Suche in den Datenbanken angewandt.

3.2.3 Leitfadenerstellung

Zur Sicherstellung der Qualität und Vergleichbarkeit der Interviewdaten sowie dem Prinzip der Offenheit wurde ein halbstrukturierter Leitfaden entwickelt (Helfferich, 2011, 2022). Dies ermöglichte Flexibilität bezüglich Reihenfolge, Formulierung und Tiefe der Fragen und ebenfalls die situative Anpassung und entsprechendes Eingehen auf die Expert*innen und deren komplexen Wissensstand (Flick, 2019). Grundlage dafür bildeten die theoretischen Erkenntnisse der Literaturrecherche.

Die Erstellung des Leitfadens erfolgte angelehnt an das SPSS-Prinzip (Sammeln, Prüfen, Sortieren, Subsumieren) gemäss Helfferich (2022). Dieses Vorgehen kombiniert die Offenheit gegenüber dem Untersuchungsgegenstand mit einer gleichzeitig übersichtlichen thematischen Strukturierung. Dazu wurden zunächst alle potenziell relevanten Fragen gesammelt und anschliessend hinsichtlich ihrer inhaltlichen Passung zum Forschungsgegenstand geprüft. Im nächsten Schritt wurden ähnliche oder überlappende Fragen zusammengeführt, sodass ein verdichteter Fragenkatalog entstand. Im letzten Schritt wurden diese Fragen thematisch geordnet, woraus schliesslich übergeordnete Themenfelder gebildet wurden. Für jedes Themenfeld wurden eine offene einleitende Frage sowie weitere vertiefende Unterfragen formuliert, um eine konsistente, aber dennoch flexible Gesprächsführung zu ermöglichen. Für die Experteninterviews wurden folgende Themenschwerpunkte gelegt: *Status quo: UXR und KI-Anwendungen; Analyse-Prozess: Menschliche vs. KI-basierte Analyse; Anforderungen an die Forschenden (Kompetenzen); Vergleich, Evaluationskriterien & Validierungsmöglichkeiten; Zukunftsperspektiven & Chancen und Risiken; Abschluss / Letzte Gedanken.*

Die Themenblöcke bestanden aus erzählgenerierenden Fragen sowie möglichen Vertiefungs- und Anschlussfragen. Als Einstieg in das Interview wurde eine offene und erzählgenerierende Frage zu den bisherigen Erfahrungen im UX-Bereich mit KI gestellt. Nach jedem Experteninterview wurde das Gespräch anhand der währenddessen erstellten Notizen reflektiert und ausgewertet. Auf dieser Grundlage erfolgte eine Anpassung des Leitfadens für das jeweils folgende Interview, um neu gewonnene Erkenntnisse zu integrieren und die Gesprächsführung gezielt zu optimieren. Zudem wurde der Leitfaden inhaltlich leicht an den jeweiligen fachlichen Hintergrund der Expertin bzw. des Experten angepasst, um deren spezifisches Fachwissen möglichst umfassend zu berücksichtigen.

Zur Überprüfung der Reliabilität des Leitfadens und der Verständlichkeit der Fragen wurden die ausgearbeiteten Leitfäden jeweils mit einem Probeinterview getestet (Flick, 2019). Die Leitfäden sind im Anhang C zu finden.

3.2.4 Durchführung Experteninterviews

Im nächsten Schritt wurden ergänzend zur Literaturrecherche Experteninterviews durchgeführt. Das Ziel dieser Experteninterviews war es, einen Zugang zum Forschungsfeld zu erhalten und weiter zu explorieren, was die Möglichkeiten sowie Potenziale aber auch Grenzen des KI-Einsatzes im UXR sein könnten. Gemäss Wassermann (2015) dient der Einsatz der Methode der Experteninterviews dazu, bislang wenig Bekanntes aufzudecken, weshalb sie sich besonders für explorative Fragestellungen eignet (vgl. Flick, 2019). Die Expert*innen verfügen dabei über Fachwissen im relevanten Forschungsfeld, welches nicht direkt und allgemein zugänglich ist (Meuser & Nagel, 2009).

Das Sampling der Experteninterviews orientierte sich an den Fragestellungen sowie dem Forschungsfeld der vorliegenden Arbeit. Insgesamt wurden sechs Expert*innen angefragt, von welchen vier zusagten und an den Interviews teilnahmen. Das Fachwissen der befragten Expert*innen umfasste die Bereiche UXR, HCD, Usability, qualitative Datenanalyse, KI sowie Informatik. Ziel war es, unterschiedliche Perspektiven aus Wissenschaft und Praxis sowie verschiedene Erfahrungsniveaus im Umgang mit KI im UXR zu berücksichtigen. Dadurch sollte ein breites und zugleich vertiefendes Bild der Potenziale, Herausforderungen und Einsatzmöglichkeiten von KI im UXR gewonnen werden.

Die vier durchgeführten Experteninterviews wurden im Zeitraum zwischen Juni und August 2025 durchgeführt. Mit einer Ausnahme wurden sämtliche Interviews online über Microsoft Teams oder Zoom durchgeführt. Die Ausnahme bildete das Interview EXP04, welches in schriftlicher Form erfolgte. Die Expertin aus Interview EXP04 hat zusätzlich noch Daten und Erkenntnisse aus einem eigenen Mensch-KI-Vergleich zur Verfügung gestellt. Weitere Daten zum Sampling der Expert*innen sind in (Anhang D) abgelegt. Zu Beginn der Interviews wurden die Expert*innen ausführlich über die Forschung und deren Zielsetzung informiert. Sie unterzeichneten zudem eine Einverständniserklärung, in welcher sie der Erhebung bzw. Verwendung und Aufzeichnung ihrer Daten zustimmten (Anhang E). Die Expert*innen wurden sowohl im Rahmen der Einverständniserklärung als auch mündlich über ihr Recht informiert, ihre Daten jederzeit aus der Studie zurückziehen zu können. Darüber hinaus bestand jederzeit die Möglichkeit, Fragen zu stellen.

3.2.5 Forschungstagebuch

Das Führen eines Forschungstagebuchs kann insbesondere bei Projekten mit mehreren Forschenden oder bei komplexen Forschungsdesigns von Vorteil sein. Die darin festgehaltenen Erkenntnisse lassen sich gezielt in den weiteren Forschungsprozess integrieren und tragen somit direkt zum Erkenntnisgewinn bei (Flick, 2019).

Das Forschungstagebuch diente in dieser Arbeit sowohl als Reflexionsinstrument, als auch als Datenquelle. Es begleitete den gesamten Forschungsprozess und ermöglichte eine fortlaufende Dokumentation und Selbstreflexion der einzelnen Arbeitsschritte. Ziel war es, sowohl methodische als auch persönliche Perspektiven und Entscheidungen im Forschungsprozess transparent und nachvollziehbar zu dokumentieren (vgl. Flick, 2019; Rädiker & Kuckartz, 2019). Die im Forschungstagebuch festgehaltenen Erkenntnisse fließen in den Ergebnisteil mit ein und sind jeweils entsprechend gekennzeichnet. Ein Ausschnitt ist zur Veranschaulichung im Anhang F abgelegt.

Das Forschungstagebuch wurde in Anlehnung an das Beispiel von Rädiker und Kuckartz (2019) in fünf Hauptkategorien gegliedert, welche eine strukturierte und zugleich flexible Dokumentation des Forschungsprozesses ermöglichten:

1. **Phase des Projekts und Aktivitäten:** Erfassung konkreter Arbeitsschritte, Einsatz von Tools, Methoden oder Theorien.
2. **Beobachtungen und Ergebnisse:** Was ist geschehen. Welche Ergebnisse oder Zwischenstände liegen vor.
3. **Probleme und Herausforderungen:** Technische Schwierigkeiten, methodische Unsicherheiten, Abweichungen vom Plan
4. **Lösungen und Anpassungen:** Umgang mit Schwierigkeiten, allfällige Änderungen oder Anpassungen am Forschungsdesign oder am Workflow
5. **Reflexion, Interpretation und Erkenntnisse:** Persönliche Eindrücke, Wahrnehmungen und Einschätzungen zu Vorgehensweisen, Situationen oder Ergebnissen

3.3 Transkription und Auswertung Experteninterviews

Nachfolgend wird das Vorgehen zur Transkription und Auswertung der Experteninterviews beschrieben.

3.3.1 Transkription

Das Datenmaterial wurde in einem ersten Schritt durch KI-Modelle bzw. Software (Microsoft Teams & NoScribe, Version 0.6.2) transkribiert und anschliessend noch einmal komplett manuell überarbeitet. Für die Auswertung der gewonnenen textbasierten qualitativen Daten wurde das

aufgezeichnete Material mithilfe von MAXQDA (Version 2024) transkribiert bzw. überarbeitet. Die Transkription erfolgte angelehnt an die Regeln von Kuckartz und Rädiker (2020), wobei einzelne Anpassungen vorgenommen wurden (vgl. Anhang G).

3.3.2 Qualitative Analyse

Die Interviews wurden jeweils während des Erhebungsprozesses in einem Schnellverfahren ausgewertet. Hierzu dienten Feldnotizen, die während der Experteninterviews zur raschen Dokumentation gemacht wurden und eine erste Überblicksanalyse ermöglichten, um die Leitfäden für die folgenden Interviews anzupassen und weiterzuentwickeln. Das Schnellverfahren umfasste die Analyse der Notizen sowie eine kurze Zusammenfassung der ersten Eindrücke und Erkenntnisse unmittelbar nach dem Interview (Flick, 2019). Bei Bedarf konnten das aufgezeichnete Audio und die erste Version der automatischen Transkription herangezogen werden, um wichtige Details zu überprüfen.

In einem weiteren Schritt wurden die Experteninterviews systematisch und strukturiert ausgewertet, um sowohl die Forschungsfragen fundiert beantworten zu können als auch zentrale Informationen für die Gestaltung und spätere Evaluation des KI-gestützten Analyseprozesses zu gewinnen. Das qualitative Datenmaterial wurde hierzu in sechs Schritten anhand der fokussierten Interviewanalyse nach Kuckartz und Rädiker (2020) und mit Hilfe der Software (MAXQDA24) ausgewertet. Dabei wurde eine deduktive-induktive Strategie zur Gestaltung des Kategoriensystems gewählt.

Nach der Exploration des Datenmaterials wurde in einem zweiten Schritt die erste Fassung thematischer Hauptkategorien deduktiv auf Basis der Leitfäden der Experteninterviews abgeleitet. Die finalen Hauptkategorien wurden in folgende Punkte eingeteilt:

- Status Quo - Aktueller Stand der Forschung zu KI
- Definition / Verständnis KI im UXR
- Mensch und KI im UXR
- KI-Einsatz im UXR
- Grenzen und Herausforderungen des KI-Einsatzes
- Evaluationskriterien/Validierungsmöglichkeiten
- (Entwicklung und Zukunftsperspektiven)

Die in den Experteninterviews erhobene Kategorie der *Zukunftsperspektiven* wurde gestrichen, da sie nicht zur Beantwortung der Fragestellung beitrug. Entsprechend kann sie nach Kuckartz und Rädiker (2020) zur Kategorie *Blumen am Wegesrand* hinzugefügt werden, da die Inhalte für zukünftige Forschung weiterhin interessant sein könnten.

Der Basiscodierung nachfolgend wurden die Subkategorien induktiv aus dem transkribierten Material herausgearbeitet. Das vollständige Kategoriensystem ist im Anhang H abgelegt. Zu jeder Kategorie wurden Code-Memos innerhalb von MAXQDA verfasst welche die nachträgliche Nachvollziehbarkeit sowie eindeutige Zuordnungen von Textteilen aus dem Datenmaterial ermöglichen. Die Code-Memos wurden jeweils mit einer Frage oder Anmerkung begonnen und durch Ankerbeispiele ergänzt. Zur finalen Auswertung der Experteninterviews wurde mit thematischen Zusammenfassungen gearbeitet, welche auf den jeweiligen Codes basierten.

Qualitative Forschung orientiert sich an bestimmten Gütekriterien (Flick, 2019). Im Folgenden wird darum dargestellt, wie diese Kriterien in der vorliegenden Arbeit angewandt wurden. Das Gütekriterium der selektiven Plausibilisierung wird durch die Verwendung von Interviewzitatens sowie der Darstellung gegensätzlicher Aussagen sichergestellt. Beides trägt gleichzeitig zur Nachvollziehbarkeit der Analyse bei. Darüber hinaus wird im Sinne von Transparenz und Nachvollziehbarkeit das verwendete Kategoriensystem im Anhang dokumentiert. Zur Sicherstellung des Gütekriteriums der Reliabilität wurde für alle Experteninterviews ein halbstrukturierter Leitfaden eingesetzt, der eine einheitliche und konsistente Datenerhebung gewährleistet. Zudem unterstützt die Anwendung der fokussierten Inhaltsanalyse eine strukturierte und übersichtliche Analyse der Daten. Neutralität und Objektivität wurden durch die offene und explorative Herangehensweise an das Forschungsfeld gewährleistet. Es besteht zudem weder eine Zugehörigkeit zur Praxispartnerin, noch gibt es finanzielle Interessen an der vorliegenden Arbeit, was zusätzlich eine neutrale und unvoreingenommene Perspektive gewährleistet.

3.4 Fallstudie

Fallstudien eignen sich besonders für Forschungsfelder, in denen bislang nur begrenzte wissenschaftliche Erkenntnisse vorliegen, weil sie ein vertieftes Verständnis ermöglichen, *wie* und *warum* Dinge passieren (Brüsemeister, 2008; Ridder, 2017). Ein solcher Fall kann unterschiedliche Untersuchungseinheiten umfassen, darunter Einzelpersonen, Gruppen, Organisationen oder spezifische Ereignisse und Problemstellungen. Ziel einer Fallstudie ist es, den betrachteten Fall detailliert zu beschreiben und ihn in seinem jeweiligen Kontext präzise zu rekonstruieren sowie nachvollziehbar darzulegen, weshalb der ausgewählte Fall als typisch oder exemplarisch für diesen Kontext gilt. Gleichzeitig schränkt die Fokussierung auf einen einzelnen Einzelfall die Möglichkeit für Verallgemeinerungen bzw. die externe Validität ein, da sich die Ergebnisse nur bedingt auf andere Kontexte oder weitere Fälle übertragen lassen (Flick, 2019; Ridder, 2017).

3.4.1 Beschreibung Fallstudie

Vor diesem Hintergrund wurde im Rahmen der vorliegenden Arbeit eine Fallstudie in Zusammenarbeit mit und anhand der Praxispartnerin sinnhaft GmbH durchgeführt. Bei der sinnhaft GmbH handelt es sich um eine UX-Agentur und ein KMU, das stellvertretend für viele

Organisationen im UX-Bereich steht. Wie aus dem Stakeholderinterview zu entnehmen war, sind KMU in der UX-Branche weit verbreitet und stehen häufig ähnlichen strukturellen Herausforderungen gegenüber, etwa begrenzten zeitlichen und personellen Ressourcen und der Notwendigkeit, Methoden effizient und pragmatisch einzusetzen (vgl. Flaherty & Dykes, 2024; Norman, 2013). Diese Bedingungen prägen den praktischen Umgang mit UXR und machen sinnhaft GmbH zu einem typischen und repräsentativen Fallbeispiel im UX-Umfeld. Das damit verbundene Fallbeispiel, der Analyse eines Usability-Tests, als etablierte und weit verbreitete Methode ermöglicht ausserdem die Untersuchung eines Kernprozesses des UXR. Usability-Tests gehören zu den häufigsten und methodisch zentralen Verfahren der Branche, wodurch sich sowohl wissenschaftliche Anschlussfähigkeit als auch praktische Relevanz herstellen lässt (vgl. Berni & Borgianni, 2021; Jacobsen & Meyer, 2019; Lewis & Sauro, 2021; Vermeeren et al., 2010).

3.4.2 Beschreibung des Datensatzes

Der Datensatz, die menschliche Referenzanalyse und die folgenden Erläuterungen basieren auf der Master-Thesis von Honegger (2023). Die Daten wurden von der Praxispartnerin zur Verfügung gestellt und entsprechen einem typischen Kundenprojekt des Unternehmens.

Im vorliegenden Fallbeispiel bezieht sich der Datensatz auf den New Application Process (NAP) von professional.ch. Das Kundenprojekt dreht sich um die Evaluierung eines neu entwickelten Online-Bewerbungsprozess der Job-Plattform. Dieses wurde im Rahmen eines Usability-Tests mithilfe eines klickbaren HTML-Prototyps in deutscher Sprache evaluiert. Einzelne Bereiche des Prototyps waren bereits mit Beispieldaten vorausgefüllt. Die Tests wurden auf mobilen Endgeräten (Smartphones) durchgeführt.

Insgesamt wurden $N = 22$ moderierte sowie $N = 22$ unmoderierte Usability-Tests erhoben. Während dieser Tests wurden Bildschirmaufzeichnungen inklusive Tonaufnahme erstellt, die anschliessend für die Auswertung und Identifikation von Usability-Problemen zur Verfügung standen. Für die vorliegende Arbeit wurden ausschliesslich die moderierten Testdaten berücksichtigt, da diese qualitative Daten aus der Anwendung der Thinking-Aloud-Methode während des Tests enthielten. Aus diesen wurden $n = 11$ Fälle zufällig ausgewählt und für die KI-Analyse sowie den Mensch-KI-Vergleich herangezogen. Die Dauer eines Tests lag zwischen 30 und 45 Minuten.

Da für den Mensch-KI-Vergleich in erster Linie der Analyseprozess bzw. die menschlichen Referenzdaten und nicht die inhaltliche Interpretation einzelner Fälle im Vordergrund stand, war das ursprüngliche Sampling des Datensatzes für diesen Zweck nicht ausschlaggebend und wurde daher nicht berücksichtigt. Die inhaltliche Relevanz der Daten sowie des Samplings war bereits durch die ursprüngliche Erhebung von Honegger (2023) sichergestellt.

Zur Identifikation von Usability-Problemen innerhalb der erhobenen Daten wurden diese von zwei UX-Expert*innen evaluiert. Die Auswertung erfolgte anhand der Videoaufnahmen bzw. Bildschirmaufzeichnungen, wobei die identifizierten Probleme systematisch in einer Excel-Tabelle festgehalten wurden. Für jedes Problem wurde eine kurze Beschreibung, eine Einschätzung des

Schweregrads und gegebenenfalls ein ergänzender Kommentar dokumentiert. Zusätzlich wurde festgehalten, bei welcher Testperson das jeweilige Usability-Problem auftrat. Diese Auswertung diente im Mensch-KI-Vergleich als Referenzbasis für den Abgleich der KI-Ergebnisse mit der menschlichen Analyse.

Der Ablauf des NAP während der Usability-Tests gestaltete sich wie folgt: Ausgehend von einem fiktiven Stelleninserat gelangten die Nutzer*innen zunächst auf eine Willkommenseite, auf der ihnen der Bewerbungsprozess kurz erläutert wurde. Anschliessend wurden persönliche Daten wie Geschlecht, Name und Adresse erfasst. Darauf folgten Fragen zu Ausbildung und beruflichem Werdegang. Um das Profil weiter zu vervollständigen, wurden Informationen wie besondere Fähigkeiten, Sprachkenntnisse oder absolvierte Kurse abgefragt. Am Ende konnten alle Eingaben durch einen automatisch generierten Lebenslauf überprüft werden. Durch die zusätzliche Option der Quick-Bewerbung sollte ein vereinfachter Bewerbungsprozess ermöglicht werden. Diese entsprach dem regulären Ablauf, verzichtete jedoch auf die erweiterten persönlichen Informationen in Schritt 7 (vgl. Honegger, 2023). Eine grafische Darstellung des NAP wird in Abbildung 6 veranschaulicht. Eine detaillierte Übersicht der einzelnen Schritte befindet sich zudem in Anhang I.

