Theoretische Einbettung, Forschungsfrage, Hypothesen und Forschungsdesign

Theoretische Einbettung

Modelltheoretisch knüpft die vorliegende Arbeit an frühere Studien in den Bereichen Vertrauen in künstliche Intelligenz/technologische Veränderungen, wahrgenommene Nützlichkeit sowie Benutzerfreundlichkeit und die daraus abgeleitete Nutzungsabsicht an. Als theoretische Grundlage dient zunächst das Technology Acceptance Model (TAM) von Fred Davis aus dem Jahr 1987, welches die Rahmenbedingungen zur Analyse von Adoptionsprozessen neuer Technologien schafft (Davis, 1987). Den zweiten Baustein liefert die Erweiterung des TAM-Modells durch Baroni et al. (2022). Diese ergänzt das Modell um zusätzliche Faktoren wie das Vertrauen in KI-gestützte Assistenten und bildet diese im Artificial Intelligence Technology Acceptance Model (AI-TAM) ab (Baroni et al., 2022).

Zuletzt wird der Framing-Effekt theoretisch beleuchtet, da dieser für die gewählte Stimulus-Wahl relevant ist. Konkret wird dabei die Form des Attribute-Framing-Effekts betrachtet (Druckman, 2001; Freling et al., 2014).

Technology Acceptance Model (Davis, 1987)

Das TAM wurde entwickelt, um die mangelnde Nutzerakzeptanz von Informationssystemen zu adressieren, die als Haupthindernis für den Erfolg neuer Technologien identifiziert wurde. Davis untersuchte 112 Angestellte und Manager eines grossen nordamerikanischen Unternehmens, die zwei unterschiedliche Softwaresysteme nutzten - ein elektronisches Mailsystem und einen Texteditor. Das Modell basiert auf der Attitude-Paradigm aus der Psychologie, speziell auf Fishbein und Ajzens Theory of Reasoned Action (Ajzen & Fishbein, 1975). TAM besagt, dass die tatsächliche Systemnutzung durch die Verhaltensintention bestimmt wird, welche von der Einstellung zur Nutzung abhängt.

Diese Einstellung wird durch zwei zentrale Konstrukte geprägt:

- Perceived Usefulness – «the degree to which an individual believes that using a particular system would enhance his or her job performance» – und

- Perceived Ease of Use – «the degree to which an individual believes that using a particular system would be free of physical and mental effort».

Die Studie zeigte, dass Perceived Usefulness etwa 50% einflussreicher auf die Nutzung war als Ease of Use, wobei das Modell 36% der Varianz in der tatsächlichen Nutzung erklären konnte.

Während TAM erfolgreich die Akzeptanz traditioneller Informationssysteme erklärt, erweist es sich für KI-basierte Systeme als unzureichend. KI-Systeme unterscheiden sich durch ihre probabilistische Natur und inhärente Unsicherheit - Eigenschaften, die Vertrauen zu einem Faktor machen, der im ursprünglichen TAM nicht berücksichtigt wird. Zudem werden KI-Systeme nicht nur als Werkzeuge, sondern oft auch als kollaborative Partner wahrgenommen, was neue Dimensionen der Mensch-Maschine-Interaktion eröffnet. Diese Lücke adressieren Baroni et al. (2022) mit ihrer Erweiterung des TAM-Modells (Baroni et al., 2022).

Artificial Intelligence-Technology Acceptance Model (Baroni et al., 2022)

Notwendigkeit der TAM-Erweiterung für KI-Systeme. Das von Davis (1987) entwickelte Technology Acceptance Model (TAM, (Davis, 1987)) basiert auf der Theory of Reasoned Action und erklärt Technologieakzeptanz durch die Faktoren der wahrgenommenen Nützlichkeit (Perceived Usefulness) und der wahrgenommenen Benutzerfreundlichkeit (Perceived Ease of Use). Während das TAM die Adoption traditioneller Informationssysteme bereits untersucht hat, erweist es sich für KI-basierte Systeme als unzureichend. Der Grund dafür liegt in den Unterschieden von KI-Systemen: Ihre probabilistische Natur, die Unsicherheit und ihre Wahrnehmung als kollaborative Partner anstelle reiner Werkzeuge. Diese Eigenschaften machen Vertrauen zu einem Faktor, der im ursprünglichen TAM nicht abgebildet wird. Darüber hinaus erfordern «Human-in-the-Loop»-Ansätze, dass Nutzer aktiv zur Verbesserung der KI beitragen – eine Dimension der Kollaboration, die das klassische Modell ebenfalls nicht berücksichtigt.

