INTERNATIONALE HOCHSCHULE

www.iu.de

IU DISCUSSION PAPERS

Business & Management

Generative KI im Kundenservice der Fondsdienstleistungsbranche

Analyse der Servicequalität und Nutzungsbereitschaft

GIUSEPPE ADORNO
CONSTANTIN SCHUBART

IU Internationale Hochschule

Main Campus: Erfurt
Juri-Gagarin-Ring 152
99084 Erfurt

Telefon: +49 421.166985.23
Fax: +49 2224.9605.115
Kontakt/Contact: kerstin.janson@iu.org

Autorenkontakt/Contact to the author(s):

Giuseppe Adorno
giuseppe.adorno@iu-study.org
ORCID-ID: 0009-0004-6760-1983 (Open Researcher und Contributer ID)

Prof. Dr. Constantin Schubart
constantin.schubart@iu.org
ORCID-ID: 0009-0008-9259-0533 (Open Researcher und Contributer ID)

IU Internationale Hochschule
Juri-Gagarin-Ring 152
99084 Erfurt
Email: constantin.schubart@iu.org

IU Discussion Papers, Reihe: Business & Management, Vol. 7, No. 13 (JUL 2026)

ISSN: 2750-0683

DOI: https://doi.org/10.56250/4139

Webseite: https://repository.iu.org

GENERATIVE KI IM KUNDENSERVICE DER FONDSDIENSTLEISTUNGSBRANCHE

ANALYSE DER SERVICEQUALITÄT UND NUTZUNGSBEREITSCHAFT

Giuseppe Adorno
Constantin Schubart

ABSTRACT:

Evolving customer expectations and increasing pressure for efficiency are accelerating the integration of generative artificial intelligence into customer services. This paper examines the perception and acceptance of AI-supported communication within the particularly trust-sensitive asset management industry. Based on an online survey incorporating a vignette experiment, the study analyzes perceived service quality, trust, and inteded use in this context. The findings indicate that AI-generated responses are perceived as equivalent in quality or slightly better than human-generated responses in certain dimensions. At the same time a preference for human operated channels seems to persist. The sample reveals different user segments ranging from digitally inclined and independent decision-makers to more traditionally oriented clients. These groups require differentiated implementation strategies. Accordingly, the primary challenge for financial service providers lies less in the technological capabilities of generative systems than in overcoming a trust- and context-dependent acceptance gap. A sustainable deployment of AI in customer service therefore does not call for a one-size-fits-all solution. This study recommends aligning the implementation with an analysis of the company's own customer structure and their individual needs, and gradually adapting the business process landscape to the new technology.

KEYWORDS:

Generative Artificial Intelligence; AI-Supported Customer Service; Perceived Service Quality; Trust in AI-Systems; Technology Acceptance Models; Asset-Management

JEL classification: G23, O33, D83

AUTOR:INNEN

Giuseppe Adorno *hat den Masterstudiengang in Digitaler Transformation mit Vertiefung in Künstlicher Intelligenz an der IU absolviert. Seine berufliche Praxis knüpft unmittelbar daran an. Als Senior Business Expert in der Fondsdienstleistungsbranche treibt er dort den Einsatz generativer und agentischer KI innerhalb der Prozesslandschaft des Fondskundenservices voran.*

Prof. Dr. Constantin Schubart *ist seit 2020 Professor für Allgemeine Betriebswirtschaftslehre an der IU Internationale Hochschule im Dualen Studium am Standort Erfurt. Seine Schwerpunkte liegen im Bereich Management- und Organisations-Entwicklung. Er verfügt über mehr als 20 Jahre Berufserfahrung in Lehre, Forschung und Praxis.*