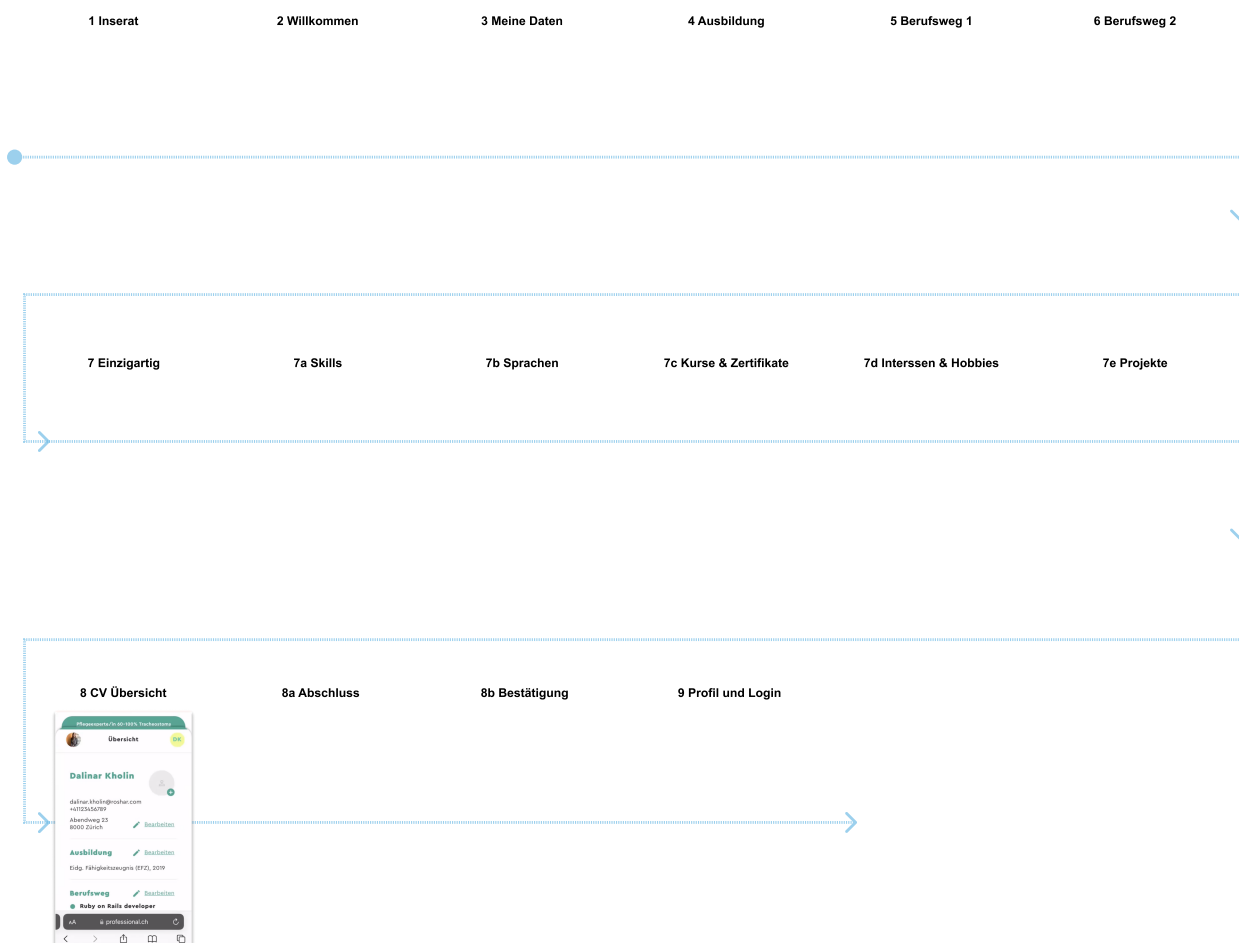


Abbildung 6. Grafische Darstellung des NAP-Ablaufs. Aus «Wirtschaftlichkeit von Usability-Evaluationsmethoden» von L. Honegger, 2023, Hochschule für Angewandte Psychologie FHNW, S. 42. Copyright 2023 bei Honegger. Wiedergabe mit Genehmigung.

3.4.3 KI-Transkription Usability-Test

Weil die Daten im ursprünglichen Analyseprozess nicht transkribiert worden waren, musste dieser Schritt zur Vorbereitung des Mensch-KI-Vergleichs nachträglich durchgeführt werden. Hierzu wurden die Aufnahmen der moderierten Usability-Tests herangezogen, welche unter Anwendung der Thinking-Aloud-Methode durchgeführt worden waren.

Im Oktober 2025 wurden die entsprechenden Videoaufzeichnungen mithilfe des KI-basierten Transkriptionsdienstes *Schweizerdeutschübersetzen.ch* transkribiert (Schweizerdeutschübersetzen.ch, 2025). Die Software erkannte Sprecherwechsel vorwiegend automatisch und verarbeitete die überwiegend auf Schweizerdeutsch geführten Usability-Tests, indem die Audiospur der Videodateien in ein Transkript umgewandelt wurde.

3.5 Mensch-KI-Vergleich

Das Konzept zur Integration eines KI-gestützten Analyseprozesses in den UXR-Prozess der sinnhaft GmbH orientiert sich an bestehenden Arbeits- und Vorgehensweisen des Unternehmens. In der Literatur wird beschrieben, dass KI grundsätzlich in allen Phasen des UXR-Prozesses eingesetzt werden kann, was auch mit dem vorliegenden Konzept verfolgt wird. Dafür wird ein AI-in-the-Loop-Ansatz angewendet, bei dem die KI den Menschen bzw. die Forschenden im UXR-Prozess unterstützend ergänzt, die Verantwortung jedoch beim Menschen verbleibt. Das Ziel dabei ist, die Kontrolle über den Analyseprozess sicherzustellen sowie Transparenz, Vertrauen und Interpretierbarkeit während des ganzen Prozesses zu gewährleisten (Komera & Manche, 2023; Natarajan et al., 2025). Entsprechend wurde jeder Schritt der KI-gestützten Analyse durch einen Menschen begleitet und überprüft. Auf den Einsatz eines agentischen, vollautomatisierten KI-Workflows wurde bewusst verzichtet, um die Kontrolle sowie die Nachvollziehbarkeit der von der KI generierten Analyseergebnisse sicherzustellen. Die daraus gewonnenen Erkenntnisse wurden fortlaufend im Forschungstagebuch dokumentiert.

Auf Basis der Experteninterviews und der Literaturrecherche wurde gemeinsam mit der Praxispartnerin entschieden, OpenAI ChatGPT Version 5.2 als KI-Tool für den Mensch-KI-Vergleich einzusetzen. Für die KI-Analyse wurde das zum Zeitpunkt der Durchführung aktuellste Modell von OpenAI in der kostenpflichtigen Premium-Version verwendet. Die Auswahl erfolgte vor dem Hintergrund, dass dieses LLM-Modell auf umfangreichen Trainingsdaten basiert und die Verarbeitung sowie Analyse komplexer sprachlicher Muster in qualitativen Daten unterstützt. Zudem wird dem aktuellen Modell eine geringere Anfälligkeit für Halluzinationen sowie eine stabilere Verarbeitung längerer Texte wie Transkripten zugeschrieben (OpenAI, 2025a). Solche Modelle werden ausserdem in wissenschaftlichen Studien und der Praxis häufig eingesetzt (Fischer, 2023; Goethals et al., 2025; Komera & Manche, 2023; Lauer et al., 2024; Ledger & Mancinni, 2024; Scott & Zuccon, 2024).

Wie bereits erwähnt, wurden die im Vorhinein erhobenen Projektdaten von der Praxispartnerin zur Verfügung gestellt und sind in Kapitel 3.4.2 näher beschrieben. Aus den insgesamt 22

moderierten Usability-Tests wurden per Zufallsstichprobe $n = 11$ Fälle für die KI-Analyse ausgewählt. Im Vorfeld der KI-Analyse war, wie auch bereits beschrieben, eine umfassende Datenaufbereitung notwendig. Da sowohl kommerzielle KI-Tools zur Verarbeitung multimodaler Daten als auch wissenschaftliche Erkenntnisse zu deren Einsatz im Usability-Kontext zum Zeitpunkt der Studiendurchführung limitiert waren (vgl. Liu, 2025), standen keine KI-Systeme zur Verfügung, welche die für Usability-Tests typischen multimodalen Daten vollständig verarbeiten konnten. Um einen möglichst fairen und vergleichbaren Mensch-KI-Vergleich zu ermöglichen, wurden die KI-generierten Transkripte manuell mit relevanten Kontextinformationen aus den Bildschirmaufnahmen der Tests ergänzt. Im selben Schritt wurden zudem Fehler aus der KI-gestützten Transkription korrigiert und bereinigt, um falsche oder fehlende Informationen für die weitere Analyse zu minimieren.

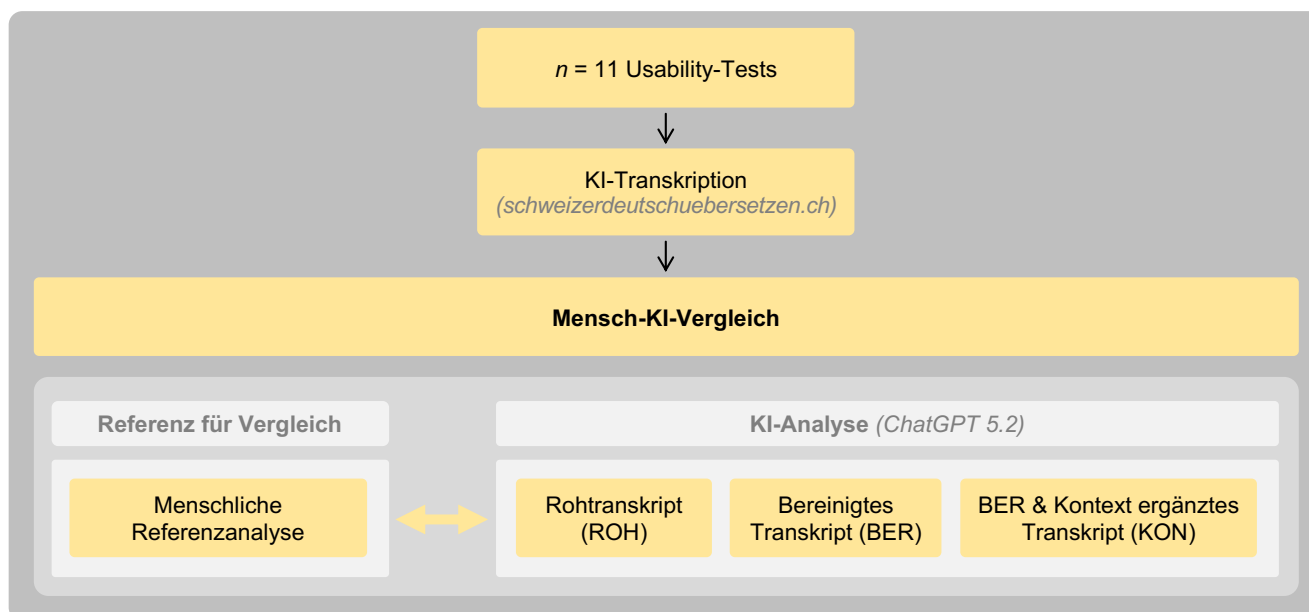


Abbildung 7. Konzept KI-gestützte Analyse und Mensch-KI-Vergleich (eigene Darstellung)

Aus dieser Vorgehensweise ergaben sich drei unterschiedliche Varianten für den Vergleich und die Messung der Zuverlässigkeit der KI-Analyse (vgl. Abbildung 7).

- Erstens wurde das Rohtranskript (ROH) direkt von der KI auf Usability-Probleme analysiert. Dieses Transkript wurde nach der KI-gestützten Transkription nicht überarbeitet und enthielt entsprechend alle dabei entstandenen Fehler und Ungenauigkeiten.
- Zweitens wurden manuell bereinigte Transkripte (BER) analysiert, um zu untersuchen, ob sich durch die Korrektur der Transkriptionsfehler Unterschiede in den Ergebnissen ergeben. Diese Transkripte wurden so angepasst, dass sie inhaltlich vollständig und korrekt der Audiospur der Videos entsprechen.
- Drittens wurde eine Variante verwendet, bei der die Transkripte sowohl manuell bereinigt als auch um zusätzliche Kontextinformationen aus den Usability-Test-Videos ergänzt wurden (KON). Diese Kontextinformationen wurden direkt in den Transkripten in eckigen Klammern kenntlich gemacht, um der KI zusätzliche Informationen aus den Bildschirmaufnahmen

bereitzustellen, die sie andernfalls nicht hätte verarbeiten oder auswerten können. Ergänzt wurden unter anderem Angaben dazu, an welcher Stelle des Prozesses sich die Proband*innen gerade befanden, sowie relevante Informationen, die nicht verbalisiert wurden und daher im Transkript nicht enthalten waren. Zur Veranschaulichung der manuell ergänzten Kontextinformationen dient Abbildung 8.

Interviewer:in: Du würdest erwarten, dass man das in irgendeiner Form speichern kann?

Antwortende:r: Genau. [Screen «Berufsweg 2»; Füllt Informationen zu Haupttätigkeit ein] Zum Beispiel, wenn ich meine letzten Jobs speichere. Oder dass ich einen kleinen Lebenslauf anlegen kann, der dann mitgeschickt wird. Letztendlich habe ich ja etwas gelernt und das wird sicherlich so bald ändern und nicht der letzte Job sein, den ich haben werde. Also es geht um Dinge wie Berufsbezeichnungen und Arbeitsgeber. [klickt auf «Speichern»; die eingetragenen Informationen verschwinden]

Interviewer:in: Was ist in deinem Verständnis passiert?

Antwortende:r: Ich habe gedacht, dass es wie vorher war, dass ich nur den letzten Arbeitsgeber angeben muss. [Ist zurück beim Screen «Ausbildung» und gibt erneut seine Daten ein]

Abbildung 8. Beispiel von manuell ergänzten Kontextinformationen in einem KI-Transkript (eigene Darstellung)

Zur Vorbereitung der KI-Analyse wurde ein Prompt als klar definierte Aufgabenstellung für die KI entwickelt, um eine zielgerichtete und für den Mensch-KI-Vergleich geeignete Auswertung der Daten in Bezug auf die Identifikation von Usability-Problemen sicherzustellen. Dabei wurde darauf geachtet, die Aufgabe präzise zu formulieren, ausreichend Kontext bereitzustellen und die gewünschte Ergebnisstruktur klar zu definieren. Zur Optimierung des Prompts wurden mehrere Iterationen und Tests mit einem ersten Datensatz durchgeführt. Der Prompt wurde mithilfe von KI (ChatGPT) selbst weiter verfeinert. Der finale und verwendete Prompt ist im Anhang K dargestellt.

Evaluation und Auswertung der KI-Analyse und Mensch-KI-Vergleich

Die Evaluationskriterien für den Vergleich zwischen KI und Mensch basierten auf den Erkenntnissen der vorangegangenen Literaturrecherche sowie den Experteninterviews. Um eine möglichst hohe Vergleichbarkeit der Analysen sicherzustellen, wurde der KI-gestützte Analyseprozess eng an den bestehenden Ablauf der menschlichen Auswertung angepasst. Zur Evaluierung der von der KI identifizierten Usability-Probleme des Usability-Tests wurde die Konfusionsmatrix als Methode eingesetzt, die häufig zur Qualitätsmessung von ML-Algorithmen verwendet wird (vgl. Abbildung 9). Diese Matrix ermöglicht im ML-Kontext eine differenzierte Betrachtung der Verteilung von tatsächlichen und vorhergesagten Werten, welche helfen, die Algorithmen bzw. Modelle anschliessend zu optimieren. Durch die Matrix können unterschiedliche Fehlerarten systematisch identifiziert und analysiert werden (Humm et al., 2022; Soares et al., 2025). Für den folgenden Mensch-KI-Vergleich trägt diese Vorgehensweise dazu bei, die Übereinstimmungen und Abweichungen zwischen der KI-gestützten Analyse und der menschlichen Referenzanalyse transparent und nachvollziehbar darzustellen.

Als Referenz bzw. Grundlage für den Vergleich dienten die Ergebnisse zweier menschlicher UX-Expert*innen aus der Auswertung der Usability-Tests des bestehenden Datensatzes nach Honegger (2023) wie bereits in Kapitel 3.4.2 beschrieben. Diese menschlichen Referenzanalysen dienten als Vergleichsgrundlage für den Mensch-KI-Vergleich und wurden den Ergebnissen der KI-gestützten Analyse gegenübergestellt. Zur systematischen Einordnung der Ergebnisse wurden drei Klassen entsprechend der Konfusionsmatrix unterschieden (vgl. Humm et al., 2022):

- **Richtig-Positiv (RP):** Die KI identifiziert ein Usability-Problem korrekt, das auch von den menschlichen UX-Expert*innen festgestellt wurde.
- **Falsch-Positiv (FP):** Die KI identifiziert ein Usability-Problem, das von den menschlichen UX-Expert*innen nicht als solches bewertet wurde.
- **Falsch-Negativ (FN):** Die KI übersieht ein Usability-Problem, das von den menschlichen UX-Expert*innen identifiziert wurde.

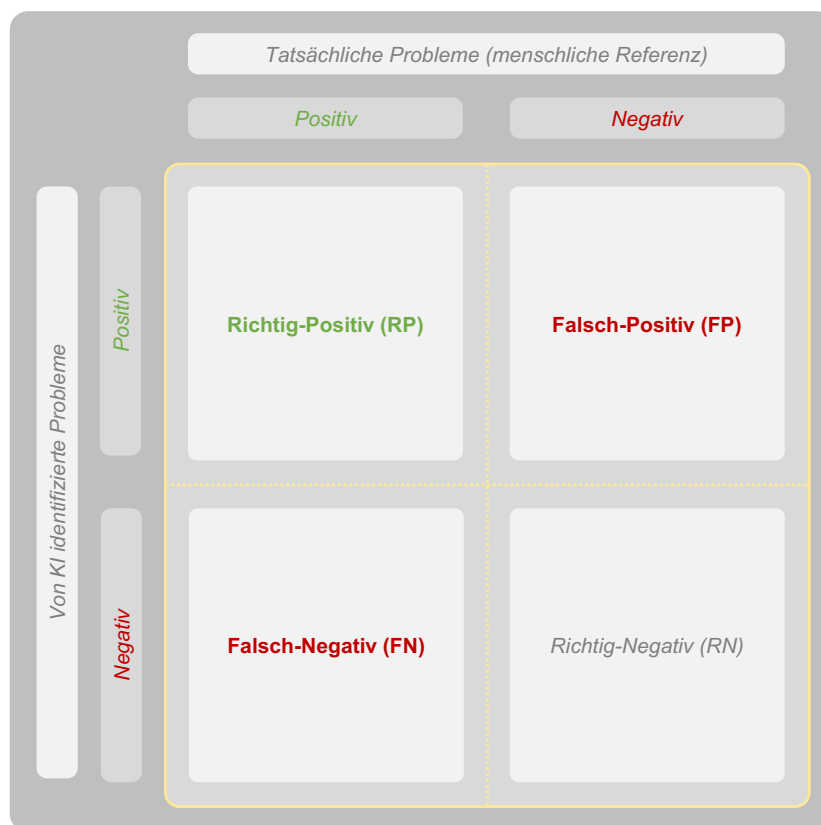


Abbildung 9. Konfusionsmatrix (in Anlehnung an Humm et al., 2022, S.21, eigene Darstellung)

In Abbildung 9 wird zusätzlich die Klasse der Richtig-Negativen (RN) dargestellt. Diese bezeichnet Fälle, in denen die KI kein Usability-Problem identifiziert und auch in der menschlichen Referenzanalyse kein entsprechendes Problem festgestellt wurde. Für die weitere Analyse und Interpretation werden Richtig-Negative Fälle jedoch nicht berücksichtigt, weil nicht eindeutig und abschliessend geklärt werden kann, welche potenziellen Usability-Probleme tatsächlich korrekt *nicht* identifiziert wurden.

Um die zuvor beschriebene Einordnung besser interpretieren zu können, wurde erneut auf etablierte Konzepte zur Bewertung von ML-Algorithmen zurückgegriffen. Für die Bewertung der KI-Analyse wurden folgende Metriken herangezogen (vgl. Soares et al., 2025):

- **Recall (Sensitivität):** Beschreibt, wie gut die KI in der Lage ist, alle relevanten Usability-Probleme (Vergleich zur Referenz) zu erkennen. Ein hoher Recall weist bspw. darauf hin, dass nur wenige von den menschlichen Expert*innen identifizierten Probleme von der KI übersehen werden (= FN).

$$\frac{\text{RichtigPositiv}}{\text{RichtigPositiv} + \text{FalschNegativ}}$$

- **Precision (Präzision):** Gibt an, wie zuverlässig bzw. treffsicher die KI bei der Identifikation von Usability-Problemen ist. Eine hohe Precision bedeutet, dass die als problematisch klassifizierten Stellen überwiegend tatsächlich Usability-Probleme darstellen und somit nur wenige FP auftreten.

$$\frac{\text{RichtigPositiv}}{\text{RichtigPositiv} + \text{FalschPositiv}}$$

Für den Mensch-KI-Vergleich bzw. systematischen Gegenüberstellung der KI-gestützten und menschlichen Usability-Analysen wurde eine konsolidierte Master-Problemliste in Excel erstellt. Darin wurden für alle $n = 11$ Fälle die identifizierbaren Usability-Probleme aus der menschlichen Referenzanalyse den Ergebnissen der KI-gestützten Analyse gegenübergestellt. Jedes Usability-Problem wurde mit einer eindeutigen Problem-ID versehen und durch eine kurze, abstrahierte Beschreibung definiert. Zusätzlich wurde die jeweilige Stelle im Prozess erfasst, an der das Problem auftrat, um die Konsolidierung der identifizierten Usability-Probleme zu erleichtern und die Nachvollziehbarkeit zu gewährleisten. Für jedes Problem wurde erfasst, ob es von der KI-gestützten Analyse erkannt wurde (1) oder nicht (0). Auf dieser Basis wurden die Kennwerte RP, FP und FN sowie Recall und Precision bestimmt. Ein exemplarischer Ausschnitt der konsolidierten Master-Excel-Liste ist in Anhang L dargestellt.