Erweiterungen des TAM zum AI-TAM. Baroni et al. (2022) erweiterten das TAM um drei zusätzliche Konstrukte: «Explainable AI Trust» (Vertrauen in KI) aus der

Literatur zu «Explainable AI» (XAI), «Collaborative Intention» (Kollaborationsabsicht) zur Messung der Bereitschaft zur Teilnahme an «Human-in-the-Loop»-Mechanismen sowie die Vertrautheit mit der Technologie und dem Anwendungskontext. Vertrauenskonstrukt basiert auf der Arbeit von Hoffman et al. (Hoffman et al., 2019). Dabei misst «Explainable AI Trust» die Zuversicht in die Ergebnisse der KI. Die «Collaborative Intention» erfasst die Bereitschaft der Nutzer, aktiv zur Verbesserung der KI beizutragen, was ein kritischer Faktor für «Human-in-the-Loop»-Systeme ist. Dieses übt einen Einfluss auf Kernkonstrukte des TAM aus.

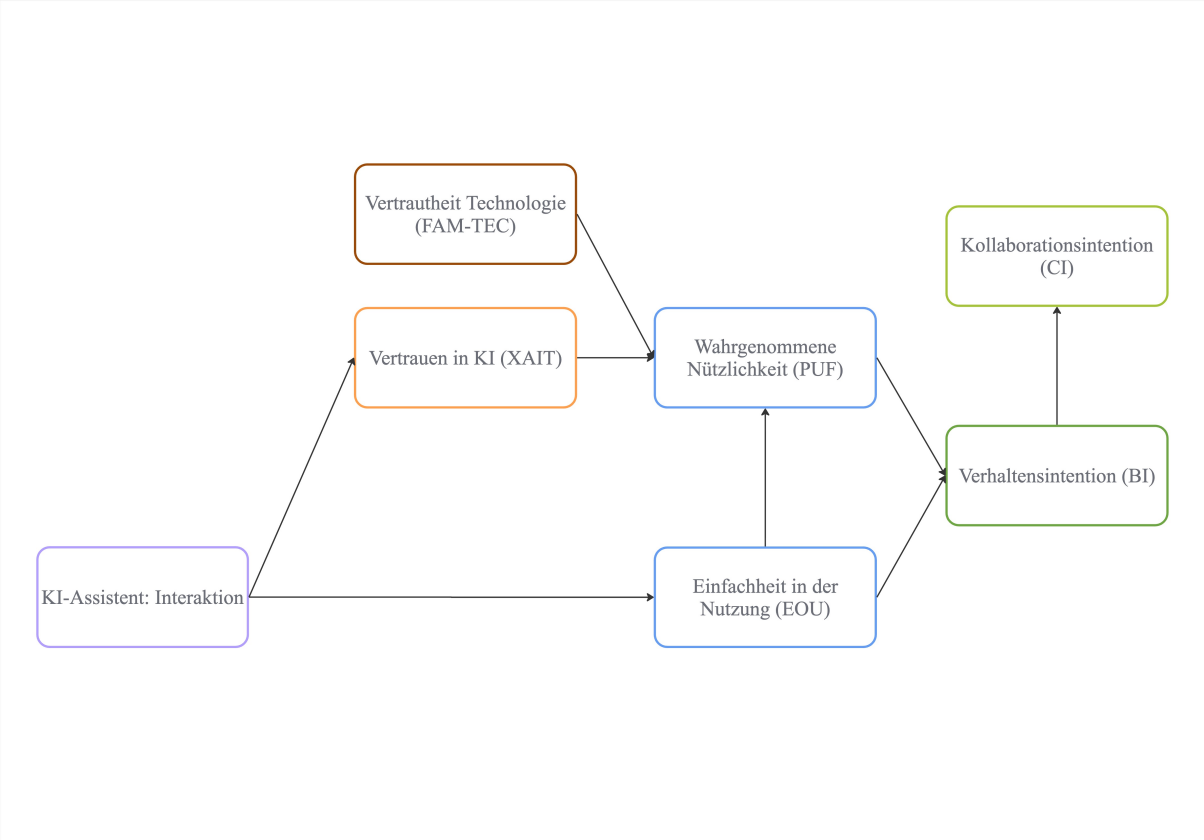

Abbildung 1

Erweitertes TAM-Modell: Artificial Intelligence-Technology Acceptance Model (Baroni et al., 2022)

Validierung durch die BumpOut-Studie (Baroni et al., 2022). Das AI-TAM wurde mit der Anwendung «BumpOut» validiert, einer KI-gestützten App zur Schadensmeldung bei Autounfällen. Die Studie umfasste 400 Teilnehmende in zwei Crowdsourcing-Kampagnen unter unterschiedlichen experimentellen Bedingungen: einer

fehlerfreien KI versus einer teilweise fehlerhaften KI. Die App analysiert dabei Schadensbilder automatisch, wobei die Nutzer die von der KI getroffenen Identifikationen bestätigen oder korrigieren können. Die Ergebnisse zeigten hohe Werte für die wahrgenommene Nützlichkeit und Benutzerfreundlichkeit, während die Funktionsfähigkeit der KI nur einen minimalen Einfluss