Einleitung

Die fortschreitende Digitalisierung verändert die Art und Weise, wie Unternehmen mit ihren Kunden interagieren. Der Kundenservice ist dabei ein zentraler Berührungspunkt, der sich auf die Zufriedenheit und Bindung auswirkt. Vor diesem Hintergrund stehen traditionelle Finanzdienstleister, insbesondere die hier betrachteten Fondsgesellschaften, vor besonders komplexen Herausforderungen. Steigende Anfragevolumina, knappe Margen und veränderte Erwartungen der Anleger erfordern eine deutliche Effizienzsteigerung in der Abwicklung (Glock, 2024, S. 34; Peters, 2025, S. 24). Gleichzeitig verlangen regulatorische Auflagen, hohe kundenseitige Vertrauensanforderungen (Malaquias & Hwang, 2016, S. 1609) und die Komplexität von Anlagethemen einen besonders verantwortungsvollen Umgang mit automatisierten Kundeninteraktionen. Wenige Technologien haben die Diskussion über die Zukunft automatisierter Serviceprozesse so umfassend geprägt, wie die jüngste Generation großer Sprachmodelle. Ihr Einsatz gilt mittlerweile als einer der wichtigsten strategischen Hebel zur Bewältigung dieser Herausforderungen. Die so generierten Inhalte werden vor allem in Form von KI-Chatbots an den Endkunden ausgespielt (Baumann, 2025, S. 57–58; Senger, 2025, S. 39). Gleichzeitig ergeben sich aus dem Technologieeinsatz deutliche Spannungsfelder. Effizienzgewinne stehen einer potenziell distanzierteren Interaktion gegenüber, höhere Konsistenz dem Risiko, dass empathische Aspekte der Servicekommunikation verloren gehen (Ferraro et al., 2024, S. 552).

Die technologischen Grundlagen dieser Modelle und ihre Potenziale für Unternehmen sind bereits umfassend betrachtet worden. Weniger systematisch untersucht erscheint dagegen die Anlegerperspektive selbst. Wie bewerten Endkunden KI-generierte Serviceantworten im direkten Vergleich zu menschlichen? Sind sie in der Lage beide Quellen voneinander zu unterscheiden? Unter welchen Bedingungen würden sie KI-gestützte Services tatsächlich nutzen? Gerade in der Fondsdienstleistungsbranche, in der Entscheidungen erhebliche finanzielle Konsequenzen nach sich ziehen und Vertrauen ein zentrales Element der Geschäftsbeziehung ist (Raff et al., 2025, S. 75), ist diese Frage von strategischer Relevanz.

Dieses Paper greift diese Diskussion auf. Es verdichtet die Erkenntnisse einer empirischen Untersuchung mit 273 Teilnehmenden. In einem quasi-experimentellen Vignettendesign wurden sowohl KI-generierte als auch von Menschen erstellte Serviceantworten, zu typischen Anliegen der Fondsdienstleistung, vorgeführt und durch die Befragten bewertet. Im Mittelpunkt steht die Beobachtung, dass die gemessene wahrgenommene Qualität KI-generierter Antworten und die berichtete Bereitschaft, solchen Systemen tatsächlich datensensible Anliegen anzuvertrauen, deutlich auseinanderfallen.

Die Studie untersucht damit weniger, wozu Sprachmodelle technologisch in der Lage sind, als welche Konsequenzen ihre Leistungsfähigkeit für die Gestaltung kundennaher Prozesse in einer hoch regulierten und vertrauenssensitiven Branche hat. Damit rückt mit der Übersetzung des technologischen Potenzials in tatsächliche Nutzungsbereitschaft in den Vordergrund. Um dieser Fragestellung nachzugehen, ordnen die folgenden Abschnitte zunächst den theoretischen Rahmen ein und beschreiben das methodische Vorgehen. Anschließend werden die zentralen empirischen Befunde aufgezeigt und kritisch diskutiert, bevor Implikationen für Forschung und Praxis abgeleitet und in einem abschließenden Fazit zusammengeführt werden.