Ergänzend wurde im Sinne des explorativen Vorgehens eine qualitative Einordnung zusätzlicher und ausschliesslich durch die KI identifizierter Usability-Probleme durchgeführt. Analysiert wurden jene Befunde, die in der menschlichen Referenzanalyse nicht enthalten waren und im quantitativen Vergleich als FP klassifiziert wurden. Ziel dieser Analyse war es zu untersuchen, ob sich systematische Muster in den zusätzlichen KI-Identifikationen erkennen lassen und inwiefern die KI die menschliche Analyse bei der möglichst vollständigen Erfassung von Usability-Problemen unterstützen kann. Hierzu wurden die von der KI zusätzlich identifizierten Usability-Probleme manuell und inhaltlich geprüft, um deren Plausibilität und potenzielle Relevanz einzuschätzen. Diese qualitative Einordnung erfolgte unabhängig von der zuvor beschriebenen quantitativen Auswertung der Kennwerte (TP, FP, FN, Recall und Precision) und hatte keinen Einfluss auf deren Ergebnisse.

4 Ergebnisse

Im folgenden Kapitel werden die Ergebnisse aus den Experteninterviews sowie die Resultate des eigentlichen Mensch-KI-Vergleichs dargestellt.

4.1 Experteninterviews

Die Ergebnisse aus den Experteninterviews sind im Folgenden in drei Themenschwerpunkte gegliedert: (1) aktueller Stand der Forschung zum Einsatz von KI, (2) Mensch und KI im UXR sowie (3) Limitationen und Herausforderungen des KI-Einsatzes.

Die Experteninterviews orientierten sich dabei an der übergeordneten Leitfrage, welche Implikationen, Herausforderungen und Potenziale sich durch die Integration KI-gestützter Prozesse in den UXR-Prozess eines KMU ergeben. Darüber hinaus wurden Anforderungen bzw. erforderliche Kompetenzen an die Anwender*innen bzw. Forschenden sowie der aktuelle Stand der Forschung thematisiert.

4.1.1 Aktueller Stand der Forschung zu KI-Einsatz

Experimentierphase

Die breite Öffentlichkeit habe durch die rasche Verbreitung und alltägliche Nutzung von ChatGPT (wie auch vergleichbaren KI-Modellen) den Eindruck, bereits mitten im KI-Zeitalter angekommen zu sein. Im Bereich der Forschung zeige sich jedoch ein deutlich differenzierteres Bild. Die Auswertung der Experteninterviews verdeutlicht, dass sich der Einsatz von KI zur Analyse qualitativer (UX-)Daten weiterhin in einer frühen, experimentellen Phase befindet, die vor allem von einer Gruppe Early Adopters (dt. frühe Anwender*innen) vorangetrieben wird. Diese Pioniere testen neue Funktionalitäten und Anwendungsmöglichkeiten, entwickeln erste Workflows und formulieren methodologische Überlegungen für den zukünftigen Einsatz.

«Ja, also was wir im Moment untersuchen, wir sind hier wirklich noch Early Adopters. Wir haben irgendwie alle das Gefühl, durch diese Bekanntheit von ChatGPT, dass wir mitten im KI-Zeitalter sind. Sind wir nicht. Wir sind alle gerade in einem grossen Feldversuch von den Early Adopters von GenAI [dt. KI-Generation].» (Experteninterview Nr. 3, Pos. 32)

*«Also respektive neue Tools schiessen wie Pilze aus dem Boden.»
(Experteninterview Nr. 1, Pos. 50)*

Begründet wird diese Einschätzung der Expert*innen durch die hohe Geschwindigkeit technologischer Fortschritte im Bereich von KI, die deutlich schneller voranschreitet als deren tatsächliche Nutzung oder Integration in bestehende Arbeitsprozesse. Obwohl aktuelle KI-Modelle und KI-Tools erheblich leistungsfähiger geworden seien, wird der tatsächliche Einsatz im Alltag bzw. der Praxis aufgrund von Unsicherheiten, offener Fragen und fehlender Standards als eher

zurückhaltend beschrieben. Besonders in der qualitativen Forschung, welche von einer interviewten Person als eher traditionelle und reflektierte Disziplin beschrieben wird, sei man vorsichtig, da Bedenken hinsichtlich wissenschaftlicher Eigenständigkeit, Validität und Nachvollziehbarkeit bestünden.

Entsprechend wird eine starke Diversität der Nutzung in der Forschungscommunity beschrieben, die von experimentierfreudiger Offenheit bis hin zu Skepsis und offener Ablehnung reiche. Der Einsatz von KI in der qualitativen Datenanalyse bleibe aktuell ein Thema, das erst im Begriff ist, sich methodisch und praktisch zu etablieren.

Fehlendes Grundverständnis über Funktionsweise von KI

Ein wiederkehrendes Thema in sämtlichen Interviews ist das allgemein als unzureichend wahrgenommene Grundverständnis über die Funktionsweise, Limitationen und Fehlerquellen aktueller KI-Modelle und -Tools. Alle Expert*innen betonen, dass zur Zusammenarbeit mit KI ein gewisses Mass an Verständnis über die Funktionsweise von KI-Modellen erforderlich ist, um eine fundierte Einschätzung treffen zu können, welche Aufgaben KI zuverlässig unterstützen kann und wann mit Fehlern gerechnet werden muss. Das fehlende Fachwissen wird ausserdem mit als Grund angesehen, dass viele Anwender*innen mit KI-Tools bzw. ihrer Integration noch zurückhaltend agieren. Welches Wissen und welche Kompetenzen Anwender*innen von KI-Tools nach Ansicht der Expert*innen mitbringen sollten, wird in Kapitel (4.1.2.1) näher erläutert.

Notwendigkeit von menschlicher Kontrolle

Über alle Interviews hinweg besteht Konsens darüber, dass der Mensch auch zukünftig wichtiger Bestandteil der Prozesse sein wird und vor allem eine Kontrollinstanz in KI-gestützten Forschungsprozessen bleiben muss. Die Expert*innen betonen, dass KI-Modelle grundlegende Unsicherheiten aufweisen, die sich auch bei zunehmender Leistungsfähigkeit der KI-Modelle nicht vollständig beseitigen lassen. Insbesondere die verbleibende stochastische Fehlerquote von etwa 4 bis 10 %, welche oft auch als Halluzinieren beschrieben wird, wird als strukturelles Problem beschrieben, das aus Sicht der Expert*innen eine vollständige Automatisierung von Analysen ausschliessen lässt. Ausserdem wird in den Interviews auf die mangelnde Nachvollziehbarkeit der Prozesse zwischen Ein- und Ausgabe verwiesen, die als Blackbox-Problem beschrieben wird. Vor diesem Hintergrund wird in den Interviews die Einbindung des Menschen in den Prozess als notwendiger Standard hervorgehoben, der sowohl der Kontrolle der Ergebnisse als auch der Sicherung wissenschaftlicher Qualitätskriterien dienen soll.

Auch im Hinblick auf die zukünftige Entwicklung agentischer KI wird die Bedeutung menschlicher Kontrolle betont, weil diese agentischen Workflows eigenständig mehrere Analyseschritte hintereinander übernehmen sollen. Hier liegt die Herausforderung in der Akkumulation von Fehlern (siehe Fehlerquote) verketteter KI-Prozesse, wobei jeder einzelne Schritt die Gesamtfehlerquote erhöhen könne. Alle Expert*innen sehen zwar das Potenzial solcher agentischer Workflows, bewerten diese Vision jedoch als noch nicht unmittelbar realisierbar. Für sie entscheidend ist das Vertrauen in die Ergebnisse, welches nur durch eine gute Zerlegung und

Kontrolle der einzelnen Schritte aufgebaut werden könne. Entsprechend wird davon ausgegangen, dass die Entwicklung stabiler agentischer Workflows noch mehrere Jahre in Anspruch nehmen dürfte.

4.1.2 Mensch und KI im UXR

Im Folgenden werden die Ergebnisse aus den Experteninterviews zum Thema Mensch und KI im UXR erläutert. Diese beinhalten die Meinungen der Expert*innen zur Rolle des Menschen und damit verbundenen Kompetenzen, sowie die Rolle der KI und durch die Integration verbundenen Potenziale und konkrete Anwendungsmöglichkeiten.

4.1.2.1 Rolle des Menschen

In den Experteninterviews zeigt sich ein klares und konsistentes Bild dahingehend, dass der Mensch mit dem Einsatz von KI im UXR noch immer eine zentrale und unverzichtbare Rolle einnimmt. Mehrere Expert*innen betonen, dass der Mensch im Forschungsprozess die Rolle einer übergeordneten Supervision übernehmen sollte. Trotz der steigenden Leistungsfähigkeit und Anwendungsmöglichkeiten aktueller KI-Modelle und -Tools soll die Verantwortung über den Prozess sowie die Bewertung bzw. Interpretation der Ergebnisse vollständig beim Menschen bleiben. Zu dieser Verantwortung gehöre es auch fachlich-konzeptionelle oder methodische Entscheidungen zu treffen und die Kontrolle bzw. Überwachung des gesamten Forschungsprozesses bis hin zum finalen Ergebnis zu gewährleisten. Durch die Integration von KI in die Forschungsprozesse wird erwartet, dass diese zukünftig interaktiver, elaborierter und kürzer bzw. schneller werden. Die Interaktion zwischen Forschenden und KI wird dabei von den meisten als partnerschaftlich und dynamisch beschrieben. Mit der KI habe man ein Gegenüber, das einem jederzeit zur Verfügung steht und einem konstruktives Feedback oder kreative Ideen bieten könne.

« [...] verändert sich natürlich auch die Rolle, der Forschenden hinzu (...) wenn es alles gut läuft, zu Überwacher des Prozesses. Und wenn es schlecht läuft, zu Sklaven des Prozesses.»
(Experteninterview No2, Pos. 61)

«Also die wichtigste Aufgabe vom Menschen ist es, Verantwortung für das Ergebnis zu übernehmen. Es ist sein Ergebnis, er ist der Autor. Das ist seine wichtigste Aufgabe. Nicht die KI ist Autor. Ich kann die Verantwortung nicht auf die KI schieben. [...] Das ist immer noch ein Stück Technik.»
(Experteninterview No3, Pos. 47)

Ein weiterer Vorteil effizienterer Prozessschritte durch den Einsatz von KI wird darin gesehen, dass Forschende mehr Zeit für den direkten Austausch mit Kund*innen bzw. Tester*innen gewinnen. Dadurch könnten sie sich stärker auf zwischenmenschliche Interaktionen und die Pflege von Beziehungen konzentrieren.

Erforderliche Kompetenzen des Menschen

Die Expert*innen sind sich einig, dass der Einsatz von KI im UXR ein breites Spektrum spezifischer Kenntnisse und Kompetenzen von den Forschenden bzw. Anwendenden erfordert.

a) Verständnis der Funktionsweise von KI

Hinsichtlich dieser Kompetenz herrschte unter allen Expert*innen Einigkeit. Als grundlegend notwendige Kompetenzen im Umgang mit KI werden einerseits ein vertieftes Verständnis ihrer Funktionsweise, insbesondere der Eigenschaften von Sprachmodellen, aber auch der zugrundeliegenden Wahrscheinlichkeitsberechnungen und der damit verbundenen Fehlerquellen (dazu genannt wurden Halluzinationen, systematisches Pleasing, Verzerrungen durch Trainingsdaten und Herkunft der Modelle), als auch die Kenntnis und das Wissen über die Limitationen bei der Generierung tatsächlich neuer oder kreativer Inhalte angesehen. Hinzu kommt das notwendige Bewusstsein für die fehlende Transparenz und die Blackbox-Problematik von KI-Systemen. Das für die Arbeit mit KI erforderliche Wissen schließt auch das Bewusstsein für Datenschutzaspekte mit ein. Dafür sei es erforderlich, die Herkunft, Verarbeitung und Speicherung der Daten zu kennen, um datenschutzrechtliche sowie Kundenvorgaben bei der Nutzung von KI einzuhalten. Angesichts der rasanten Entwicklung von KI-Tools wird eine kontinuierliche Auseinandersetzung mit der KI-Technologie und neuen -Tools sowie eine stetige Weiterbildung der Forschenden als zwingend angesehen.

b) Prompting

Ein entscheidender Kompetenzbereich ist das Prompting, wobei die Einschätzungen der Expert*innen hierzu variieren; dies, weil aktuelle KI-Tools meist sogenannte One-Click-Funktionalität (dt. Ein-Klick) integrieren würden, durch welche Prompting nicht mehr unbedingt notwendig sei. Andere Expert*innen betonen, dass Forschende trotzdem lernen müssten, Eingaben präzise zu formulieren, ausreichend Kontext bereitzustellen und die gewünschte Ergebnisstruktur zu definieren. Die Fähigkeit, Prompts zu entwickeln und zu verfeinern, wird von ihnen als essenziell angesehen, um die KI gezielt zu steuern und reichhaltige, qualitativ hochwertige Ergebnisse zu erhalten. Eine Expertin schlägt vor, Prompting aktiv in die Lernprozesse des jeweiligen Unternehmens zu integrieren und erzielte Erfolge sowie Strategien teamübergreifend zu teilen. Hierfür könnten sogenannte Research-Repositories (dt. Forschungs- und Wissensdatenbanken) genutzt werden, die zugleich die Nachvollziehbarkeit der Ergebnisse über den gesamten Forschungsprozess hinweg erleichtern.

c) Methodenwissen und fachliche Erfahrung

Ein weiterer relevanter Kompetenzbereich umfasst Methodenwissen und fachliche Erfahrung. Obschon KI gewisse Arbeitsschritte wie Transkription oder Inhaltsanalysen oder gar das Schreiben von Berichten übernehmen könne, ersetze sie nicht die grundlegenden methodischen Überlegungen bei der Planung von UXR oder die finale Interpretation und Einordnung der Ergebnisse. Forschende müssten daher sowohl über fundierte Kenntnisse verschiedener Methoden als auch über praktisches

Fachwissen im jeweiligen UX-Kontext verfügen, um die Ergebnisse validieren oder die Auswertungen und Tests weiterentwickeln zu können. Die Expert*innen betonen, dass etablierte und gängige Methoden weiterhin die Grundlage für Auswertungen mit KI-gestützten Tools bilden sollten und verweisen darauf, dass KI menschliche Prozesse nicht unbedingt ersetzt, sondern lediglich effizienter durchführt. Gleichzeitig erfordere der Umgang mit KI eine neue Kompetenz im Umgang mit der speziellen Art von Daten und Inhalten (KI-Ausgaben), welche durch die KI generiert werden. Anwender*innen bzw. Forschende müssten verstehen, wie sie mit generierten Daten bzw. KI-Ergebnissen umgehen, diese in die eigenen Analysen integrieren und die Informationsqualität bewerten können.

« [...] Methoden sind in der KI-Zeit wichtiger denn je, da ansonsten möglicherweise die KI überhand nimmt. » (Experteninterview No2, Pos. 50)

d) Frustrationstoleranz und Kritisches Denken

Wie in den vorherigen Abschnitten bereits erwähnt, betonen die Expert*innen die Notwendigkeit der Implementierung von Kontrollmechanismen im Umgang mit KI. Diese Routinen seien einerseits zeitaufwendig und andererseits emotional herausfordernd, da die Überprüfung der KI-Ergebnisse oft als mühsam empfunden werden könne. Ein hohes Mass an Frustrationstoleranz sei daher notwendig, um die wiederholten Iterationen und Korrekturschleifen produktiv zu bewältigen. Gleichzeitig ermögliche die Einbindung der KI eine Entlastung von anderen repetitiven und zeitaufwendigen Aufgaben, welche ihrerseits mehr Zeit in Anspruch nehmen würden.

Nicht zuletzt heben die Expert*innen die Bedeutung des kritischen und reflektierten Denkens hervor. Aufgrund der Tendenz von KI-Modellen zum Pleasing würden die von der KI generierten Ausgaben häufig überzeugend und passend wirken. Dennoch sollten Forschende, so die Expert*innen, stets die Grundannahme vertreten, dass die Ergebnisse fehlerhaft sein könnten und sie konsequent mit eigenem Wissen und durch Auseinandersetzung mit dem vorliegenden Datenmaterial überprüfen und validieren. Die Nutzung der KI sollte dabei als unterstützendes Werkzeug verstanden werden, das Vereinfachung, Effizienzsteigerung und Reflexion sowie kreatives Denken fördert, anstatt Prozesse oder menschliche Kompetenzen ganz zu ersetzen.

4.1.2.2 Potenziale und Anwendungsfelder von KI im UXR

Die Rolle der KI bei der Integration in den UXR-Prozess wird in erster Linie als unterstützend oder ergänzend angesehen. Der Analyseprozess entwickelt sich dadurch zu einem interaktiven und dynamischen Austausch zwischen Mensch und KI, wobei die KI als Buddy bzw. Partnerin fungiert. Die Verantwortung für die Validität, Relevanz und Qualität der Ergebnisse solle aber weiterhin beim Menschen bzw. den Forscher*innen liegen. Im Folgenden werden die während der Experteninterviews besprochenen Potenziale der KI-Integration sowie die erwarteten Unterschiede zur menschlichen Analyse dargestellt.

«[...] KI nicht als irgendwie, als das Böse anschauen, sondern kann auch als Tool, als Unterstützung. Ich sage häufig auch als Buddy zu verstehen. Wie beispielsweise beim Tauchen. [...] dass du als Experte etwas anschaust, analysierst und dann KI parallel dazu, dann kannst du ja diese Sachen vergleichen und Feedback erhalten.»
(Experteninterview No1, Pos. 19)

Effizienzgewinne und Zeitersparnis

Der am häufigsten genannte Faktor von KI im UXR liegt im Gewinn an Effizienz. Der zentrale Vorteil des Einsatzes von KI liege darin, dass erheblich Zeit- und Kosten gespart werden können. Konkret zeigt sich dies bei verschiedenen Aufgaben wie etwa bei der Erstellung von Personas, der Transkription, der Datenexploration, der Erstellung von Zusammenfassungen oder der qualitativen Datenanalyse. Durch die Verkürzung einzelner Prozessschritte können Forschende mehr Kapazitäten auf die direkte Arbeit mit Nutzer*innen verwenden und sich auf den fachlichen Austausch und die Pflege von Kundenbeziehungen konzentrieren. Gleichzeitig gehen die Expert*innen davon aus, dass eine gesteigerte Geschwindigkeit und Effizienz es auch kleineren Teams oder Unternehmen mit begrenzten Budgets und Ressourcen ermöglicht, komplexe und umfangreiche Projekte durchzuführen und die gesamte Produktentwicklung zu beschleunigen.

Objektivität

Ein weiteres Potenzial von KI und zugleich auch eine ihrer Schwächen liegt in ihrer Unvoreingenommenheit. Die Expert*innen beschreiben, dass KI-Modelle keine sozialen Berührungspunkte besitzen würden und daher Aspekte im Datenmaterial identifizieren können, die menschlichen Forscher*innen möglicherweise entgehen könnten. Dies könne dazu beitragen, individuelle menschliche Verzerrungen in der Interpretation qualitativ erhobener Daten zu reduzieren. Einige Expert*innen berichten zudem, dass bestimmte Modelle durchaus in der Lage seien, subtile sprachliche Nuancen wie etwa Sarkasmus zu erkennen. Diese Fähigkeit bzw. Funktionalität sei jedoch stark kultur- und kontextabhängig und werde wesentlich durch die zugrunde liegenden Trainingsdaten geprägt. Die Expert*innen verweisen in diesem Zusammenhang auch auf damit verbundene Risiken (vgl. Kapitel 4.1.3). Entsprechend bleibe eine sorgfältige menschliche Überprüfung dennoch unerlässlich, um Fehlinterpretationen zu vermeiden und die Validität der Ergebnisse sicherzustellen.

Elaborierte Sprache und Berichte

Ein weiterer Vorteil KI-gestützter Tools, welcher die Expert*innen anführen, liegt in ihrer Fähigkeit, in sehr kurzer Zeit kohärente, elaborierte und strukturiert aufbereitete Texte zu erzeugen. Sprachmodelle können umfangreiche Zusammenfassungen und erste analytische Vorschläge liefern, die Forschende dabei unterstützen, relevante Muster im Datenmaterial schneller zu identifizieren. Dies erleichtere insbesondere Aufgaben wie die Kategorisierung, Typenbildung oder die Erstellung früher Analysefassungen.

Einige Expert*innen betonen, dass KI-generierte Texte häufig überraschend differenziert formuliert sind und neue sprachliche Kombinationen hervorbringen, welche die eigene Perspektive erweitern können. Teilweise wird die Ansicht vertreten, dass KI-generierte Texte in bestimmten Kontexten die Qualität menschlicher Texte erreichen oder sogar übertreffen können. Gleichzeitig machen sie auch an dieser Stelle darauf aufmerksam, dass die Qualität dieser Ergebnisse stark von den zugrunde liegenden Daten und der Formulierung der Prompts abhängt. Einige Expert*innen erkennen zwar das beschriebene Potenzial, betonen jedoch, dass es KI-generierten Texten mitunter an inhaltlicher Tiefe fehlen kann und sich unzutreffende oder nicht im Datenmaterial begründete Inhalte einschleichen können, wie es bei Halluzinationen oder Pleasing vorkommen kann.