hatte. Besonders bedeutsam war die starke Korrelation zwischen der Nutzungsabsicht («Behavioral Intention») und der Kollaborationsabsicht («Collaborative Intention»). Dies bestätigt, dass Nutzer, die bereit sind, die App zu verwenden, auch bereit sind, zur Verbesserung der KI beizutragen. Das AI-TAM eignet sich daher auch für die Untersuchung der Akzeptanz von Large Language Models (LLMs), da diese Systeme die gleichen kritischen Charakteristika aufweisen: probabilistische Ausgaben, inhärente Unsicherheit und die Notwendigkeit von Nutzervertrauen. LLMs werden zunehmend als kollaborative Partner wahrgenommen, deren Ergebnisse oft Nutzerfeedback erfordern. Insbesondere die XAI-Konstrukte sind hier relevant, da Nutzer nachvollziehen müssen, warum ein LLM eine bestimmte Antwort generiert.

Die Verbindung des AI-TAM mit dem Konzept des Attribute Framings eröffnet neue Forschungsperspektiven. Framing-Effekte könnten insbesondere das XAIT-Konstrukt beeinflussen. So dürfte die Präsentation von KI-Fähigkeiten als Gewinn («95 % Genauigkeit») im Vergleich zu einer Darstellung als Unsicherheit («5 % Fehlerrate») das Vertrauen in die KI («Explainable AI Trust») direkt verändern. Gemäss dem AI-TAM-Modell beeinflusst dieser Faktor wiederum die Nutzungsabsicht («Behavioral Intention»). Für Experimente mit Large Language Models (LLMs) bedeutet dies, dass die Art der Leistungsdarstellung, beispielsweise durch das Attribute Framing der Modellfähigkeiten, die Nutzerakzeptanz beeinflussen könnte. Das AI-TAM bietet hierbei den methodischen Rahmen, um diese Effekte auf den relevanten Dimensionen präzise zu messen.