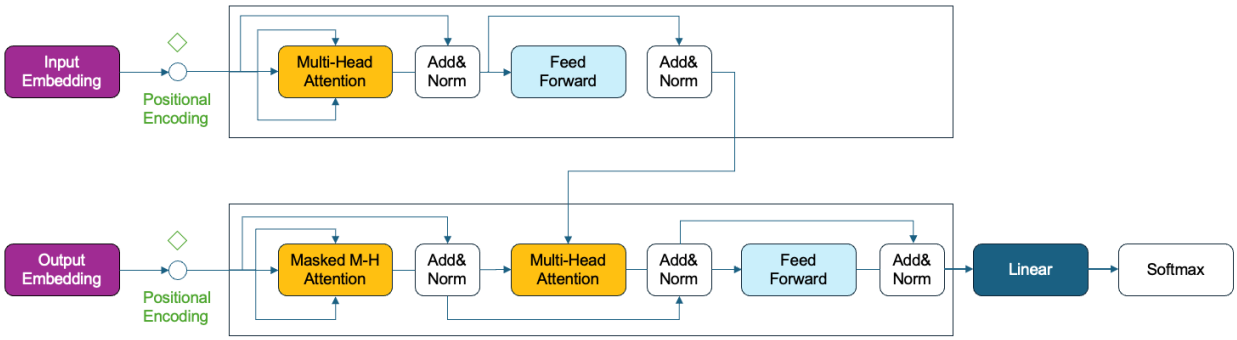
Theoretische Grundlagen

GENERATIVE KI UND LLMS

Large Language Models (LLMs) sind statistische Modelle, die sprachliche Muster aus umfangreichen Trainingsdaten erlernen und auf dieser Grundlage neue Inhalte generieren können. Anders als regelbasierte Chatbot-Systeme früherer Generationen sind sie nicht auf vordefinierte Antwortpfade angewiesen, sondern können kontextsensitiv, formal anspruchsvolle Antworten in natürlicher Sprache produzieren. Die Qualität der generierten Inhalte hat sich in den vergangenen Jahren so stark verbessert, dass deren Ergebnisse menschlicher Kommunikation immer näherkommen (Hafner & Hundertmark, 2024, S. 5).

Ermöglicht wird dies durch eine architekturelle Weiterentwicklung zur sogenannten Transformer-Architektur. Deren Self-Attention-Mechanismus erkennt wichtige Zusammenhänge zwischen Wörtern eines eingegebenen Textes und gewichtet diese unterschiedlich, abhängig vom jeweiligen Kontext. Ergänzend sorgt das sogenannte Positional Encoding dafür, dass die Reihenfolge der Wörter berücksichtigt wird, damit sprachliche Strukturen korrekt verstanden werden können (Vaswani et al., 2017, S. 2–5). Dadurch wird eine kohärente und kontextsensitive Verarbeitung natürlicher Sprache in bisher nicht gekannter Qualität zum neuen Standard.

Abbildung 1: Encoder-Decoder-Transformer-Architektur



Quelle: Eigene Darstellung (geändert) in Anlehnung an Vaswani et al.

Im Rahmen dieser Studie wurde das Modell GPT-4o von OpenAI eingesetzt. Es basiert auf einer Decoder-only-Transformer-Architektur. Das Modell wurde zunächst im Pre-Training mit großen Datenmengen vortrainiert, mit Beispielantworten im Supervised Fine-Tuning weiterentwickelt (Blüthgen, 2025, S. 231) und schließlich durch menschliches Feedback im Reinforcement Learning from Human Feedback optimiert (Seemann, 2022, S. 48). Die Bezeichnung GPT ist ein Akronym für Generative Pre-trained Transformer und leitet sich damit also direkt aus dieser Architektur- und Trainingslogik ab.

Die auf dieser Basis erzielte Leistungssteigerung wird allerdings von für die Technologie charakteristischen Einschränkungen begleitet. Dazu zählen plausibel formulierte aber faktisch falsche Inhalte, sogenannte Halluzinationen (Scheuer, 2020, S. 23–24), aber auch die mangelnde Nachvollziehbarkeit von Entscheidungen (Voge, 2020, S. 7). Erweiternde Verfahren wie Retrieval-Augmented Generation (RAG),

die das Modell auf verifizierte interne Wissensquellen stützen (Kamath et al., 2024, S. 107), oder Explainable AI (XAI), die anstreben die Entscheidungslogik weitestgehend transparent zu machen (Höhne, 2021, S. 453–455) adressieren diese Schwächen, können sie aber nicht vollständig auflösen.

SERVICEQUALITÄT UND TECHNOLOGIEAKZEPTANZ

Zur strukturierten Einordnung und Bewertung der Leistungsfähigkeit solcher Modelle bieten sich etablierte Ansätze aus der Servicequalitäts- und Technologieakzeptanzforschung an. Die Erhebung der Servicequalität wird in der Theorie vor allem durch das SERVQUAL-Modell und seinen Weiterentwicklungen geprägt. Es operationalisiert Servicequalität entlang der Dimensionen Zuverlässigkeit, Reaktionsfähigkeit, Sicherheit, Empathie sowie materielle Aspekte (A. P. Parasuraman et al., 1988, S. 14). Mit der Verlagerung der Servicekontakte in den digitalen Raum ergänzen E-S-QUAL und vergleichbare Ansätze diese um Dimensionen wie Effizienz, Erfüllung, Systemverfügbarkeit und Datenschutz (A. Parasuraman et al., 2005). Neuere KI-spezifische Adaptionen wie das AICSQ-Modell betonen darüber hinaus Informationsqualität und systemische Kompetenz als wichtige Bewertungsdimensionen (Warjiyono et al., 2025, S. 1432).