Konkrete Anwendungsmöglichkeiten von KI im UXR

Alle Expert*innen sehen vielfältige Anwendungsmöglichkeiten von KI im UXR, insbesondere zur Effizienzsteigerung von Prozessen und Kosteneinsparungen in Projekten. Die genannten Anwendungsbereiche umfassen sowohl vorbereitende, analytische und nachgelagerte Tätigkeiten. Besonders häufig erwähnt wurden der Einsatz von KI zur Transkription sowie zur inhaltlichen Auswertung qualitativer Daten. KI könne weiter bei der Entwicklung und Erweiterung von Fragenkatalogen, der Erstellung und Optimierung von Interviewleitfäden, Pretests, der Exploration von Daten, der Entwicklung von Kategoriensystemen und der Kodierung eingesetzt werden. Auch in der Berichterstellung sehen die Expert*innen grosses Potenzial, da KI schnell strukturierte Zusammenfassungen generieren und komplexe Inhalte sprachlich aufbereiten kann.

Darüber hinaus könne KI für die Generierung von Personas und User Journeys (dt. Reise von Nutzer*innen) sowie zur Analyse visueller Daten eingesetzt werden. Genannt wurden insbesondere Funktionen zur Mustererkennung in Bildern, etwa für Anwendungen im Eyetracking oder zur Auswertung von Emotionen, die bei der Interpretation nonverbaler Hinweise in Interviews unterstützen können. Für die Auswertung von Videodaten aus Usability-Tests hingegen wurden keine etablierten KI-Tools genannt. Eine Expertin berichtet, dass viele aktuelle KI-Tools lediglich das Transkript der Audiospur analysieren und keine umfassende Auswertung visueller Interaktionsdaten ermöglichen.

Bei der Integration in die UX-Prozesse mahnen die Expert*innen auch zur Vorsicht. Da grosse LLMs primär auf sprachliche Muster spezialisiert sind, kann es bei der Analyse quantitativer Daten zu Fehlinterpretationen kommen. Eine Analyse quantitativer Daten sei mit aktuellen LLMs eher nicht angezeigt. Als in der Praxis besonders häufig verwendetes Modell wird ChatGPT des Unternehmens OpenAI genannt, wobei die Expert*innen betonen, dass je nach Aufgabenstellung auch andere Modelle geeigneter sein können.

4.1.3 Limitationen und Herausforderungen des KI-Einsatzes

Die Expert*innen gehen in den Interviews auf eine Reihe technischer und methodischer Limitationen ein, welche den Einsatz von KI im UXR derzeit einschränken können. Wiederholt hervorgehobene Probleme, welche zuvor schon benannt wurden, sind Halluzinationen, Biases sowie das Pleasing. Als zusätzliche Fehlerquellen hinzukommen können Qualitätsprobleme bei Transkripten auftreten, vor allem bei Dialekten wie Schweizerdeutsch, sowie mögliche Kapazitätsgrenzen der KI-Modelle bei der Verarbeitung grosser Datenmengen. Letztere wird in den Interviews auch als Zustandslosigkeit von Chatbots beschrieben. Damit ist, das Vergessen oder nicht mehr Berücksichtigen früherer Informationen durch das Überschreiten der Token-Grenze der verwendeten Modelle gemeint. Zudem verweisen die Expert*innen auf die strukturelle Fehleranfälligkeit gegenwärtiger KI-Modelle. Genannt wurden Fehlerraten im Bereich von rund 4 bis 10 %, welche durch die auf Wahrscheinlichkeitsrechnungen aufbauende KI auftreten können. Vor diesem Hintergrund wird die automatisierte Aneinanderreihung mehrerer KI-basierter Prozessschritte, wie sie in agentischen Workflows typisch wäre, als problematisch eingeschätzt. Die Expert*innen betonen, dass dadurch die für Kontrollprozesse notwendige Nachvollziehbarkeit und Überprüfbarkeit der Ergebnisse zusätzlich erschwert werden würde.

Zwei Expert*innen thematisieren ein für den UXR-Kontext weitere Herausforderung, nämlich die begrenzte Fähigkeit von KI-Systemen, visuelle Daten zu verarbeiten. Viele KI-Tools würden lediglich Transkripte oder Textdateien auswerten, während Möglichkeiten zur Analyse von Videomaterial, insbesondere bei Usability-Tests, weitgehend fehlen würden. Zwar würden Ansätze zur Bild- und Mustererkennung, etwa in Form von Emotionserkennung oder Eye-Tracking (dt. Blickbewegungen) existieren, diese seien jedoch nicht direkt für eine Usability-Analyse einsetzbar.

Methodisch bleibe der Einsatz von KI klar begrenzt. KI könne menschliche Forschungsprozesse nicht ersetzen, insbesondere nicht die Datenerhebung oder den Kontakt zu realen Nutzer*innen. Kritisch sehen die Expert*innen den Einsatz synthetischer Nutzer*innen oder mit KI generierte Interviewdaten. Solche Praktiken werden als qualitativ unzureichend und potenziell schädlich beschrieben, da sie die Gefahr einer *genetischen Verarmung* des Wissens bzw. von Daten bergen würden und zur Reproduktion bereits vorhandener Muster und Datenpunkte führten.

*«Also wir dürfen nicht, nur weil wir diese Tools jetzt haben, das Gefühl haben, das [das Tool] ersetzt mir jetzt den Umgang oder respektive das Rausgehen und mit den realen Nutzer*innen zu arbeiten.»*

(Experteninterview No1, Pos. 54)

Schliesslich weisen die Expert*innen darauf hin, dass KI aufgrund ihrer Trainingsdaten oftmals eine globale und generalisierte Perspektive einnimmt, die nicht ohne Weiteres auf spezifische Zielgruppen übertragbar ist. Um kulturelle bzw. lokale Unterschiede adäquat zu berücksichtigen, bleibe die Einbindung realer Nutzer*innen weiterhin unverzichtbar.

Hürden für KMU

Eine zentrale Herausforderung für KMU bei der Implementierung von KI liegt, laut Expert*innen, in der Abhängigkeit von spezifischen Plattformen oder Anbietern. Hinzu kommt die rasche technologische Entwicklung, während die Anpassung in Unternehmen häufig langsamer erfolgt, was die Einführung oder Auswahl neuer KI-Lösungen zusätzlich erschwert. Darüber hinaus mangle es in Organisationen und UX-Teams häufig an einem ausreichenden Verständnis für die Möglichkeiten und Grenzen von KI. Dies führe oft dazu, dass menschliche Ergebnisse im Vergleich zur KI überschätzt würden. Fehlende gemeinsame Nutzung von Prompts und ein geringer Diskurs über die KI-Arbeit innerhalb von Teams würden zudem weiter eine effektive Nutzung der KI-Technologie behindern.

Ein weiterer damit verbundener Aspekt wird im fehlenden Vertrauen in KI-generierte Ergebnisse gesehen. Vertrauen könne durch transparente Offenlegung der Arbeits- und Lernprozesse sowie durch die Implementierung von Kontrollmechanismen aufgebaut werden. Das Vertrauen ist jedoch nicht nur innerhalb von Teams oder Unternehmen entscheidend, sondern spiele auch in der Geschäftsbeziehung mit der Kundschaft eine wichtige Rolle. So seien eine transparente Aufklärung sowie das Einverständnis der Kundschaft bzw. der Nutzer*innen zentrale Voraussetzungen dafür, dass deren Daten mithilfe von KI verarbeitet werden dürften. Besonders bei sensiblen Daten können der Datenschutz und die Anonymisierung eine zusätzliche Hürde bei der KI-Integration darstellen, so die Expert*innen.

Allgemeingültige Herausforderungen der KI-Technologie

Ergänzend zu den bereits erläuterten Punkten sehen die Expert*innen auch allgemeine Risiken im Umgang mit und in der Implementierung von KI in Unternehmensprozesse, die von allen ähnlich bewertet werden. Ein häufig genanntes Risiko besteht in der Abhängigkeit von Unternehmen, die KI-basierte Leistungen anbieten. So wird beschrieben, dass wenn es zu Ausfällen der Dienste komme oder die zugrundeliegenden Start-ups inkl. der von ihnen entwickelten Modelle gar ganz vom Markt verschwänden, müssten Unternehmen wieder auf manuelle Prozesse oder alternative Anbietende zurückgreifen, was planerische Herausforderungen mit sich bringt. Gleichzeitig bestehe eine Abhängigkeit von wenigen grossen und globalen Technologieunternehmen wie Google, Anthropic, OpenAI oder Apple, welche die Entwicklung und Finanzierung grosser KI-Modelle dominieren würden, so die Expert*innen. Dies könne zu einer Konzentration von Marktmacht führen und theoretisch den Einsatz bestimmter Filter oder Einschränkungen begünstigen, wie sie beispielsweise in China bei Deep-Seek-Modellen existieren würden, auch wenn letzteres den Expert*innen für westliche Länder weniger wahrscheinlich erscheint. Die Nutzung internationaler KI-Modelle erfordere trotzdem ein besonderes Bewusstsein für die Themen Datenschutz und Privatsphäre. Eine konsequente Anonymisierung sensibler Kunden- und Nutzerdaten sei dabei essenziell. Als mögliche Lösung werden von einer Expert*in lokal gehostete KI-Modelle angeführt, jedoch sei der Aufwand dafür noch zu hoch im Vergleich zu den webbasierten KI-Tools.

Darüber hinaus betonen die Expert*innen den hohen Kompetenzbedarf und fundiertes Methodenwissen im Umgang mit KI (vgl. Kapitel 4.1.2.1). Die Expert*innen stellen sich die Frage, auf

welche Weise Kompetenzen wie kritisches Denken in Gegenwart der KI-Technologie heutzutage noch erlernt oder ausgebildet werden können. Einige Expert*innen warnen zudem, dass der zunehmende Einsatz von KI dazu führen könnte, dass Menschen eigene Kompetenzen weniger aktiv einsetzen, was langfristig zu einem Verlust von wichtigen menschlichen Fähigkeiten führen könnte.

4.2 Mensch-KI-Vergleich

Für den Mensch-KI-Vergleich sowie zur Analyse der Unterschiede zwischen menschlicher und KI-gestützter Usability-Problemidentifikation wurden insgesamt $n = 11$ Usability-Tests ausgewertet. Hierzu wurde ein für die Praxispartnerin angepasster und entwickelter KI-gestützter Analyseprozess eingesetzt und evaluiert. Die Usability-Tests wurden in drei Transkriptvarianten (Roh-, bereinigtes und kontextualisiertes Transkript) durch die KI analysiert und auf Usability-Probleme hin ausgewertet. Die resultierenden KI-Analysen wurden anschliessend miteinander sowie mit der menschlichen Referenzanalyse verglichen. Auf dieser Basis wurden für die Einordnung der Ergebnisse die Kennzahlen RP, FN, FP sowie Recall und Precision bestimmt.

Die Ergebnisse werden im Folgenden zunächst quantitativ anhand absoluter Häufigkeiten und daraus abgeleiteter Kennzahlen wie Recall und Precision dargestellt. Abschliessend wird die ergänzend durchgeführte qualitative Einordnung jener Usability-Probleme dargestellt, die ausschliesslich durch die KI identifiziert wurden.

4.2.1 Vorbereitungen und Zeitaufwand

Im Rahmen der Vorbereitung mussten die erhobenen Usability-Testdaten, wie bereits erwähnt, zunächst für die KI-gestützte Analyse aufbereitet werden. Dieses Kapitel stützt sich unter anderem auf die im Forschungstagebuch dokumentierten Erkenntnisse. In einem ersten Schritt wurden die Aufzeichnungen der Usability-Tests automatisiert durch die KI transkribiert. Anschliessend wurden diese Transkripte manuell überprüft, bereinigt und um relevante Kontextinformationen aus den Bildschirmaufzeichnungen ergänzt (vgl. Kapitel 3.5). Die Überarbeitung eines Transkripts nahm im Durchschnitt etwa eine Stunde in Anspruch. Begleitend wurde im Forschungstagebuch die Qualität der KI-generierten Transkripte dokumentiert. Es zeigten sich wiederkehrende Auffälligkeiten, darunter teilweise fehlende Anonymisierung, fehlerhafte oder inkonsistente Sprechererkennung (z. B. Verwechslungen der Sprecher oder fehlende Zuordnung), ganz ausgelassene Textpassagen und Abschnitte sowie Wiederholungen einzelner Inhalte (vgl. Abbildung 10). Zudem traten vereinzelt fehlerhafte Übersetzungen vom Schweizerdeutschen ins Hochdeutsche auf. Zusätzlich war der für die Transkription genutzte KI-Dienst während eines gewissen Zeitfensters nur eingeschränkt verfügbar oder nutzbar. Infolge von Serverproblemen verlängerte sich die Dauer der automatisierten Transkription in diesem Zeitraum von durchschnittlich etwa 3 Minuten auf teilweise bis zu 20 Minuten pro Test. Darüber hinaus war es zeitweise nicht möglich, die fertiggestellten Transkriptionen herunterzuladen.

Interviewer:in: Sehr gut. Dann schicke ich dir jetzt den Link für die Vorbefragung. Du kannst einfach ausfüllen, das musst du noch nicht kommentieren.

Befragte/r: Ja, das ist gut.

Interviewer:in: Dann schicke ich dir den Link.

Befragte/r: Ja, das ist gut.

Interviewer:in: Dann schicke ich dir den Link.

Befragte/r: Ja, das ist gut.

Interviewer:in: Dann schicke ich dir den Link.

Befragte/r: Ja, das ist gut.

Interviewer:in: Dann schicke ich dir den Link.

Befragte/r: Ja, das ist gut.

Interviewer:in: Dann schicke ich dir den Link.

Befragte/r: Ja, das ist gut.

Interviewer:in: Dann schicke ich dir den Link.

Befragte/r: Ja, das ist gut.

Abbildung 10. Exemplarischer Ausschnitt eines KI-Transkripts mit fehlerhafter Transkription (eigene Darstellung)

Die Entwicklung und Überprüfung der Prompts für die KI-Analyse erfolgte iterativ und nahm insgesamt rund zwei Stunden in Anspruch. Für die Analyse der Roh- und bereinigten Transkripte sowie für die kontextualisierte Analyse wurden zwei unterschiedliche Prompts eingesetzt. In frühen Testläufen wurde zusätzlich der Schweregrad der identifizierten Usability-Probleme durch die KI erfasst. Um die Komplexität der Prompts zu reduzieren und potenzielle Fehlzuordnungen während der KI-Analyse zu minimieren, wurde diese Dimension im weiteren Verlauf nicht mehr berücksichtigt. Im Forschungstagebuch wurde in diesem Zusammenhang festgehalten, dass sich die Einschätzung Schweregrads sowohl zwischen den drei Transkriptvarianten als auch im Vergleich zur menschlichen Analyse teilweise deutlich unterschied und als unbrauchbar eingeschätzt wurde. Darüber hinaus wurde beobachtet, dass die Qualität der KI-Analyse mit der Qualität der zugrunde liegenden Transkripte variierte. Auffällig war dabei, dass sich Unterschiede unter anderem in Abhängigkeit davon zeigten, wie konsequent das Thinking-Aloud während der Tests umgesetzt wurde und ob es während der Durchführung zu Störungen, etwa Audioproblemen, kam. In solchen Fällen wurde das Gesagte zu den auftretenden Problemen von der KI in vielen Fällen als Usability-Problem des zu testenden Prototyps interpretiert. Zudem enthielten die Transkripte hypothetische Formulierungen der Testpersonen (z. B. «ich würde hier klicken», «ich stelle mir vor, dass...»), welche für die KI trotz eines entsprechend angepassten Prompts zur Steuerung dieser Tendenzen nicht eindeutig interpretierbar waren. Solche Äusserungen wurden von der KI wiederholt fälschlicherweise als Usability-Probleme klassifiziert.

Zur Einordnung der in der Literatur sowie in den Experteninterviews beschriebenen Effizienzgewinne und Zeitersparnisse werden im Folgenden, die im Rahmen des KI-gestützten Analyseprozesses dokumentierten, zeitlichen Aufwände dargestellt. Dies trägt ausserdem dazu bei, den Analyseprozess nachvollziehbar zu dokumentieren.

Der eigentliche KI-gestützte Analyseprozess dauerte pro Durchlauf durchschnittlich 10 Minuten. Hinzu kommt die Zeit zur manuellen Überprüfung der einzelnen Kontrollroutinen, welche jedoch jeweils nur rund 5 Minuten dauerten. Die anschließende Zusammenführung der KI-Ergebnisse mit der menschlichen Referenzanalyse sowie die Gegenüberstellung der identifizierten Usability-Probleme erforderten pro Datei einen zusätzlichen Zeitaufwand von rund einer Stunde. Zur Übersicht der zeitlichen Aufwände dient Tabelle 2:

Tabelle 2

Überblick über den Zeitaufwand der einzelnen Arbeitsschritte des Mensch-KI-Vergleichs (Zeitangaben gerundet)

Arbeitsschritt	Zeitaufwand pro Einheit	Anzahl	Gesamtzeit
KI-Transkription	3min (o. 20min)	11 Tests	0.55h
Transkriptüberarbeitung	60min	11 Transkripte	11h
Prompt-Entwicklung	120min	-	2h
KI-Analyse (Durchführung)	10min (+5min)	11 x 3 Analysen	8.25h
Ergebnisabgleich	60min	11 x 3 Analysen	33h
Gesamt			54.8h

4.2.2 Zuverlässigkeit und Aussagekraft KI-Analyseergebnisse

In einem nächsten Schritt wurden die Ergebnisse der KI-gestützten Analysen mit der menschlichen Referenzanalyse zusammengeführt. Hierzu wurden die Resultate aller Analysen in einem zentralen Masterdokument in Excel konsolidiert. Über alle Usability-Tests und identifizierten Usability-Probleme hinweg umfasste der Datensatz insgesamt 3960 Datenzeilen.

Die menschliche Identifikation von Usability-Problemen wurde als Referenz herangezogen und den jeweiligen KI-Ergebnissen gegenübergestellt. Auf dieser Grundlage erfolgte ein manueller und qualitativer Abgleich der identifizierten Usability-Probleme, um die absoluten Häufigkeiten von Richtig-Positiven (RP), Falsch-Negativen (FN) und Falsch-Positiven (FP) zu bestimmen. Fälle, in denen die KI kein Usability-Problem identifizierte und auch keine entsprechende menschliche Identifikation vorlag (Richtig-Negative, RN), wurden in der weiteren Auswertung nicht berücksichtigt (vgl. Kapitel 3.5). Fälle, in welchen die KI neue Usability-Probleme identifizierte, die keinem zuvor durch den Menschen identifizierten Problem zugeordnet werden konnten, wurden diese als eigenständige Einträge erfasst und methodisch den FP zugeordnet.

Die Auswertung erfolgte getrennt nach den drei Transkriptvarianten bzw. Analysearten der KI (ROH: Rohtranskript; BER: bereinigtes Transkript; KON: um Kontextinformationen ergänztes Transkript). Tabelle 3 zeigt die Ergebnisse der Analyse in absoluten Häufigkeiten.

Tabelle 3

Absolute Häufigkeiten der Übereinstimmung zwischen menschlicher und KI-Analyse in RP, FN und FP pro Analyseart (ROH, BER, KON)

Analyseart	RP	FN	FP
ROH	52	49	51
BER	64	37	35
KON	68	33	32

Über alle drei Analysearten hinweg zeigt sich ein klarer Anstieg der RP, von 52 korrekt identifizierten Usability-Problemen in der Analyse auf Basis von ROH über 64 in BER bis hin zu 68 in KON. Parallel dazu ist eine Abnahme der FN in derselben Reihenfolge zu beobachten (ROH = 49, BER = 37 und KON = 33). Daraus wird ersichtlich, dass sich der Anteil korrekt identifizierter sowie falsch identifizierter Usability-Probleme mit dem Grad bzw. Qualität der Transkriptaufbereitung verändert.

Gleichzeitig ist auf der anderen Seite eine Reduktion der FP zu beobachten. Während bei der Analyse auf Basis von ROH = 51 noch eine vergleichsweise hohe Anzahl zusätzlicher, nicht durch den Menschen identifizierter Usability-Probleme erkannt wurde, verringert sich diese Anzahl bei BER = 35 und KON = 32 zunehmend.

Aufbauend auf dieser ersten Analyse wurden für jede Analyseart die Bewertungsmasse Recall und Precision berechnet, um die Übereinstimmung bzw. Vollständigkeit zwischen KI-gestützter Analyse und menschlicher Referenz quantitativ zu beschreiben. Der Recall gibt dabei an, wie gross der Anteil der durch den Menschen identifizierten Usability-Probleme auch durch die KI korrekt erfasst wurde. Die Precision andererseits beschreibt, welcher Anteil der von der KI erfassten Usability-Probleme mit der menschlichen Referenzanalyse übereinstimmt. Die berechneten Ergebnisse pro Analyseart sind in Tabelle 4 ersichtlich.