Framing-Effekt

Der Framing-Effekt, erstmals von Kahneman und Tversky in ihrer Prospect Theory beschrieben, zeigt, dass Entscheidungen davon beeinflusst werden, wie Informationen

präsentiert werden (Tversky & Kahneman, 1986). Der Framing-Effekt zeigt unter anderem, wie identische Szenarien zu unterschiedlichen Präferenzen führen, je nachdem ob sie in Gewinn- oder Verlustbegriffen formuliert werden. Während sich die frühe Forschung auf riskante Entscheidungen konzentrierte, erweiterte sich das Konzept auf verschiedene Framing-Typen wie Risky Choice Framing, Goal Framing und Attribute Framing.

Zusätzlich untersuchte Freling et al. (2014) in ihrer Meta-Analyse 107 Studien zum Attribute Framing und entwickelten dabei eine theoretische Integration mittels Construal Level Theory (CLT). Ihre zentrale Erkenntnis: Die Effektivität von Attribute Framing hängt von der Kongruenz zwischen dem Abstraktionsniveau (Construal Level) des Frames und der psychologischen Distanz des Bewertenden zum geframten Event ab (Freling et al., 2014).

Attribute Framing nach Levin & Gaeth (1988) und Dolgopolova et al. (2022). Attribute Framing unterscheidet sich von anderen Framing-Typen, da hier ein einzelnes Attribut in äquivalenten aber unterschiedlich valenten Begriffen beschrieben wird. Levin und Gaeth demonstrierten dies mit Hackfleisch, das entweder als «75% mager» oder «25% fett» beschrieben wurde (Levin & Gaeth, 1988). Der Attribute Framing-Effekt manifestiert sich in einer valenz-konsistenten Verschiebung: Positive Frames führen zu günstigeren Bewertungen als negative. Ihre Studie zeigte zudem, dass direkte Produkterfahrung den Framing-Effekt abschwächt - ein Befund, der durch ein Averaging-Modell erklärt wird, bei dem zusätzliche Informationsquellen den relativen Einfluss einzelner Frames reduzieren.

Dolgopolova et al. (2022) spezifisch auf Lebensmittelentscheidungen und fanden Effekte für Einstellungen versus Intentionen. Während Gain-Frames signifikant positivere Einstellungen erzeugten, war der Effekt auf Kaufintentionen nahe null und nicht signifikant. Mehrere Moderatoren wurden identifiziert: Gain-Frames, Interaktionsterme, spezifische Produkte und Studentenstichproben beeinflussten signifikant die Ergebnisse. Diese Befunde unterstreichen die Komplexität des Attribute Framing bei Lebensmitteln, wo zeitliche Diskontierung und die Verzögerung zwischen Konsum und

Gesundheitskonsequenzen eine Rolle spielen (Dolgopolova et al., 2022).
Der Attribute Framing-Effekt ist für die Untersuchung der KI-Akzeptanz relevant, da KI-Systeme durch ihre Fähigkeiten (Gain-Frame: «95% Genauigkeit») oder Limitationen (Loss-Frame: «5% Fehlerrate») charakterisiert werden können. Im Kontext des erweiterten TAM-Modells (Davis, 1987) könnte Attribute Framing die Wahrnehmung von Perceived Usefulness und Vertrauen in KI beeinflussen. Die Präsentation von KI-Funktionen als Gewinne («erhöht Produktivität um 30%») versus Verluste («30% manuelle Arbeit bleibt erforderlich») könnte unterschiedliche Akzeptanzmuster erzeugen.

Forschungsfrage

Wie beeinflusst die Framingdarstellung von KI-Konfidenzwerten (positiv vs. negativ) das Vertrauen in KI-generierte Antworten und die daraus resultierende Technologieakzeptanz in LLM-basierten Assistenzsystemen?

Latente Konstrukte

Die latenten Konstrukte werden mittels einer Online-Befragung vor, während und nach der Nutzung der KI-Assistenz erhoben. Die verwendeten Konstrukte stammen grossteils aus dem Technology Acceptance Model (TAM) von Davis sowie aus der Erweiterung dieses Modells durch Baroni et al. (2022). Diese Erweiterung ergänzt das bestehende TAM um KI-relevante Faktoren wie das Vertrauen in erklärbare KI («Explainable AI Trust», XAIT), die Kollaborationsabsicht («Collaborative Intention, CI») und die technologische Vorerfahrung («Familiarity with Technology», FAM-TEC») (Baroni et al., 2022).