Tabelle 4

Kennzahlen Recall und Precision pro Analyseart (gerundete Werte)

Analyseart	Recall	Precision
ROH	.52	.51
BER	.63	.65
KON	.67	.68

In der Analyse des Recall-Kennwerts auf Basis von ROH erfasst die KI rund 52 % der durch den Menschen identifizierten Usability-Probleme. Durch die Bereinigung der Transkripte steigt dieser Anteil auf etwa 63 % (BER), während in der kontextualisierten Analyse (KON) ein Recall von rund 67 % erreicht wird. Anders formuliert bedeutet dies, dass die KI mit KON gut 67 % der durch den Menschen identifizierten Usability-Probleme korrekt bzw. vollständig erfassen konnte.

Ein vergleichbares Muster zeigt sich für die Precision. Während in der ROH-Analyse gut die Hälfte (51 %) der durch die KI erfassten Usability-Probleme mit den menschlichen Referenzwerten übereinstimmt, erhöht sich dieser Anteil in der BER-Analyse auf 65 % und erreicht in der KON-Analyse mit 68 % den höchsten Wert. Bezogen auf die kontextualisierten Analyse bedeutet dies, dass 68 % der von der KI erfassten Usability-Probleme auch in der menschlichen Referenzanalyse enthalten sind bzw. mit ihr übereinstimmen.

4.2.3 Qualitative Einordnung zusätzlicher KI-identifizierter Usability-Probleme

Neben der quantitativen Übereinstimmung zwischen menschlicher Referenzanalyse und KI-Ergebnissen wurden zusätzlich jene Usability-Probleme analysiert, die ausschliesslich durch die KI identifiziert wurden und so nicht in der menschlichen Referenzanalyse enthalten waren (vgl. Kapitel 3.5). Die Ergebnisse dieser Analyse sind in Tabelle 5 ersichtlich.

Tabelle 5

Qualitative Einordnung zusätzlicher KI-identifizierter Usability-Probleme

Analyseart	FP	Als plausibel eingeordnet	Anteil plausibler FP (%)
ROH	51	15	29.4 %
BER	35	15	42.9 %
KON	32	16	50.0 %

Ausgehend von dieser Analyse zeigt sich, dass die Anzahl der als plausibel eingeschätzten zusätzlichen Usability-Probleme über alle Analysearten hinweg auf einem vergleichbaren Niveau liegen (ROH = 15, BER = 15, KON = 16). Im relativen Vergleich zeigt sich jedoch, dass sich der Anteil plausibler identifizierter Usability-Probleme an der Gesamtzahl der FP zwischen den Analysearten unterscheidet. Während dieser Anteil in der ROH-Analyse mit 15 Fällen bei rund 30 % liegt, steigt er in der BER-Analyse mit ebenfalls 15 Fällen auf 42.9 % und erreicht in der KON-Analyse mit 16 Fällen einen Wert von 50 %.

Zur Veranschaulichung der beschriebenen Fälle werden exemplarisch zwei ausschliesslich durch die KI identifizierte Usability-Probleme dargestellt, die im Rahmen der inhaltlichen Überprüfung und Einordnung als plausibel eingeschätzt wurden:

- **Beispiel 1: Nicht auswählbare Option «bis heute»**
Die KI identifizierte aus dem Transkript, dass die Option «bis heute» nicht auswählbar ist (z. B. «Wenn ich ‚bis heute‘ anklicken möchte, geht das nicht.»). Das Problem bezieht sich auf eine nicht ausführbare Interaktion mit dem Prototypen und wurde in allen drei Analysearten (ROH, BER und KON) erkannt.
- **Beispiel 2: Nicht bearbeitbare E-Mail-Adresse**
In den Analysen ROH und BER hatte die KI erfasst, dass die E-Mail-Adresse nicht bearbeitet werden konnte (z. B. «Bei der E-Mail-Adresse kann ich nichts angeben.»).

5 Diskussion

In diesem Kapitel werden die Ergebnisse der vorliegenden Arbeit im Hinblick auf die zugrunde liegenden Fragestellungen reflektiert und kritisch diskutiert. Die Struktur des Kapitels orientiert sich dabei an den zwei Leitfragen sowie der übergeordneten Hauptfragestellung dieser Studie. Ziel der vorliegenden Arbeit war es sowohl die Potenziale als auch die Herausforderungen der Integration von KI im UXR zu untersuchen und die damit verbundenen Anforderungen an Anwender*innen bzw. Forschenden herauszuarbeiten. Darüber hinaus wurde der entwickelte KI-gestützte Analyseprozess getestet und evaluiert, um Hinweise auf die Zuverlässigkeit und Aussagekraft der durch KI generierten Erkenntnisse im Vergleich zu menschlichen Referenzanalysen zu gewinnen

5.1 Potenziale und Herausforderungen der KI-Integration im UXR

Die Beantwortung der ersten Leitfrage stützt sich hauptsächlich auf die Auswertung der durchgeführten Experteninterviews sowie auf die herangezogene wissenschaftliche Literatur.

Leitfrage 1: *Welche Potenziale und Herausforderungen ergeben sich bei der Integration von KI-gestützten Prozessen im UX-Research eines KMU?*

5.1.1 Potenziale der KI-Integration im UXR

Die Rolle der KI bei der Integration im UXR wird von den Expert*innen in erster Linie als unterstützend für den Menschen verstanden. KI arbeitet dabei parallel zur menschlichen Analyse und wird als ergänzendes intelligentes Werkzeug in die Prozesse integriert (vgl. Korteling et al., 2021; Li et al., 2024). Der Analyseprozess entwickelt sich dadurch zu einem interaktiven und dynamischen Austausch zwischen Mensch und KI, wobei die Verantwortung für die Bewertung und Nutzung der Ergebnisse klar beim Menschen verbleibt. Diese Einschätzung deckt sich mit Positionen aus der wissenschaftlichen Literatur, in welcher der Fokus weniger auf einer vollständigen Substitution menschlicher Arbeit liegt, sondern vielmehr auf der Entwicklung neuer Formen der Interaktion zwischen Mensch und KI (vgl. Brandao, 2025; Brynjolfsson et al., 2025; Fragiadakis et al., 2024). Der HAIC beschreibt einen kollaborativen Ansatz für eine gegenseitig vorteilhafte Zusammenarbeit, bei der Aufgaben und Verantwortlichkeiten entlang der jeweiligen Stärken und Schwächen von Mensch und KI verteilt werden. Ziel ist es, die komplementären Fähigkeiten beider Parteien so zu kombinieren, dass ihre jeweiligen Potenziale synergetisch genutzt werden können. Vor diesem Hintergrund lassen sich im UXR verschiedene relevante Potenziale des Einsatzes KI-gestützter Technologien identifizieren (vgl. Fragiadakis et al., 2024; Song et al., 2024).

Sowohl die Literatur als auch die Expert*innen betonen vor allem Effizienz- und Zeitgewinne als zentrale Vorteile der KI-Integration. Sie verweisen darauf, dass KI dazu beitragen kann, UXR-Prozesse schneller, adaptiver und ökonomischer zu gestalten, was insbesondere in dynamischen und digital geprägten Unternehmen und Aufgabengebieten zur Sicherung der Wettbewerbsfähigkeit beitragen kann (vgl. Najana et al., 2024; Brynjolfsson et al., 2025). Besonders für KMU, die häufig

mit begrenzten Ressourcen konfrontiert sind, kann dies einen entscheidenden Vorteil bringen, um dennoch umfangreiche UXR-Projekte umzusetzen und wettbewerbsfähig zu bleiben (vgl. Flaherty & Dykes, 2024; Norman, 2013). Aus den Experteninterviews wird klar, dass KI vor allem bei zeitintensiven, operativen Aufgaben wie Transkription, Datenexploration, qualitativer Analyse, Persona-Erstellung oder Berichtserstellung als entlastend wahrgenommen und eingeschätzt wird. Diese Erkenntnisse aus Literatur und Experteninterviews spiegeln sich auch in der Durchführung des Mensch-KI-Vergleichs der vorliegenden Arbeit wider. Für die Transkription eines Usability-Tests benötigte die KI durchschnittlich rund 3 Minuten. Die anschließende KI-gestützte Analyse zur Identifikation von Usability-Problemen dauerte etwa 10 Minuten pro Test. Selbst unter Einbezug zusätzlicher Kontrollroutinen von rund 5 Minuten kann davon ausgegangen werden, dass dieser zeitliche Aufwand deutlich unter dem einer rein menschlichen Analyse liegt (vgl. Aitim & Abdulla, 2024). Zum Vergleich der zeitlichen Aufwände liegen jedoch keine Daten aus der menschlichen Analyse vor. Die weitere Übereinstimmung der KI-Ergebnisse mit der menschlichen Referenzanalyse sowie die Vollständigkeit der KI-Identifikationen werden in Kapitel 5.3 vertieft diskutiert.

Durch die Verkürzung einzelner Prozessschritte entstünden, gemäss den Expert*innen, zudem neue Freiräume, die für die direkte Arbeit mit Nutzer*innen, den fachlichen Austausch oder strategische Diskussionen genutzt werden können. Es kann angenommen werden, dass die Reduktion repetitiver und manueller Tätigkeiten sowie die daraus entstehenden zusätzlichen Kapazitäten zu einer Entlastung der Mitarbeitenden beitragen. Dies wiederum könnte mit positiven Effekten auf die Arbeitszufriedenheit einhergehen (vgl. Brynjolfsson et al., 2025). Auf der anderen Seite weisen die Expert*innen darauf hin, dass durch den Einsatz von KI zwar repetitive Tätigkeiten reduziert werden können, die notwendigen Kontrollroutinen zur Überprüfung der KI-Ergebnisse jedoch als aufwendig, mühsam und frustrierend erlebt werden können (vgl. Lauer et al., 2024).

KI wird in der Literatur als Unterstützung und Treiber von Innovationsprozessen verstanden, der neue Perspektiven eröffnen und kreative Denkanstösse liefern kann (vgl. Zhuhadar & Lytras, 2023). Ergänzend ermöglichen KI-gestützte Analysen die effiziente Exploration grosser Datenmengen sowie die Identifikation von Mustern und Zusammenhängen, die in rein manuellen Analysen möglicherweise übersehen werden würden (vgl. Bano et al., 2024; Humm et al., 2022). Diese Potenziale werden auch von einigen Expert*innen diskutiert. Insbesondere wird von ihnen hervorgehoben, dass KI dazu beitragen kann, die Objektivität von Auswertungen zu verbessern. Da KI-Systeme keine sozialen Berührungspunkte oder Hemmungen aufweisen, können sie Aspekte im Datenmaterial identifizieren, die menschlichen Forscher*innen unter Umständen entgehen könnten.

Ein weiteres Potenzial wird in der Fähigkeit generativer KI-Systeme gesehen, in kurzer Zeit kohärente und sprachlich elaborierte Texte zu erzeugen. Dies erleichtert die Erstellung von Zusammenfassungen, ersten Analyseentwürfen oder Berichten und kann den Zugang zu komplexem Datenmaterial beschleunigen. Zugleich weisen sowohl Literatur als auch Expert*innen darauf hin, dass diese oft überzeugende sprachliche Qualität nicht direkt mit inhaltlicher Korrektheit einhergehen muss. Eine sorgfältige menschliche Überprüfung bleibt daher unerlässlich, um Fehlinterpretationen

zu vermeiden und die Validität der Ergebnisse sicherzustellen (vgl. Farquhar et al., 2024; Fischer, 2023; Ledger & Mancini, 2024).

Trotz der genannten Vorteile beschreibt die Literatur, dass diese Potenziale nicht automatisch durch den blossen Einsatz von KI-Technologie entstehen. Vielmehr hängt der tatsächliche Nutzen stark von den jeweiligen Rahmenbedingungen ab (vgl. Aitim & Abdulla, 2024; Brynjolfsson et al., 2025; Paass & Hecker, 2020). Die ausgeführten Vorteile lassen sich daher nicht pauschal auf jedes Unternehmen oder jedes Einsatzgebiet übertragen. Vielmehr erfordert die Integration von KI-Tools in bestehende Prozesse strategische sowie methodisch klare Überlegungen und Abwägungen, um einen effizienten und sinnvollen Einsatz zu gewährleisten. Diese Einschätzung wird durch die Experteninterviews gestützt, in denen wiederholt auf die Bedeutung von Methodenkompetenz, klaren Zielsetzungen und einer bewussten Aufgabenabgrenzung zwischen Mensch und KI hingewiesen wird.

5.1.2 Herausforderungen und Risiken der KI-Integration im UXR

Ein wesentliches Risiko ergibt sich aus der Funktionsweise grosser Sprachmodelle selbst. Da LLMs, vereinfacht ausgedrückt, auf statistischen Wahrscheinlichkeiten basieren und kein tatsächliches inhaltliches Verständnis besitzen, können sie die Richtigkeit ihrer eigenen Ergebnisse nicht selbst oder nur begrenzt überprüfen (vgl. Ayyamperumal & Ge, 2024; Bano et al., 2024; Gethmann, 2022; Korteling et al., 2021). Besonders relevant ist in diesem Zusammenhang das Phänomen der sogenannten Halluzinationen, welches sowohl in der Literatur als auch in den Experteninterviews als häufige Fehlerquelle in KI-Ergebnissen beschrieben wird (vgl. Bano et al., 2024; Farquhar et al., 2024; Fischer, 2023; Ledger & Mancini, 2024). Damit verbunden ist das sogenannte Pleasing, bei dem KI-Systeme dazu tendieren, überzeugend klingende, jedoch faktisch falsche oder nicht im Datenmaterial begründete Inhalte zu generieren. Dieses Phänomen wurde in den Experteninterviews thematisiert, in der explorierten Literatur jedoch nicht direkt diskutiert. Dadurch wird eine menschliche Kontrolle bzw. Supervision der Ergebnisse zwingend erforderlich, um die Validität der Ergebnisse sicherstellen zu können. Insbesondere im Kontext von Datenanalysen können diese Limitationen ein erhebliches Risiko darstellen, da dadurch fehlerhafte oder verzerrte Informationen in den Analyseprozess einfließen können. Dieses Risiko kann ausserdem zusätzlich verstärkt werden, wenn sich Anwender*innen bzw. Forschende auf KI-generierte Ergebnisse verlassen, ohne den zugrunde liegenden Datensatz ausreichend gut zu kennen, um die KI-Ergebnisse kritisch auf ihre Validität hin überprüfen zu können.

Ein weiteres relevantes Thema betrifft die begrenzte Reproduzierbarkeit sowie die fehlende Transparenz von KI-Ergebnissen. Aufgrund der stochastischen Funktionsweise von LLMs können identische Eingaben zu unterschiedlichen Ergebnissen führen, was die Vergleichbarkeit von Analysen erschwert. Gleichzeitig behindert diese sogenannte Blackbox-Problematik die Nachvollziehbarkeit von KI-Ergebnissen, da zugrunde liegende Entscheidungsprozesse nur eingeschränkt verständlich oder rekonstruierbar sind (vgl. Komera & Manche, 2023; Sarker, 2024). Sowohl die Literatur als auch die Expert*innen weisen darauf hin, dass diese Thematik das

Vertrauen in KI-gestützte Analysen beeinträchtigen kann, was sowohl innerhalb von UX-Teams als auch gegenüber der Kundschaft ein entscheidender Faktor bei der Integration von KI in Unternehmensprozesse darstellt (vgl. Brandao, 2025; Sarker, 2024).

Weiter zeigen sowohl die Literatur als auch die Ergebnisse des Mensch-KI-Vergleichs eine Abhängigkeit der Ergebnisqualität von der zugrunde liegenden Datenbasis sowohl in Bezug auf die Trainingsdaten der KI-Modelle als auch auf den verwendeten Datensätze. Verzerrte oder unvollständige Trainingsdaten können dazu führen, dass gesellschaftliche oder kulturelle Biases nicht nur reproduziert, sondern potenziell auch verstärkt werden (Bano et al., 2024; Bender et al., 2021). Die Expert*innen sprechen in diesem Zusammenhang auch von möglichen Qualitätsproblemen bei Transkripten, etwa bei der Verarbeitung von Dialekten wie Schweizerdeutsch, die sich negativ auf nachgelagerte Analyseschritte auswirken könnten. Ebenfalls wurden Kapazitätsgrenzen der KI-Modelle bei der Verarbeitung grosser Datenmengen thematisiert, welche die Ergebnisqualität beeinflussen können. In den Interviews wurde die Problematik des Vergessen oder Nicht-Berücksichtigen früherer Informationen infolge der Überschreitung von Token-Grenzen als Zustandslosigkeit von Chatbots beschrieben. Zwar können diese Einschränkungen teilweise durch den Einsatz leistungsfähigerer oder kostenpflichtiger Premium-Modelle reduziert werden. Dies ist jedoch mit zusätzlichem finanziellen Aufwänden verbunden und bindet gerade für KMU relevante Ressourcen (vgl. OpenAI, 2025a, 2025b). Vor diesem Hintergrund erscheint eine sorgfältige Kosten-Nutzen-Analyse bei der Integration von KI-Tools in bestehende Prozesse sinnvoll zu sein.

Schliesslich wird die zunehmende Autonomie von KI-Systemen, insbesondere im Kontext agentischer Workflows, kritisch betrachtet. Sowohl Literatur als auch Expert*innen warnen davor, Empfehlungen oder Resultate von KI-Systemen unkritisch zu übernehmen, da diese aufgrund ihrer systemimmanenten Fehleranfälligkeit sowie möglicher Fehlerakkumulation in automatisierten Prozessen verzerrte oder unzuverlässige Ergebnisse liefern können (Brynjolfsson et al., 2025; Brandao, 2025; Li et al., 2024). Eine ausschliessliche Orientierung an synthetischen Nutzer*innen wird kritisch beurteilt, da die Gefahr besteht, dass diese reale menschliche Erfahrungen verzerren oder nur unzureichend abbilden (vgl. Lauer et al., 2024; Li, 2024). Daraus ergeben sich erhöhte Anforderungen an Anwender*innen und Forschende, speziell in Bezug auf kritisches Denken und den reflektierten Umgang mit KI-Ergebnissen. Zudem unterstreicht dies erneut die zentrale Bedeutung guter Datenqualität und verdeutlicht, dass im UXR der direkte Kontakt mit realen Nutzer*innen unverzichtbar bleibt. Ferner sollte eine sorgfältige Balance zwischen Effizienzgewinnen durch KI-Automation und gründlichen, verlässlichen Analyseergebnissen gefunden werden, die den Zielen des Unternehmens entspricht.

Neben technischen und methodischen Herausforderungen verdeutlichen Literatur sowie die Experteninterviews, dass die Integration KI-gestützter Prozesse im UXR mit unternehmens- und kompetenzbezogenen Risiken verbunden ist. Ein erhöhter Einarbeitungs- und Lernaufwand, der hohe zeitliche Aufwand für die Datenaufbereitung sowie ein häufig fehlendes Verständnis der Möglichkeiten und Grenzen von KI können die Integration in Unternehmensprozesse erschweren

(vgl. Li et al., 2024; Wallraff et al., 2023). Diese Unsicherheiten könnten mit internen Widerständen sowie mit Bedenken auf Seiten der Kundschaft in Bezug auf Datenschutz einhergehen. Bei sensiblen Nutzerdaten sind Transparenz, Einwilligung zur KI-Nutzung und konsequente Anonymisierung zentrale Voraussetzungen für den Einsatz von KI. Ebenfalls weisen die Expert*innen darauf hin, dass der rasche technologische Fortschritt der KI-Technologie häufig eher langsamen Implementierungsprozesse in Organisationen gegenüberstehen. Fehlender Austausch über KI-Nutzung, Prompts und Erfahrungen sowie unzureichendes Verständnis der Funktionsweise von KI können dazu führen, dass KI-Ergebnisse über- oder unterschätzt werden. Laut den Expert*innen entsteht das Vertrauen in KI-Ergebnisse vor allem durch transparente Prozesse und klar definierte Kontrollmechanismen sowie eine klare Strategie.

Eng damit verknüpft und ein weiteres relevantes Thema ist die Abhängigkeit von eingesetzten KI-Tools und den dahinterstehenden Unternehmen. Temporäre Ausfälle oder Updates von KI-Modellen können Prozesse verzögern und die Stabilität von Workflows beeinträchtigen, wie auch im Rahmen dieser Arbeit beobachtet werden konnte (vgl. Kapitel 4.2.1). Für KMU kann diese Abhängigkeit eine relevante Hürde bei der KI-Integration darstellen, da alternative Lösungen wie lokal gehostete Modelle zwar mehr Kontrolle und Stabilität ermöglichen könnten, jedoch erhöhte Anforderungen an technische Infrastruktur und zusätzliches internes Fachwissen und Verständnis für komplexe Systeme erfordern.

Wie bereits bei den identifizierten Potenzialen lassen sich auch die Risiken nur teilweise direkt auf KMU bzw. die Praxispartnerin sinnhaft GmbH oder vergleichbare UX-Agenturen übertragen. Dies ist darauf zurückzuführen, dass die Risiken in der vorliegenden Arbeit bewusst auf einer eher allgemeinen Ebene betrachtet werden. Es kann angenommen werden, dass sie je nach eingesetzten KI-Tools sowie der Art ihrer Integration in Unternehmensprozesse variieren können, etwa in Bezug auf kollaborative Mensch-KI-Ansätze mit Fokus auf Zusammenarbeit im Vergleich zu vollständig automatisierten Lösungen. Ziel war es daher, einen möglichst umfassenden Überblick über potenzielle Risiken zu geben, der auf unterschiedliche Anwendungsfelder des KI-Einsatzes übertragbar ist. Die dargestellten Herausforderungen und Risiken dienen entsprechend als Orientierungsrahmen und sollen Denkanstöße liefern, welche Aspekte bei der Integration von KI in der Praxis zu berücksichtigen sind.

5.2 Anforderungen an Anwender*innen bzw. Forschende

Wie im vorherigen Kapitel stützt sich die Beantwortung der zweiten Leitfrage auf die Auswertung der durchgeführten Experteninterviews und auf die herangezogene wissenschaftliche Literatur.

Leitfrage 2: *Welche Anforderungen stellt die Integration von KI in den UX-Research Prozess an die Anwender*innen bzw. Forschenden?*

Auf die unterstützende Rolle der KI sowie auf deren Stärken und Schwächen wurde in Kapitel 5.1 eingegangen. Zur Beantwortung der zweiten Leitfrage rückt der Mensch in den Fokus und adressiert die Anforderungen an Anwender*innen und Forschende im Umgang mit und beim Einsatz von KI. Sowohl die wissenschaftliche Literatur als auch die Expert*innen betonen, dass die Integration von KI nicht zur Abgabe der methodischen und inhaltlichen Verantwortung an automatisierte KI-Systeme einhergehen dürfe. Vielmehr sollen grundlegende Tätigkeiten wie Planung, Steuerung, Interpretation sowie die Verantwortung für die Ergebnisse klar beim Menschen verbleiben (vgl. Brynjolfsson et al., 2025; Gethmann, 2022; Natarajan et al., 2025). Die Rolle der Anwender*innen und Forschenden verschiebt sich somit von der Durchführung einzelner Arbeitsschritte hin zu einer übergeordneten Funktion der Supervision sowie der Kontrolle und Einordnung automatisierter KI-gestützter Ergebnisse. Vor diesem Hintergrund erscheint die Wahl einer kollaborativen Strategie im Sinne eines AI-in-the-Loop- oder HAIC-Ansatzes gegenüber einer vollständigen Automatisierung angezeigt. Besonders im UXR-Kontext, erfordern viele Aufgaben ein differenziertes Verständnis des Nutzungskontexts sowie umfassendes Fachwissen und situatives Urteilsvermögen (vgl. Fragiadakis et al., 2024; Natarajan et al., 2025). Gleichzeitig ermöglicht dieser Ansatz, die komplementären Stärken von Mensch und KI gezielt zu kombinieren und synergetisch zu nutzen (Fragiadakis et al., 2024).

Eine erforderliche Kompetenz von Anwender*innen und Forschenden im Kontext KI-gestützter UXR-Prozesse betrifft die Entwicklung einer KI- und Medienkompetenz. Sowohl die Literatur als auch Expert*innen betonen hierzu, dass ein grundlegendes Verständnis über die Funktionsweise, Fähigkeiten und Limitationen aktueller KI-Modelle erforderlich ist, um deren Einsatz möglichst reflektiert gestalten und begleiten zu können (vgl. Bender et al., 2021; Korteling et al., 2021; Najana et al., 2024; Scott & Zuccon, 2024). Ein vertieftes Verständnis umfasst Kenntnisse über die stochastische Funktionsweise von KI-Modellen wie LLMs sowie über typische Fehlerquellen, wie etwa Halluzinationen, Pleasing oder Verzerrungen durch Trainingsdaten, um KI-Ergebnisse angemessen einordnen und verstehen zu können. Gleichermassen relevant ist das Bewusstsein für die fehlende Transparenz vieler KI-Modelle und die damit verbundene Blackbox-Problematik (vgl. Komera & Manche, 2023). Diese Aspekte sind im UXR äusserst relevant, da sie die Nachvollziehbarkeit und Bewertung von Analyseergebnissen beeinflussen können (vgl. Ayyamperumal & Ge, 2024; Brandao, 2025; Komera & Manche, 2023; Sarker, 2024). Ebenfalls schliesst KI-Kompetenz auch Kenntnisse zu Datenschutz mit ein. Anwender*innen müssen wissen,

wie und wo die eingegebenen Daten verarbeitet und allenfalls gespeichert werden, um daraus datenschutzrechtliche sowie kundenspezifische Vorgaben einhalten zu können.

In Anbetracht der rasanten technologischen Entwicklung von KI verweisen Expert*innen auf die Notwendigkeit einer kontinuierlichen Auseinandersetzung mit neuen KI-Tools und einer fortlaufenden Weiterbildung, um den Einsatz von KI im UXR dauerhaft verantwortungsvoll und methodisch fundiert zu gestalten. Auch in der Literatur wird das Risiko einer schleichenden Erosion methodischer Expertise diskutiert. Es wird vermutet, dass eine starke Abhängigkeit von KI-Assistenzen und die damit verbundene Delegation komplexer kognitiver Aufgaben an KI-Systeme langfristig dazu führen könnte, dass menschliche analytische Fähigkeiten verkümmern könnten (Brynjolfsson et al., 2025; Li et al., 2024). Gerade vor diesem Hintergrund erscheint es wichtig, menschliche Kompetenzen und Fähigkeiten gezielt zu fördern und weiterzuentwickeln sowie bei der Integration von KI in Unternehmensprozesse einen kollaborativen Ansatz einer vollständigen Automatisierung vorzuziehen.

Um KI im UXR effizient und nutzenstiftend einzusetzen sowie die generierten Ergebnisse angemessen überprüfen und bewerten zu können, zeigen sich Methodenwissen und fachliche Erfahrung als weitere zentrale Kompetenzbereiche. Obwohl KI zahlreiche Prozessschritte unterstützen oder teilweise übernehmen kann, ersetzt sie weder grundlegende methodische Überlegungen bei der Planung von UXR noch die finale Interpretation und Einordnung der Ergebnisse. Forschende müssen daher sowohl über ein solides methodisches Verständnis als auch über praktisches Fachwissen verfügen, um KI-gestützte Analysen fundiert einordnen, Ergebnisse kritisch prüfen und Prozesse oder Tests gezielt weiterentwickeln zu können. Die Expert*innen betonen, dass etablierte und gängige Methoden weiterhin die Grundlage für UXR-Prozesse mit KI-gestützten Tools bilden sollten. Auch die Literatur hebt hervor, dass im Umgang mit grossen Datenmengen und KI die Anwendung standardisierter Methoden sowie eine angemessene Auswahl von Werkzeugen in Abhängigkeit von der jeweiligen Phase des Forschungsprozesses entscheidend sein können für die Qualität und den Erfolg von UXR sind (vgl. Aitim & Abdulla, 2024; Bender et al., 2021; Brown, 2024; Brynjolfsson et al., 2025; Fischer, 2023; Li et al., 2024; Paass & Hecker, 2020).

Zur Anwendung von KI-gestützten Prozessen und Interpretation heben sowohl Literatur als auch Expert*innen die Bedeutung eines kritischen und reflektierten Denkens hervor (vgl. Brandao, 2025; Fischer, 2023; Korteling et al., 2021; Najana et al., 2024; Scott & Zucco, 2024). Anwender*innen und Forschende sollten grundsätzlich davon ausgehen, dass KI-Ergebnisse potenziell fehlerhaft sind, und diese konsequent mit dem eigenen Fachwissen sowie dem zugrunde liegenden Datenmaterial abgleichen. Die Überprüfung und Validierung KI-generierter Ergebnisse kann zeitaufwendig und durch wiederholte Iterations- und Korrekturschleifen als herausfordernd empfunden werden. Ein hohes Mass an Frustrationstoleranz erweist sich daher als wichtige Voraussetzung, um diese Prozesse produktiv zu bewältigen.

Da KI-Tools häufig über chatbasierte Schnittstellen mit Anwender*innen interagieren, stellt Prompting einen weiteren wichtigen Kompetenzbereich dar. Einige Expert*innen erwähnen, dass moderne KI-Tools zunehmend One-Click-Funktionalitäten anbieten und Prompting dadurch an Bedeutung verliert. Insbesondere bei der Interaktion mit chatbasierten KI-Tools, wie ChatGPT, das in dieser Arbeit angewendet wurde, wird dringend die Fähigkeit benötigt, Eingaben präzise zu formulieren, ausreichend Kontext bereitzustellen und das gewünschte Ergebnisformat klar zu definieren. Die gezielte Entwicklung und iterative Anpassung von Prompts ermöglicht es, KI-Systeme effektiv zu steuern und qualitativ hochwertige Ergebnisse zu generieren (vgl. Brown, 2024; Stockman, 2025). Um diese Fähigkeiten im Unternehmen zu verankern und die Integration von KI-Tools in bestehende Prozesse zu unterstützen, empfiehlt eine Expertin, Prompting systematisch in organisationsinterne Lernprozesse zu integrieren und erfolgreiche Strategien untereinander im Team zu teilen, um voneinander zu lernen.

5.3 Einsatz von KI im UXR und Validität der KI-Ergebnisse

Im Folgenden werden die Ergebnisse des Mensch-KI-Vergleichs kritisch eingeordnet und mit den Erkenntnissen aus den Experteninterviews sowie der bestehenden Literatur in Beziehung gesetzt, um unter Einbezug der beiden Leitfragen die Hauptforschungsfrage zu beantworten.

Hauptforschungsfrage:

Wie kann Künstliche Intelligenz (KI) genutzt werden, um den UX-Research-Prozess in kleinen Unternehmen (KMU) zu unterstützen und welche Unterschiede bestehen zwischen einer KI-gestützten und einer menschlichen Analyse?

Die Ergebnisse aus Experteninterviews und Literatur zeigen konsistent, dass KI den UXR-Prozess von KMU durch Effizienzsteigerungen, den Ausgleich menschlicher Limitationen, die Unterstützung der qualitativen Analyse grosser Datenmengen sowie die Reduktion zeitlicher und personeller Aufwände sinnvoll unterstützen kann (vgl. Kapitel 5.1.1). Dies ist für UX-Agenturen wie sinnhaft GmbH von besonderer Relevanz, da UXR häufig durch begrenzte Ressourcen eingeschränkt ist und gerade die Analysephase häufig einen wesentlichen Engpass im Produktentwicklungsprozess darstellt (vgl. Aitim & Abdulla, 2024; Flaherty & Dykes, 2024; Najana et al., 2024; Norman, 2013). Gleichzeitig wird betont, dass KI im UXR primär als ergänzendes, kollaboratives Werkzeug verstanden werden soll, das parallel und unterstützend zum Menschen arbeitet. Eine kritische menschliche Überprüfung der KI-Ergebnisse bleibt dabei unerlässlich, um Limitationen bzw. Einschränkungen von KI zu erkennen und auszugleichen (vgl. Kapitel 5.1.2). Die bisherige Literatur legte den Schwerpunkt vor allem auf den Einsatz von KI zur Gestaltung von Benutzeroberflächen und zur Erstellung von Inhalten, während das Potenzial für automatisierte und

remote Usability-Evaluationsverfahren bislang nur begrenzt untersucht worden ist (vgl. Liu, 2025). Die vorliegende Arbeit hat unter anderem diese Forschungslücke adressiert.

Grundsätzlich kann KI praktisch in allen Phasen des UXR-Prozesses eingesetzt werden (vgl. Li et al., 2024; Liu, 2025; Soares et al., 2025). In frühen Phasen unterstützt sie vor allem die Planung und Exploration. Konkret können beispielsweise bestehende Datenquellen analysiert werden, wobei grosse Mengen unstrukturierter Daten systematisch ausgewertet werden, um erste Nutzungsmuster, Bedürfnisse oder Problembereiche zu identifizieren. KI kann ebenfalls unterstützend bei der Vorbereitung, Erstellung und Optimierung von Interviewleitfäden, Fragenkatalogen oder Beobachtungsinstrumenten sowie für Pretests eingesetzt werden (vgl. Li, 2024; Rane et al., 2024).

Ein besonders häufig genannter Anwendungsbereich für UXR betrifft die Aufbereitung und Vorbereitung von Daten. Automatisierte KI-Transkriptionstools ermöglichen eine erhebliche Reduktion des manuellen Aufwands bei der Umwandlung von Audiodaten in eine Textform. Zusätzlich kann KI bei der Bereinigung, Strukturierung und ersten inhaltlichen Aufbereitung der Daten unterstützen, etwa durch Zusammenfassungen oder durch Identifikation relevanter Themen und Abschnitte, für anschliessende Kategorisierungen. Diese Funktionen sind speziell für KMU relevant, da sie zeitintensive Routinetätigkeiten reduzieren und damit Ressourcen für andere Aufgaben schaffen. KI kann auch bei nachgelagerten UXR-Phasen unterstützen, etwa bei der Erstellung von Berichten, der Aufbereitung von Ergebnissen für Kund*innen oder der Generierung von Personas und User Journeys. Generative KI, wie bspw. ChatGPT ermöglicht für diese Aufgabenstellungen eine schnelle sprachliche Strukturierung und Zusammenfassung komplexer Inhalte.

Während der Analysephase können KI-gestützter Tools zur Unterstützung bei der Auswertung qualitativer Daten wie Interviews oder Thinking-Aloud-Transkripten aus Usability-Tests eingesetzt werden. Ebendies wurde auch für die Analyse in der vorliegenden Arbeit genutzt. Sprachbasierte Modelle sind in der Lage, wiederkehrende Muster, Themen oder Stimmungen zu identifizieren und zur Strukturierung umfangreicher Datensätze beizutragen (vgl. Penkert, 2025; Soares et al., 2025; Liu, 2025). Der Einsatz multimodaler KI, zur Analyse visueller Daten, wird in der Literatur zwar beschrieben und in einigen Anwendungsfeldern bereits eingesetzt, befindet sich im UXR-Kontext jedoch bislang überwiegend noch auf einer konzeptionellen oder experimentellen Ebene, vor allem im Hinblick auf die Auswertung von Usability-Testdaten (vgl. Li, 2024; Liu, 2025; OpenAI, 2025a; Zaghdoudi & Glomann, 2021; Zhang et al., 2024). Für diesen spezifischen Anwendungsfall wurden auch in den Experteninterviews keine etablierten KI-Tools genannt. Vielmehr wurde berichtet, dass viele der derzeit verfügbaren KI-Tools, primär das Transkript der Audiospur analysieren und keine umfassende Auswertung visueller Interaktionsdaten ermöglichen. Zudem zeigt die Recherche zu geeigneten KI-Tools für die KI-gestützte Analyse und deren Integration in UXR-Prozesse, dass zahlreiche KI-Lösungen bereits vor der eigentlichen Analysephase ansetzen, indem sie Designs direkt mit künstlichen Personas oder simulierten Nutzer*innen testen und dadurch die Datenerhebung mit realen Nutzer*innen umgehen. Dies macht klassische oder KI-unterstützte

Analyseschritte teilweise obsolet und erschwerte gleichzeitig die Identifikation eines geeigneten Tools für den, in dieser Arbeit vorgesehene, Mensch-KI-Vergleich. Gerade der Einsatz solcher vollständig automatisierter oder KI-gestützter Ansätze, die reale Nutzer*innen durch synthetische Personas oder simulierte Interviews vollständig ersetzen, wurden in den Experteninterviews, wie auch in der Literatur kritisch diskutiert. Es besteht die Gefahr, dass die generierten Ergebnisse verzerrt sind oder reale menschliche Erfahrungen unzureichend abgebildet werden (vgl. Lauer et al., 2024; Li, 2024). Für die KI-Integration in ein KMU erscheint daher ein AI-in-the-Loop-Ansatz sinnvoll, bei dem KI als unterstützendes Werkzeug eingesetzt wird, während Verantwortung, Interpretation und Bewertung der Ergebnisse weiterhin dem Menschen obliegen (Natarajan et al., 2025; Stockman, 2025).

Die Auswertung des Mensch-KI-Vergleichs bei der Analyse von Usability-Tests zur Identifikation von Usability-Problemen, ergibt ein differenziertes Bild der Leistungsfähigkeit KI-gestützter Analyseverfahren. Die Ergebnisse dieser KI-Analyse zeigen einen konstanten Anstieg von Recall und Precision über die drei Analysearten vom Rohtranskript (ROH) über die bereinigte Analyse (BER) bis hin zur kontextualisierten Analyse (KON). Dies gibt Hinweise darauf, dass die Übereinstimmung zwischen KI-Analyse und menschlicher Referenz unter anderem vom Grad der Datenaufbereitung und der Datenqualität abhängen könnte. Diese Beobachtung steht im Einklang mit den Erwartungen und Einschätzungen der befragten Expert*innen. Möglicherweise könnten jedoch weitere Faktoren, wie das eingesetzte KI-Modell, dessen Auslastung sowie das verwendete Prompting eine Rolle spielen. Für die KI-Analysen der vorliegenden Arbeit auf Grundlage der Roh- und bereinigten Transkripte wurden daher identische Prompts verwendet. In der kontextualisierten Analyse wurde derselbe Prompt verwendet, ergänzt lediglich um Informationen zu den zusätzlichen Kontextinformationen aus den Bildschirmaufnahmen während der Usability-Tests. Ein weiterer Einflussfaktor auf die KI-Ergebnisse liegt in der Durchführung der Usability-Tests selbst. Hypothetische Formulierungen der Testpersonen (z. B. «ich würde hier klicken») oder eine nicht konsequent umgesetzte Thinking-Aloud-Methode bei der Datenerhebung können die Interpretation durch die KI zusätzlich erschweren und zu möglichen Fehlklassifikationen führen.

Trotz dieser stetigen Verbesserungen von Recall und Precision über die drei Analysearten hinweg erreicht die KI auch in der kontextualisierten Analyse keine vollständige Übereinstimmung mit der menschlichen Analyse. Selbst unter den günstigsten Bedingungen werden rund 67 %, der durch den Menschen identifizierten Usability-Probleme, durch die KI korrekt erkannt. Im ungünstigsten Fall sind es mit 51% jedoch nur etwa die Hälfte. Für bestimmte praxisorientierte Fragestellungen könnte die erzielte Identifikationsleistung der KI dennoch ausreichend sein, insbesondere wenn die Geschwindigkeit gegenüber der menschlichen Analyse priorisiert wird. Die Ergebnisse deuten somit auf die Notwendigkeit einer bewussten Abwägung zwischen Schnelligkeit der KI-gestützten Auswertung und der gewünschten analytischen Tiefe.

In allen Analysearten traten zudem Falsch-Negative (FN) Fälle auf. Also Usability-Probleme, die von der KI nicht erkannt wurden. Zwar nimmt deren Anzahl von ROH über BER, bis hin zu KON

kontinuierlich ab, dennoch machen sie auch in KON weiterhin rund einen Drittel der, durch den Menschen identifizierten Usability-Probleme aus. Die Ergebnisse zeigen zudem, dass die Verbesserungen von Recall und Precision zwischen BER und KON deutlich geringer ausfallen als der Zuwachs zwischen ROH und BER bzw. KON. Daraus lässt sich ableiten, dass der zusätzliche Nutzen der Kontextualisierung bzw. dem Ergänzen der zusätzlichen Kontextinformationen aus den Bildschirmaufzeichnungen, im Vergleich zur einfachen Transkriptbereinigung als gering eingeschätzt werden kann. Vor diesem Hintergrund stellt sich deshalb die Frage, ob für bestimmte Anwendungsfälle die Analyse eines qualitativ hochwertigen und inhaltlich vollständigen Transkripts bereits ausreichend sein könnte und eine zusätzliche Auswertung der Bildschirmaufzeichnungen nicht zwingend erforderlich wäre. In Verbindung mit den bisherigen Erkenntnissen zur Relevanz der Datenqualität kann davon ausgegangen werden, dass der Mehrwert visueller Informationen abnimmt, je strukturierter, vollständiger und aussagekräftiger die erhobenen Daten sind. Entsprechend lässt sich schlussfolgern, dass eine hohe Qualität der Datenerhebung und -aufbereitung, eine relevante Voraussetzung für eine erfolgreiche und nutzenstiftende Integration von KI in UXR-Prozesse darstellt.