Hypothesenübersicht

Ausformulierte Hypothesen

Haupthypothesen (Framing-Effekte).

- H1a: Die positive Darstellung des Accuracy Scores (z.B. «Diese Antwort ist zu 80% korrekt») führt zu einem höheren AI Output Trust als die Kontrollbedingung ohne Score-Anzeige.

- H1b: Die negative Darstellung des Accuracy Scores (z.B. «Diese Antwort hat eine 20% Fehlerwahrscheinlichkeit») führt zu einem niedrigeren AI Output Trust als die Kontrollbedingung ohne Score-Anzeige.

AI-TAM Kernbeziehungen.

- H2: AI Output Trust hat einen positiven Einfluss auf die Perceived Usefulness. Nutzer, die den AI-Ausgaben vertrauen, bewerten das System als nützlicher für ihre Aufgaben.
- H3: AI Output Trust hat einen positiven Einfluss auf die Perceived Ease of Use. Vertrauen in die AI-Ausgaben reduziert die wahrgenommene kognitive Belastung bei der Systemnutzung.

TAM-Standardbeziehungen.

- H4: Die Perceived Usefulness hat einen positiven Einfluss auf die Behavioral Intention. Je nützlicher Nutzer Alva einschätzen, desto höher ist ihre Absicht, das System zukünftig zu nutzen.
- H5: Die Perceived Ease of Use hat einen positiven Einfluss auf die Behavioral Intention. Eine als einfach wahrgenommene Nutzung erhöht die Intention zur zukünftigen Systemnutzung.
- H6: Die Perceived Ease of Use hat einen positiven Einfluss auf die Perceived Usefulness. Systeme, die einfach zu nutzen sind, werden als nützlicher wahrgenommen.
- H7: Die Behavioral Intention hat einen positiven Einfluss auf die Collaborative Intention. Nutzer, die beabsichtigen Alva zu nutzen, zeigen auch eine höhere Bereitschaft zur kollaborativen Zusammenarbeit mit dem AI-System.
- H8: Familiarity with Technology hat einen positiven Einfluss auf die Perceived Usefulness. Nutzer, die mit KI-Technologie vertraut sind, schätzen die Nützlichkeit der digitalen Assistenz höher ein.

Mediation.

- H9: Der Effekt des Framings auf die Behavorial Intention wird durch Explainable AI Trust partiell oder vollständig mediiert. Bei niedrigen Explainable AI Trust-Werten ist der Unterschied zwischen positivem und negativem Framing grösser als bei hohen Explainable AI Trust-Werten.