Eine weitere Erkenntnis betrifft die Falsch-Positiven Identifikationen. Die nachträgliche qualitative Einordnung zeigt, dass ein Teil, der als Falsch-Positive klassifizierten KI-Ergebnisse inhaltlich plausibel ist und potenzielle Usability-Probleme adressiert, welche von den menschlichen Expert*innen nicht explizit benannt oder ganz übersehen wurden. Auffällig ist dabei, dass die absolute Anzahl dieser als plausibel eingeschätzten Identifikationen, über alle Analysearten hinweg (ROH, BER, KON) auf einem vergleichbaren Niveau bleibt, während die Gesamtzahl der Falsch-Positiven mit zunehmender Aufbereitung der Daten deutlich abnimmt. Dies legt nahe, dass die KI im vorliegenden Fallbeispiel nicht zufällig zusätzliche Befunde generiert bzw. halluziniert, sondern tatsächlich plausible, wiederkehrende Muster erkennt, welche in der menschlichen Referenzanalyse teilweise unberücksichtigt geblieben sind. Diese Erkenntnis deckt sich mit den Aussagen aus den Experteninterviews, wonach KI Aspekte oder Muster im Datenmaterial identifizieren kann, die menschlichen Forscher*innen unter Umständen entgehen. Insgesamt verdeutlicht dies den Mehrwert eines kollaborativen Ansatzes bei der Integration von KI, bei dem menschliche Expertise und KI-gestützte Analyse kombiniert werden, um die jeweiligen Potenziale beider Seiten bestmöglich zu nutzen und bekannte Limitationen auszugleichen.

In Bezug auf den zeitlichen Aufwand zeigt sich, dass die KI-Analyse vergleichsweise schnell durchgeführt werden kann. In der vorliegenden Arbeit nahm die KI-Analyse durchschnittlich rund zehn Minuten pro Usability-Test in Anspruch. Die zuvor diskutierten Vorteile von Effizienz- und Zeitgewinn zeigen sich entsprechend auch in der untersuchten Fallstudie und der Evaluation des KI-gestützten Analyseprozesses. Gleichzeitig relativiert sich dieser Vorteil teilweise durch den zusätzlichen Aufwand für notwendige Schulungen für betroffene Mitarbeitende der KI-Integration, Datenaufbereitung sowie Kontrolle der KI-Ergebnisse. Durch eine vollständige Automatisierung einzelner Prozessschritte durch die KI, etwa durch agentische Workflows, liesse sich dieser

Zeitaufwand möglicherweise noch weiter reduzieren. Allerdings könnte dies die Nachvollziehbarkeit und Kontrolle weiter einschränken. Vor dem Hintergrund bekannter struktureller Fehleranfälligkeiten aktueller KI-Modelle erscheint eine solche Automatisierung derzeit jedoch kritisch.

6 Fazit und Ausblick

Zusammenfassend zeigt sich, dass KI den UXR-Prozess von KMU durch potenzielle Effizienz- und Zeitgewinne, die Unterstützung diverser Arbeitsschritte und der damit verbundenen Reduktion personeller Aufwände sinnvoll unterstützen kann. Dies betrifft vornehmlich die Datenaufbereitung, die Exploration und Analyse umfangreicher Datensätze und den Ausgleich menschlicher Limitationen. Ausserdem kann KI neue Perspektiven eröffnen und kreative Denkanstösse anregen sowie dabei helfen sprachlich elaborierte Ergebnisberichte zu schreiben. Die Möglichkeiten der KI-Integration scheinen schier unbegrenzt zu sein.

Gleichzeitig wird deutlich, dass der Mehrwert der KI-Integration stark vom jeweiligen Anwendungskontext sowie von den jeweiligen Aufgabenstellungen und Zielsetzungen bestimmt wird. Dazwischen muss ein angemessenes Gleichgewicht gefunden werden, wobei auch potenzielle Risiken und Limitationen zu berücksichtigen sind. Dazu zählen die fehleranfällige und teilweise intransparente Funktionsweise von KI-Systemen, die begrenzte Reproduzierbarkeit und Nachvollziehbarkeit von Ergebnissen sowie die Abhängigkeit von den eingesetzten KI-Tools und den dahinterstehenden Unternehmen. Letztere ist zudem tendenziell mit zusätzlichen finanziellen Aufwänden durch kostenpflichtige Premium-Modelle verbunden.

Ferner zeigt sich, dass die Integration von KI im UXR mit neuen Anforderungen an die Anwender*innen und Forschenden einhergeht. Dazu zählt die Übernahme einer neuen Rolle der Supervision und Kontrolle KI-gestützter Prozesse, welche ihrerseits ein fundiertes Verständnis der Funktionsweise, Stärken und Limitationen aktueller KI-Modelle voraussetzt. Nicht zuletzt sind ausgeprägte Methoden- und Fachkompetenzen zur kritischen Einordnung der Ergebnisse, sowie die Fähigkeit zu einem reflektierten und verantwortungsvollen Umgang mit potenziell fehlerhaften KI-Ergebnissen erforderlich, einschliesslich der Berücksichtigung von Datenschutzaspekten. Über die genannten menschlichen Kompetenzen hinweg besteht die Gefahr, sie schrittweise an automatisierte KI-Systeme zu delegieren und dadurch auf langfristige Sicht zu schwächen oder gar zu verlieren. Dies macht zugleich deutlich, wie wichtig die gezielte Förderung des Erlernens und der nachhaltige Erhalt methodischer sowie grundlegender menschlicher Kompetenzen sind. Aufbauend auf den zuvor diskutierten Ergebnissen, empfiehlt die vorliegende Arbeit einen kollaborativen Ansatz für die Integration von KI, bei welchem Mensch und KI partnerschaftlich zusammenarbeiten und ihre jeweiligen komplementären Stärken gezielt und gewinnbringend kombiniert werden können.

KI ermöglicht Effizienzgewinne, eröffnet neue analytische und kreative Perspektiven, und kann menschliche Limitationen teilweise ausgleichen. Sie ersetzt jedoch weder menschliche Fähigkeiten, wie Erfahrung, Kreativität, situatives Urteilsvermögen oder Kommunikationsfähigkeit, noch die Verantwortung und kritische Bewertung der Ergebnisse durch den Menschen. Eine bewusste Balance zwischen Automatisierung und dem Einsatz menschlicher Fähigkeiten und Kontrolle sollte daher die Grundlage für einen nachhaltigen und verantwortungsvollen Einsatz von KI im UXR sowie in anderen Kontexten bilden.

Die Ergebnisse des Mensch-KI-Vergleichs verdeutlichen zugleich, dass sich KI-gestützte Analysen in ihrer Leistungsfähigkeit von menschlichen Analysen unterscheiden, und unter anderem von der Datenqualität sowie dem Aufbereitungsgrad der Daten geprägt sein dürften. Zwar ist die KI

in der Lage, in kurzer Zeit einen relevanten Anteil menschlich identifizierter Usability-Probleme zu erkennen und überdies zusätzliche plausible Usability-Probleme zu identifizieren, sie erreicht jedoch keine vollständige Übereinstimmung mit der menschlichen Analyse. Für bestimmte Anwendungsfälle in der Praxis könnte diese Identifikationsleistung ausreichen, hauptsächlich wenn Geschwindigkeit gegenüber analytischer Tiefe im Vordergrund steht. Entscheidend ist, im Vorfeld klar zu definieren, wie zuverlässig und umfassend die KI-generierten Ergebnisse sein müssen, und die KI-Integration auf dieser Grundlage strategisch und reflektiert für spezifische Aufgaben und Anwendungsfelder zu planen. Zur Sicherstellung von Kontrolle und Vertrauen in die Ergebnisse wird empfohlen, auf methodisch gut etablierte und validierte Verfahren bzw. Methoden zurückzugreifen und eine konsequente menschliche Kontrolle in den Prozess zu integrieren.

Die vorliegende Arbeit leistet einen Beitrag zur Bearbeitung einer Forschungslücke und liefert zugleich wissenschaftliche und praxisrelevante Implikationen für den UXR-Kontext bei der Integration von KI in Unternehmensprozesse. Der Schwerpunkt liegt dabei auf den daraus resultierenden Anforderungen an Anwender*innen bzw. Forschende, den Potenzialen und Risiken sowie auf konkreten Mess- und Vergleichswerten zwischen menschlicher und KI-gestützter Analyse von Usability-Tests. Dazu wurden drei Forschungsfragen explorativ bearbeitet und beantwortet. Trotz der daraus gewonnenen Erkenntnisse bleiben dennoch einige Fragen offen, welche von zukünftiger Forschung adressiert und untersucht werden könnten. Es wurde nicht untersucht, ob es systematische Muster hinsichtlich jener Usability-Probleme gibt, die durch KI nicht erkannt werden, und inwieweit sich die Identifikationsleistung durch gezielte Prompting-Strategien oder den Einsatz tatsächlich multimodaler KI weiter verbessern liesse. Weiter sind die im Mensch-KI-Vergleich beobachteten Effekte ausschliesslich deskriptiver Natur und stellen keine statistischen Korrelationen bzw. Abhängigkeiten dar. Sie können daher lediglich als praxisorientierte Implikationen und erste explorative Hinweise verstanden werden. Entsprechend sind die Ergebnisse mit Vorsicht zu interpretieren und müssten in weiterführenden Studien mithilfe geeigneter statistischer Verfahren empirisch überprüft werden. Darüber hinaus fehlen bislang empirische Erkenntnisse zur Leistungsfähigkeit vollständig automatisierter bzw. agentischer Workflows im Vergleich zu kollaborativen Mensch-KI-Ansätzen im UXR. Offen bleibt ebenfalls, ob die in dieser Arbeit beobachtete Leistungsfähigkeit zur Erkennung von Usability-Problemen für den praktischen Einsatz tatsächlich ausreicht. Entscheidend bleibt eine unternehmensspezifische Abwägung zwischen Geschwindigkeit der Auswertung und erwarteter analytischer Tiefe.

7 Limitationen

Die vorliegende Arbeit weist mehrere Limitationen auf, die bei der Einordnung der Ergebnisse zu berücksichtigen sind.

Zuallererst ein Transparenzhinweis zur Durchführung der vorliegenden Arbeit. Die Kosten für den Einsatz der KI-Tools wurden von der Praxispartnerin übernommen. Die Praxispartnerin hatte jedoch keinen direkten Einfluss auf die Auswahl der eingesetzten KI-Tools. Diese erfolgte auf Grundlage einer Analyse der bestehenden Unternehmensprozesse sowie mit Fokus auf KI-Tools, die entweder bereits im Unternehmen verwendet wurden oder sich sinnvoll in bestehende Workflows integrieren liessen. Die Auswahl der KI-Tools war somit praxis- und kontextbezogen, weshalb keine systematische Evaluation eines breiten Spektrums verfügbarer KI-Lösungen vorgenommen wurde.

Eine weitere Limitation ergibt sich aus der ursprünglichen Studienplanung. Zu Beginn der Arbeit wurde mit der Praxispartnerin vereinbart, qualitative Interviewdaten zu analysieren. Aufgrund von Datenschutzbedenken und mangels passender Datensätze konnte dieses Vorhaben jedoch nicht umgesetzt werden, sodass stattdessen auf einen Datensatz bestehend aus Usability-Tests zurückgegriffen wurde. Zu diesem Zeitpunkt waren die Experteninterviews bereits durchgeführt und inhaltlich auf die Analyse qualitativer Interviewdaten ausgerichtet. Entsprechend konnten die Erkenntnisse aus den Experteninterviews nur eingeschränkt auf den konkreten Kontext der Usability-Test-Analyse übertragen werden. Zwar stehen durch Usability-Tests auch qualitative Daten zur Verfügung, dennoch ist die Aussagekraft der Experteninterviews für diesen spezifischen Anwendungsfall begrenzt.

Die externe Validität der Ergebnisse ist ebenfalls eingeschränkt. Die vorliegende Arbeit ist als Fallstudie konzipiert und bezieht sich auf den konkreten Kontext der UX-Agentur sinnhaft GmbH. Obwohl dieser Fall beispielhaft für vergleichbare kleine UX-Agenturen bzw. KMU stehen kann, lassen sich die Ergebnisse nicht zwingend direkt auf andere Organisationsformen oder Branchen übertragen. Eine weitere Limitierung besteht auch darin, dass lediglich eine KI-Lösung (ChatGPT) sowie eine spezifische Analysemethode (Usability-Tests) untersucht wurden.

Weiter ist zu berücksichtigen, dass sich KI-Technologien und der Markt für KI-Tools derzeit sehr dynamisch entwickeln. Modelle sowie Funktionen unterliegen kontinuierlichen Veränderungen, sodass die in dieser Arbeit gewonnenen Ergebnisse nur als Momentaufnahme verstanden werden können. Aussagen zur Leistungsfähigkeit einzelner KI-Tools sind daher zeitlich begrenzt gültig.

Eine methodische Einschränkung betrifft zudem die begrenzte Fähigkeit aktueller KI-Tools zur Verarbeitung visueller Daten, was gerade im Kontext der Analyse von Usability-Tests interessant gewesen wäre. Die untersuchten Lösungen analysieren primär Textdaten oder Transkripte. Zu erwähnen ist hier auch, dass viele KI-Tools setzen bereits vor der eigentlichen UXR-Analysephase ansetzen, indem sie Designs direkt mit künstlichen Personas oder simulierten Nutzer*innen evaluieren und damit auf die Datenerhebung mit realen Nutzer*innen verzichten.

Eine weitere methodische Limitation besteht darin, dass die menschliche Analyse als Referenzwert für den Mensch-KI-Vergleich herangezogen wurde. Obwohl dieser Referenzwert auf

der Analyse von zwei UX-Expert*innen basiert, ist davon auszugehen, dass die menschlichen Usability-Analysen nicht frei von Subjektivität oder nicht erfassten Aspekten sind. Dieser Einschränkung wurde teilweise durch eine nachfolgende qualitative Einordnung jener Usability-Probleme begegnet, die ausschliesslich durch die KI identifiziert wurden. Auch ist zu berücksichtigen, dass der verwendete Datensatz ursprünglich nicht mit dem Ziel erhoben wurde, KI-gestützt analysiert zu werden, da der Einsatz von KI zum Zeitpunkt der Datenerhebung noch nicht vorgesehen war. Dies könnte die Passung der Daten für eine KI-gestützte Auswertung limitiert haben. Der Abgleich und die Vergleichsanalyse zwischen menschlicher und KI-Analyse erfolgten manuell. In diesem Schritt könnten erneut subjektive Interpretationen in die Auswertung eingeflossen sein, die sich nicht vollständig kontrollieren liessen. Zudem erlaubt die Analyse der verwendeten Kennzahlen (RP, FP, FN, Recall, Precision) keine Ableitung statistischer Korrelationen oder Kausalitäten. Die beobachteten Zusammenhänge und Effekte sind daher als deskriptive Hinweise zu verstehen und nur eingeschränkt interpretierbar. Es bedarf einer empirischen Überprüfung in weiterführenden Studien mittels geeigneter statistischer Verfahren, um die beobachteten Effekte verlässlich einordnen zu können.

Zusammenfassend sind darum die Ergebnisse dieser Arbeit kontextabhängig zu interpretieren und dienen primär für praktische Implikationen zur Integration von KI im UXR sowie als explorative Grundlage für weiterführende Forschung.

Literaturverzeichnis

- Aitim, A. & Abdulla, M. (2024). Data processing and analysing techniques in UX research. *Procedia Computer Science*, 251, 591–596. <https://doi.org/10.1016/j.procs.2024.11.154>
- Ayyamperumal, S. G. & Ge, L. (2024, Juni 16). *Current state of LLM Risks and AI Guardrails*. arXiv. Manuscript submitted for publication. <https://doi.org/10.48550/arXiv.2406.12934>
- Bano, M., Zowghi, D. & Whittle, J. (2024). AI and Human Reasoning: Qualitative Research in the Age of Large Language Models. *AI Ethics Journal*, 4. <https://doi.org/10.47289/AIEJ20240122>
- Bender, E. M., Gebru, T., McMillan-Major, A. & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (S. 610–623). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
- Berni, A. & Borgianni, Y. (2021). Making Order in User Experience Research to Support Its Application in Design and Beyond. *Applied Sciences*, 11(15), 6981. <https://doi.org/10.3390/app11156981>
- Brandao, P. R. (2025). The Impact of Artificial Intelligence on Modern Society. *AI*, 6(8), 190. <https://doi.org/10.3390/ai6080190>
- Brown, N. B. (2024, Juni 4). *Enhancing Trust in LLMs: Algorithms for Comparing and Interpreting LLMs*. arXiv. Manuscript submitted for publication. <https://doi.org/10.48550/arXiv.2406.01943>
- Brüsemeister, T. (2008). Qualitative Forschung: Ein Überblick. In T. Brüsemeister (Hrsg.), *Qualitative Forschung: Ein Überblick* (S. 53–228). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91182-3_2
- Brynjolfsson, E., Li, D. & Raymond, L. (2025). Generative AI at Work*. *The Quarterly Journal of Economics*, 140(2), 889–942. <https://doi.org/10.1093/qje/qjae044>
- Bundesamt für Statistik (BFS). (2024). Kleine und mittlere Unternehmen. *Statistiken: Kleine und mittlere Unternehmen*. Verfügbar unter: <https://www.bfs.admin.ch/content/bfs/de/home/statistiken/industriedienstleistungen/unternehmen-beschaefigte/wirtschaftsstruktur-unternehmen/kmu.html>
- Deutsches Institut für Normung e.V. (2010). EN ISO 9241-210. (ISO 9241-210:2010). (2010). *Ergonomie der Mensch-System-Interaktion: Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme deutsche Fassung*. Berlin: Beuth.
- Fantapié Altobelli, C. (2023). *Marktforschung - Methoden - Anwendungen - Praxisbeispiele* (4. überarb. u. erw. Aufl.). Tübingen: UVK Verlag. <https://elibrary.utb.de/doi/epdf/10.36198/9783838588162>
- Farquhar, S., Kossen, J., Kuhn, L. & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature*, 630(8017), 625–630. <https://doi.org/10.1038/s41586-024-07421-0>
- Fischer, G. (2023). A Research Framework Focused on AI and Humans instead of AI versus Humans. *Interaction Design and Architecture(s)*, (59), 17–36. <https://doi.org/10.55612/s-5002-059-001sp>

- Flaherty, K. & Dykes, T. (2024). *The Biggest Challenges Practitioners Encounter Working in UX - Nielsen Norman Group*. Verfügbar unter: <https://www.nngroup.com/articles/ux-challenges/>
- Flick, U. (2019). *Qualitative Sozialforschung. Eine Einführung* (9. Auflage). Reinbek b. Hamburg: Rowohlt.
- Fragiadakis, G., Diou, C., Kousiouris, G. & Nikolaidou, M. (2024, Juli 9). *Evaluating Human-AI Collaboration: A Review and Methodological Framework*. Manuscript submitted for publication. <https://doi.org/10.48550/arXiv.2407.19098>
- Gardey, J. C., Grigera, J., Rodriguez, A. & Garrido, A. (2024). UX-Analyzer: Visualizing the interaction effort for web analytics. In *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing* (S. 1774–1780). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3605098.3636013>
- Geis, T. & Tesch, G. (2023). *Basiswissen Usability und User Experience: Aus- und Weiterbildung zum UXQB Certified Professional for Usability and User Experience (CPUX) - Foundation Level (CPUX-F)* (2., überarbeitete und aktualisierte Auflage.). Heidelberg: dpunkt.verlag.
- Gethmann, C. F. (2022). Zur Frage der Ersetzbarkeit des Menschen durch KI in der Forschung. In C.F. Gethmann, P. Buxmann, J. Distelrath, B.G. Humm, S. Lingner, V. Nitsch et al. (Hrsg.), *Künstliche Intelligenz in der Forschung: Neue Möglichkeiten und Herausforderungen für die Wissenschaft* (S. 43–78). Berlin: Springer. https://doi.org/10.1007/978-3-662-63449-3_3
- Goethals, S., Luther, J. & Matz, S. (2025). Words reveal wants: How well can simple LLM-based AI agents replicate people's choices based on their social media posts. In *Adjunct Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization* (S. 126–131). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3708319.3733689>
- Graner, M. (2015). *Methodeneinsatz in der Produktentwicklung: Bessere Produkte, schnellere Entwicklung, höhere Gewinnmargen*. Wiesbaden: Springer Fachmedien. <https://doi.org/10.1007/978-3-658-08582-7>
- Hassenzahl, M. (2003). The Thing and I: Understanding the Relationship Between User and Product (Human-Computer Interaction Series). In M.A. Blythe, K. Overbeeke, A.F. Monk & P.C. Wright (Hrsg.), *Funology* (Band 3, S. 31–42). Dordrecht: Springer Netherlands. https://doi.org/10.1007/1-4020-2967-5_4
- Hassenzahl, M. & Tractinsky, N. (2006). User experience - a research agenda. *Behaviour & Information Technology*, 25(2), 91–97. <https://doi.org/10.1080/01449290500330331>
- Helfferrich, C. (2011). *Die Qualität qualitativer Daten: Manual für die Durchführung qualitativer Interviews* (4. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften
- Helfferrich, C. (2022). Leitfaden- und Experteninterviews. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 875–892). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-37985-8_55
- Hokkanen, L. & Väänänen-Vainio-Mattila, K. (2015). UX Work in Startups: Current Practices and Future Needs. In C. Lassenius, T. Dingsøyr & M. Paasivaara (Hrsg.), *Agile Processes in*