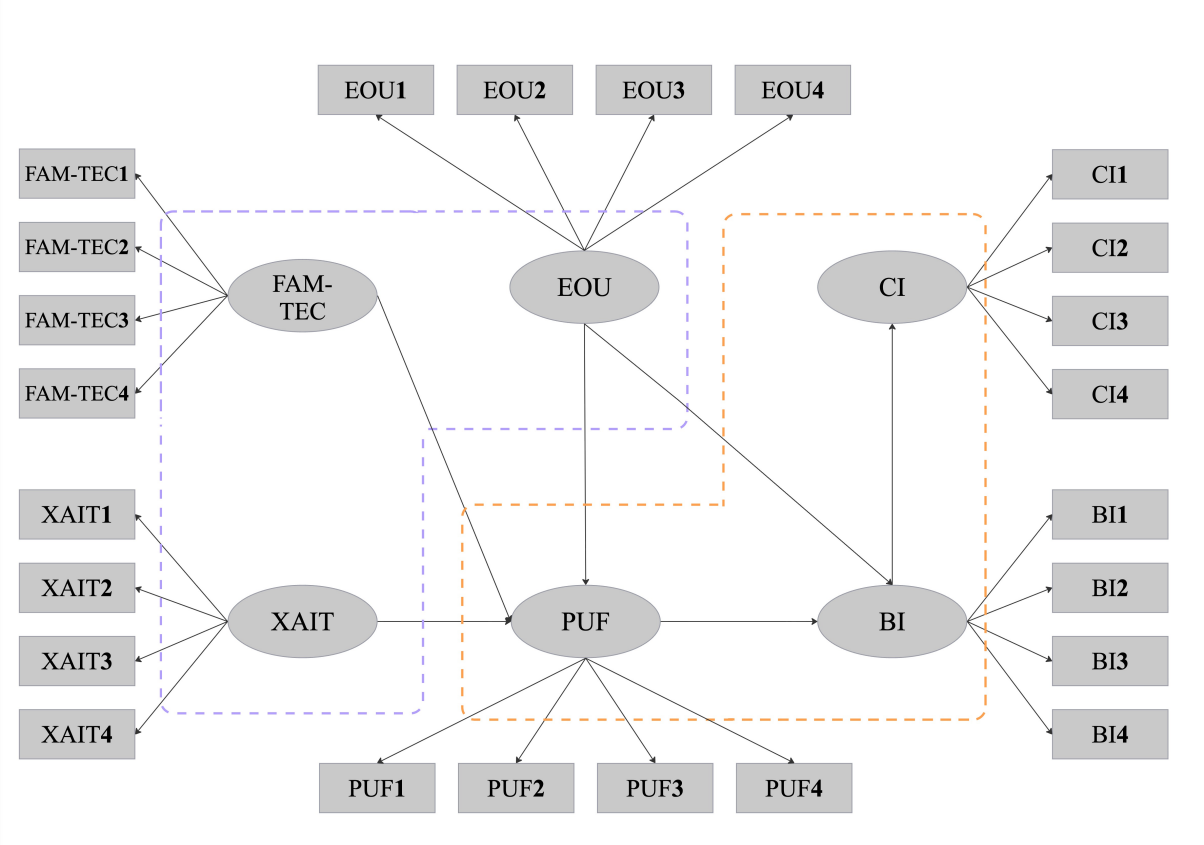

Abbildung 2

Strukturmodell für AI-TAM, Teil des Strukturgleichungsmodell, inkl. Hypothesen, inkl. Stimulus

Tabelle 1

Identifizierte latente Konstrukte zum Einsatz von AI-TAM

Abkürzung	Name	Definition	Quelle
XAIT	Explainable AI trust	Vertrauen in die generierte Antwort und die LLM-Lösung	(Baroni et al., 2022)
BI	Behavorial Intention	Verhaltensintention die LLM-Lösung zu Nutzen	(Baroni et al., 2022; Davis, 1987)
CI	Collaborative Intention	Kollaborationsintention zur digitalen Assistenz	(Baroni et al., 2022; Davis, 1987)
PUF	Perceived Usefulness	Wahrgenommene	(Baroni et al., 2022; Davis, 1987)

Tabelle 2

Aufgelistete Hypothesen im Rahmen der Bachelor-Arbeit-Vorstudie

Hypothese	**Pfad**	**Richtung**	**Theorie**
H1a	Dummy_Pos (Stimulus) → XAIT	+	Attribute Frame
H1b	Dummy_Neg (Stimulus) → XAIT	-	Attribute Frame
H2	XAIT → PUF	+	AI-TAM-Modell
H3	XAIT → EOU	+	AI-TAM-Modell
H4	PUF → BI	+	AI-TAM-Modell
H5	EOU → BI	+	AI-TAM-Modell
H6	EOU → PUF	+	AI-TAM-Modell
H7	BI → CI	+	AI-TAM-Modell
H8	FAMTEC → PUF	+	AI-TAM-Modell
Kovariate			
ACTS	ACTS → XAIT	kontrolliert	-
Mediation			
H9	Framing → XAIT → (PUF, EOU) → BI	Indirekt	-

Anmerkung: Alle Pfade werden simultan im Strukturgleichungsmodell (SEM) geschätzt