- Software Engineering and Extreme Programming* (S. 81–92). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-18612-2_7
- Honegger, L. (2023). *Wirtschaftlichkeit von Usability-Evaluationsmethoden*. Hochschule für Angewandte Psychologie FHNW. <https://doi.org/10.26041/FHNW-4908>
- Humm, B. G., Buxmann, P. & Schmidt, J. C. (2022). Grundlagen und Anwendungen von KI. In C.F. Gethmann, P. Buxmann, J. Distelrath, B.G. Humm, S. Lingner, V. Nitsch et al. (Hrsg.), *Künstliche Intelligenz in der Forschung: Neue Möglichkeiten und Herausforderungen für die Wissenschaft* (S. 13–42). Berlin: Springer. https://doi.org/10.1007/978-3-662-63449-3_2
- Jacobsen, J. & Meyer, L. (2019). *Praxisbuch Usability und UX* (2. aktualisierte Auflage). Bonn: Rheinwerk Verlag.
- Kieffer, S., Rukonić, L., Kervyn de Meerendré, V. & Vanderdonckt, J. (2019). Specification of a UX Process Reference Model towards the Strategic Planning of UX Activities. In *14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* (S. 74–85). Prag: HUCAPP. <https://doi.org/10.5220/0007693600740085>
- Knearem, T., Khwaja, M., Gao, Y., Bentley, F. & Kliman-Silver, C. E. (2023). Exploring the future of design tooling: The role of artificial intelligence in tools for user experience professionals. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (S. 1–6). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3544549.3573874>
- Komera, O. & Manche, R. (2023). Black-Box Behavior in Large Language Models: Challenges and Implications. In *Conference: International Conference of Renewable Energy, Green Computing and Sustainable Development*. Hyderabad
- Kondapaka, P., Khanra, S., Malik, A., Kagzi, M. & Hemachandran, K. (2023). Finding a fit between CXO's experience and AI usage in CXO decision-making: evidence from knowledge-intensive professional service firms. *Journal of Service Theory and Practice*, 33(2), 280–308. <https://doi.org/10.1108/JSTP-06-2022-0134>
- Korteling, J. E., van de Boer-Visschedijk, G. C., Blankendaal, R. a. M., Boonekamp, R. C. & Eikelboom, A. R. (2021). Human- versus Artificial Intelligence. *Frontiers in Artificial Intelligence*, 4. <https://doi.org/10.3389/frai.2021.622364>
- Kuckartz, U. & Rädiker, S. (2020). *Fokussierte Interviewanalyse mit MAXQDA: Schritt für Schritt*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-31468-2>
- Kunz, W. H., Sajtos, L. & Flavián, C. (2025). Beyond replacement: human-machine collaboration in the age of AI. *Journal of Service Management*, 36(4), 477–494. <https://doi.org/10.1108/JOSM-04-2025-0194>
- Kusuma, W. A., Jantan, A. H., Admodisastro, N. I. & Binti Mohd Norowi, N. (2024). Toward Adoption User Experience Variables for Solo Software Development in Academic and Industry. *International Journal on Informatics Visualization*, 8(4), 2090. <https://doi.org/10.62527/joiv.8.2.2172>

- Lauer, C., Storey, D. & Soley, R. (2024). Vector Personas: How UX Researchers Can Use AI to Bring New Dimension to Traditional Persona Development. In *Proceedings of the 42nd ACM International Conference on Design of Communication* (S. 151–157). Fairfax, VA: Association for Computing Machinery. <https://doi.org/10.1145/3641237.3691663>
- Ledger, G. & Mancinni, R. (2024). Detecting LLM Hallucinations Using Monte Carlo Simulations on Token Probabilities. *TechRxiv*. <https://doi.org/10.36227/techrxiv.171822396.61518693/v1>
- Lewis, J. R. & Sauro, J. (2021). Usability and User Experience: Design and Evaluation. *HANDBOOK OF HUMAN FACTORS AND ERGONOMICS* (S. 972–1015). Hoboken, NJ: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119636113.ch38>
- Li, J. (2024). How Far Can We Go with Synthetic User Experience Research? *interactions*, 31(3), 26–29. <https://doi.org/10.1145/3653682>
- Li, Z., Wang, Y.-L., Zhang, B. & Klein, A. (2024). The Evolving Landscape of UX Research in Industry: How AI is Transforming Manual Practices. In *Mensch und Computer 2024 - Usability Professionals*. Karlsruhe: Gesellschaft für Informatik e.V.
- Licklider, L. (1960). Man–computer symbiosis. *IRE Transactions on Human Factors in Electronics*, 4, 4–11.
- Liu, J. (2025). AI-Powered Automated and Remote UX Evaluation Methods: A Systematic Literature Review. In *Proceedings of the 43rd ACM International Conference on Design of Communication* (S. 10–16). Lubbock, TX: ACM. <https://doi.org/10.1145/3711670.3764614>
- Manning, B. S., Zhu, K. & Horton, J. J. (2024, April 25). *Automated Social Science: Language Models as Scientist and Subjects*. arXiv. Manuscript submitted for publication. <https://doi.org/10.48550/arXiv.2404.11794>
- Marti, C. L., Liu, H., Kour, G., Bilgihan, A. & Xu, Y. (2024). Leveraging artificial intelligence in firm-generated online customer communities: a framework and future research agenda. *Journal of Service Management*, 35(3), 438–458. <https://doi.org/10.1108/JOSM-10-2023-0443>
- Martinelli, S., Lopes, L. & Zaina, L. (2024). UX Research practices related to Long-Term UX: A Systematic Literature Review. *Information and Software Technology*, 170, 107431. <https://doi.org/10.1016/j.infsof.2024.107431>
- Meuser, M. & Nagel, U. (2009). Das Experteninterview - konzeptionelle Grundlagen und methodische Anlage. In S. Pickel, G. Pickel, H.-J. Lauth & D. Jahn (Hrsg.), *Methoden der vergleichenden Politik- und Sozialwissenschaft: Neue Entwicklungen und Anwendungen* (S. 465–479). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91826-6_23
- Najana, M., Bhattacharya, S., Kewalramani, C. & Pandiya, D. K. (2024). AI and Organizational Transformation: Navigating the Future. *International Journal of Global Innovations and Solutions*. <https://doi.org/10.21428/e90189c8.03fab010>
- Natarajan, S., Mathur, S., Sidheekh, S., Stammer, W. & Kersting, K. (2025). Human-in-the-loop or AI-in-the-loop? Automate or Collaborate? In *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(27), 28594–28600. <https://doi.org/10.1609/aaai.v39i27.35083>

- Neha, F., Bhati, D., Shukla, D. K. & Amiruzzaman, M. (2024, Dezember 6). *From classical techniques to convolution-based models: A review of object detection algorithms*. arXiv. Manuscript submitted for publication. <https://doi.org/10.48550/arXiv.2412.05252>
- Norman, D. A. (2013). *The design of everyday things* (Rev. and expanded ed.). Cambridge, MA: MIT press.
- OpenAI. (2025b). *Entdecke GPT-5*. Verfügbar unter: <https://openai.com/de-DE/index/introducing-gpt-5/>
- OpenAI. (2025c). *Entdecke GPT-5.2*. Verfügbar unter: <https://openai.com/de-DE/index/introducing-gpt-5-2/>
- OpenAI. (2025a). *Sycophancy in GPT-4o: What happened and what we're doing about it*. Verfügbar unter: <https://openai.com/index/sycophancy-in-gpt-4o/>
- Paass, G. & Hecker, D. (2020). *Künstliche Intelligenz: Was steckt hinter der Technologie der Zukunft?* Wiesbaden: Springer Fachmedien. <https://doi.org/10.1007/978-3-658-30211-5>
- Penkert, L. (2025). AI-moderated Interviews as an Alternative to Surveys?. In *Mensch und Computer 2025 - Usability Professionals*. Karlsruhe: Gesellschaft für Informatik e.V.
- Rädiker, S. & Kuckartz, U. (2019). *Analyse qualitativer Daten mit MAXQDA: Text, Audio und Video*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-22095-2>
- Rane, N. L., Choudhary, S. P. & Rane, J. (2024). Artificial Intelligence and Machine Learning in Business-toBusiness (B2B) Sales and Marketing: A Review. *International Journal of Data Science and Big Data Analytics*, 4(1), 17–33.
- Redler, J. & Ullrich, S. (2021). *Marketing klipp & klar (WiWi klipp & klar)*. Wiesbaden: Springer Fachmedien. <https://doi.org/10.1007/978-3-658-34538-9>
- Richter, M. & Flückiger, M. D. (2016). *Usability und UX kompakt*. Berlin: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-49828-6>
- Ridder, H.-G. (2017). The theory contribution of case study research designs. *Business Research*, 10(2), 281–305. <https://doi.org/10.1007/s40685-017-0045-z>
- Robinson, J., Lanius, C. & Weber, R. (2018). The past, present, and future of UX empirical research. *Communication Design Quarterly*, 5(3), 10–23. <https://doi.org/10.1145/3188173.3188175>
- Sarker, I. H. (2024). LLM potentiality and awareness: a position paper from the perspective of trustworthy and responsible AI modeling. *Discover Artificial Intelligence*, 4(1), 40. <https://doi.org/10.1007/s44163-024-00129-0>
- Sauer, J., Sonderegger, A. & Schmutz, S. (2020). Usability, user experience and accessibility: towards an integrative model. *Ergonomics*, 63(10), 1207–1220. <https://doi.org/10.1080/00140139.2020.1774080>
- Schweizerdeutschübersetzen.ch. (2025). *Schweizerdeutsch Übersetzer App - von Dialekt auf Deutsch*. Verfügbar unter: <https://schweizerdeutsch-uebersetzen.ch/>
- Scott, I. A. & Zuccon, G. (2024). The new paradigm in machine learning – foundation models, large language models and beyond: a primer for physicians. *Internal Medicine Journal*, 54(5), 705–715. <https://doi.org/10.1111/imj.16393>

- Soares, T. S., Costa, R. L. H., Soares, E., Calderon, I., Lunardi, G. M., Valle, P. H. D. et al. (2025). Machine Learning-Assisted Tools for User eXperience Evaluation: A Systematic Mapping Study. In *Simpósio Brasileiro de Sistemas de Informação (SBSI)* (S. 379–388). Recife: SBC. <https://doi.org/10.5753/sbsi.2025.246512>
- Song, B., Zhu, Q. & Luo, J. (2024). Human-AI collaboration by design. *Proceedings of the Design Society*, 4, 2247–2256. <https://doi.org/10.1017/pds.2024.227>
- Stockman, C. (2025). Generative AI and the Ethical Risks Associated with Human-Computer Symbiosis. *Weizenbaum Journal of the Digital Society*, 5(1). <https://doi.org/10.34669/wi.wjds/5.1.2>
- Unterkalmsteiner, M., Abrahamsson, P., XiaoFeng, W., Nguyen-Duc, A., Shah, S., Shahid Bajwa, S. et al. (2016). Software Startups - A Research Agenda. *e-Informatica Software Engineering Journal*, 10(1), 89–123. <https://doi.org/10.5277/E-INF160105>
- UXQB e.V. (Hrsg.). (2025). *CPUX-UT Curriculum: Certified Professional for Usability and User Experience – Usability Testing and Evaluation (Version 1.21 DE)*. Verfügbar unter: https://uxqb.org/public/documents/CPUX-UT_DE_Curriculum.pdf
- Vermeeren, A. P. O. S., Law, E. L.-C., Roto, V., Obrist, M., Hoonhout, J. & Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: current state and development needs (NordiCHI '10). In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries* (S. 521–530). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/1868914.1868973>
- Wallraff, B., Brossmann, F., Hochlenert, C., Bihn, K., Pohl, M. & Nickstadt, N. (2023). *Künstliche Intelligenz in kleinen und mittleren Unternehmen - Hürden in der Implementierung und wie man sie überwinden kann*. Verfügbar unter: https://www.researchgate.net/publication/366965284_Kunstliche_Intelligenz_in_kleinen_und_mittleren_Unternehmen_-_Hurden_in_der_Implementierung_und_wie_man_sie_uberwinden_kann
- Wassermann, S. (2015). Das qualitative Experteninterview. In M. Niederberger & S. Wassermann (Hrsg.), *Methoden der Experten- und Stakeholdereinbindung in der sozialwissenschaftlichen Forschung* (S. 51–67). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-01687-6_4
- Weichert, S., Quint, G. & Bartel, T. (2021). Worum geht es? In S. Weichert, G. Quint & T. Bartel (Hrsg.), *Quick Guide UX Management: So verankern Sie Usability und User Experience im Unternehmen* (S. 1–32). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-34726-0_1
- Zaghdoudi, S. & Glomann, L. (2021). Artificial Intelligence Enabled User Experience Research. In T. Ahram (Hrsg.), *Advances in Artificial Intelligence, Software and Systems Engineering* (S. 187–193). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-51328-3_27
- Zhang, X., Zeng, Y., Li, Q., Chen, G., Xu, Q., Hu, X. et al. (2024). DesignWatch: Analyzing Users' Operations of Mobile Apps Based on Screen Recordings. In *26th International Conference*

on Mobile Human-Computer Interaction (S. 1–7). Melbourne, VIC: ACM.

<https://doi.org/10.1145/3640471.3680231>

Zheng, Q., Tang, Y., Liu, Y., Liu, W. & Huang, Y. (2022). UX Research on Conversational Human-AI Interaction: A Literature Review of the ACM Digital Library. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (S. 1–24). New York, NY:

Association for Computing Machinery. <https://doi.org/10.1145/3491102.3501855>

Zhuhadar, L. & Lytras, M. (2023). The Application of AutoML Techniques in Diabetes Diagnosis: Current Approaches, Performance, and Future Directions. *Sustainability*, 15, 13484.

<https://doi.org/10.3390/su151813484>

Abbildungsverzeichnis

Abbildung 1. Übersicht über UX und Usability als Teilbereich (in Anlehnung an Jacobsen & Meyer, 2019, S.33, eigene Darstellung)	6
Abbildung 2. Die Komponenten von Usability: Nutzer*innen, Werkzeuge, Aufgaben und Kontext (in Anlehnung an Richter & Flückiger, 2016, S.11, eigene Darstellung).....	7
Abbildung 3. Human-Centered Design Prozess nach ISO 9241-210 (in Anlehnung an Weichert et al., 2021, S.25, eigene Darstellung)	10
Abbildung 4. Künstliche Intelligenz und verwandte Teilbereiche (in Anlehnung an Zhuhadar & Lytras, 2023, S.4, eigene Darstellung)	15
Abbildung 5. Forschungsdesign: Mensch und KI im UXR (eigene Darstellung).....	27
Abbildung 6. Grafische Darstellung des NAP-Ablaufs. Aus «Wirtschaftlichkeit von Usability-Evaluationsmethoden» von L. Honegger, 2023, Hochschule für Angewandte Psychologie FHNW, S. 42. Copyright 2023 bei Honegger. Wiedergabe mit Genehmigung.	35
Abbildung 7. Konzept KI-gestützte Analyse und Mensch-KI-Vergleich (eigene Darstellung)	37
Abbildung 8. Beispiel von manuell ergänzten Kontextinformationen in einem KI-Transkript (eigene Darstellung).....	38
Abbildung 9. Konfusionsmatrix (in Anlehnung an Humm et al., 2022, S.21, eigene Darstellung)	39
Abbildung 10. Exemplarischer Ausschnitt eines KI-Transkripts mit fehlerhafter Transkription (eigene Darstellung).....	51

Tabellenverzeichnis

Tabelle 1 Kriterien zur Beurteilung von Benutzungsschnittstellen nach ISO 9241 für eine nutzerfreundliche Anwendung (Jacobsen & Meyer, 2019).....	13
Tabelle 2 Überblick über den Zeitaufwand der einzelnen Arbeitsschritte des Mensch-KI-Vergleichs (Zeitangaben gerundet)	52
Tabelle 3 Absolute Häufigkeiten der Übereinstimmung zwischen menschlicher und KI-Analyse in RP, FN und FP pro Analyseart (ROH, BER, KON)	53
Tabelle 4 Kennzahlen Recall und Precision pro Analyseart (gerundete Werte).....	54
Tabelle 5 Qualitative Einordnung zusätzlicher KI-identifizierter Usability-Probleme.....	54

Glossar

Begriff	Definition
Blackbox-Problem	Eingeschränkte Nachvollziehbarkeit interner Prozesse von KI-Systemen, bei denen für Nutzer*innen nicht klar und transparent ist, wie Eingaben zu bestimmten Ausgaben führen.
Deep Learning (DL)	Teilbereich von Machine Learning; basiert auf künstlichen neuronalen Netzen, diese Algorithmen können aus unstrukturierten Daten lernen.
Falsch-Negative (FN)	<i>Hier:</i> Usability-Problem, das in der menschlichen Referenzanalyse enthalten ist, von der KI jedoch nicht identifiziert wurde.
Falsch-Positive (FP)	<i>Hier:</i> Usability-Problem, das von der KI identifiziert wurde, jedoch nicht in der menschlichen Referenzanalyse enthalten ist.
Generative KI	Teilbereich von Deep Learning; KI-Systeme, die mithilfe von Wahrscheinlichkeitsmodellen eigenständig neue Inhalte wie Texte oder Bilder generieren können.
Halluzinationen	Generierung scheinbar plausibler, jedoch nicht daten- oder faktenbasierter KI-Inhalte.
Künstliche Intelligenz (KI)	Überbegriff für Verfahren und Systeme, die menschliches Verhalten nachahmen und auf Basis von Daten lernen, Muster erkennen sowie Entscheidungen treffen können.
Large Language Models (LLMs)	Grosse Sprachmodelle (KI-Modelle) zur Verarbeitung und Generierung natürlicher Sprache, bspw. ChatGPT.
Natural Language Processing (NLP)	Teilbereich von KI, der sich mit der automatisierten Verarbeitung, Analyse und Generierung natürlicher Sprache durch Computer befasst
Machine Learning (ML)	Teilgebiet von KI, aus lernfähigen Algorithmen, die aus Daten Muster und Zusammenhänge ableiten, ohne explizit dafür programmiert zu sein.
Persona	Prototypische Anwenderprofile, die typische Merkmale, Bedürfnisse und Nutzungsmuster einer Zielgruppe beinhalten und im UX-Bereich zur Evaluation von Design- und Funktionsentscheidungen genutzt werden.
Pleasing / Sycophancy	Übermässig zustimmendes Verhalten oder schmeichelnde Antworten von KI-Systemen, die bestrebt sind möglichst hilfreich, unterstützend und respektvoll zu sein.

Precision (Übereinstimmung)	Gibt an, wie zuverlässig bzw. treffsicher die von der KI identifizierten Usability-Probleme sind.
Prompting	Formulierung von Eingaben (Prompt) an ein KI-System zur Aufforderung und Steuerung von KI-Ausgaben, um gewünschte Ergebnisse zu erhalten.
Recall (Vollständigkeit)	Beschreibt, wie gut die KI in der Lage ist, alle relevanten Usability-Probleme (Vergleich zur Referenz) zu erkennen.
Richtig-Positive (RP)	Usability-Problem, das sowohl von der KI als auch in der Referenzanalyse identifiziert wurde.
User Journey	Alle Schritte, Interaktionen und Berührungspunkte, die Nutzende durchlaufen, um ein bestimmtes Ziel mit einem Produkt oder einer Dienstleistung zu erreichen.
(KI-)Workflow	Abfolge von (KI-gestützten) Arbeits- oder Prozessschritten.
Tool	<i>Hier:</i> Software oder digitales Werkzeug, das zur Unterstützung spezifischer Aufgaben eingesetzt wird.

Hilfsmittelverzeichnis mit Verwendungszweck

KI-Assistenzsystem	Teile / Stelle(n) in der Arbeit	Einsatz
ChatGPT	Mensch-KI-Vergleich bzw. KI-gestützte Analyse	Auswertung von KI-Transkripten zur Evaluation der Leistungsfähigkeit von KI im Vergleich zu einem Menschen
ChatGPT	Prompterstellung	Zur Erstellung und Optimierung des Prompts für die KI-Analyse der Usabilitydaten
ChatGPT	Zusammenfassung / Abstract	Übersetzungshilfe für englisches Abstract sowie kürzen und umformulieren
ChatGPT	Alle Teile des Berichtes	Unterstützung für alternative Formulierungen, Optimieren der Lesbarkeit, Grammatikalische Korrektheit Jeder KI-generierte Inhalt wurde geprüft und überarbeitet
ChatGPT	Leitfäden	Formulierungshilfe und Brainstorming zu möglichen Fragen
DeepL	Abstract / Alle Teile des Berichtes	Übersetzungshilfe für englisches Abstract sowie Übersetzungen von Fachbegriffen aus der englischen Literatur
Connected Paper / Litmaps	Literaturrecherche	Identifikation relevanter wissenschaftlicher Literatur auf Basis bereits gefundener Paper
Noscribe Software / Microsoft Teams	Transkription Experteninterviews	Unterstützung Transkription Experteninterviews, erstes grobes Gerüst
MaxQDA	Analyse Experteninterviews	Qualitatives Analysetool für Inhaltsanalyse nach Kuckartz und Rädiker (2020)
schweizerdeutsch-übersetzen.ch	Datensatz Praxispartner	Transkription Usability-Tests