Die Technologieakzeptanzforschung wiederum bezieht sich besonders häufig auf das Technology Acceptance Model (TAM) und seiner Weiterentwicklung UTAUT2. Beide Modelle erklären die Nutzungsabsicht über die wahrgenommene Nützlichkeit und die wahrgenommene Benutzerfreundlichkeit, ergänzt um soziale Einflüsse und weitere individuelle Moderatoren (Davis, 1989; Venkatesh & Bala, 2008; Venkatesh & Davis, 2000). KI-spezifische Erweiterungen, wie das KI-Akzeptanzmodell (KIAM), unterscheiden zwischen der Wahrnehmung der KI als Werkzeug und als sozialem Akteur (Scheuer, 2020, S. 61–65) oder beziehen neue Risikodimensionen wie Halluzinationsanfälligkeit und Datenschutzbedenken mit ein, wie das Extended TAM for Generative AI (Lim et al., 2025, S. 186–187).

FONDSDIENSTLEISTUNGSBRANCHE

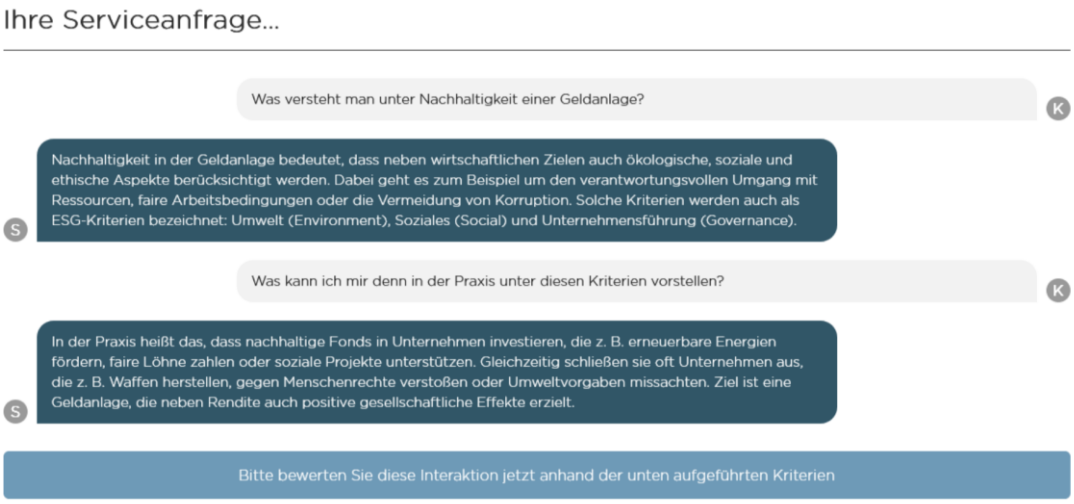
Die Fondsdienstleistungsbranche wird von besonders hoher Regulierungsdichte geprägt. Darunter fallen beispielsweise die Datenschutzgrundverordnung, das Wertpapierhandels- und Geldwäschegesetz aber auch der europäische AI-Act. Damit sind enge Grenzen für autonome KI-Entscheidungen gesetzt und umfassende Dokumentation und menschliche Aufsicht vorausgesetzt. Weiterhin betreffen die Kundenanliegen häufig persönliche Vermögenswerte, sodass fehlerhafte Auskünfte unmittelbare finanzielle Konsequenzen haben können. Schließlich basiert die Geschäftsbeziehung selbst auf Vertrauen. Fondsgesellschaften verwalten treuhänderisch Vermögen, mit dem Ziel dieses für ihre Kunden zu vermehren (Raab, 2019, S. 17). Innerhalb dieses Rahmens besteht eine ausgeprägte Informationsasymmetrie, da vor allem Privatkunden finanzwirtschaftlicher Entscheidungen häufig nur eingeschränkt beurteilen können. Vertrauen wird damit zu einem wichtigen Differenzierungsmerkmal (Hartmann-Wendels et al., 2019, S. 93, 273). Diese Rahmenbedingungen prägen die Erwartungen der Kunden daher auch an scheinbar nachgelagerten Berührungspunkten wie dem Kundenservice. KI-gestützte Serviceleistungen müssen deshalb nicht nur funktional überzeugen, sondern auch den besonderen Servicequalitäts- und Vertrauenserwartungen des regulierten Finanzkontexts genügen.

Methodik

ONLINEBEFRAGUNG UND VIGNETTENEXPERIMENT

Die empirische Grundlage des Papers bildet eine teilstandardisierte Online-Befragung mit integriertem Vignettenexperiment (Miko-Schefzig, 2022) im Within-Subjects-Design (n = 273). Im Zentrum standen sechs realitätsnahe Chatdialoge zu typischen Anliegen der Fondsdienstleistung, die als Live-Chatverläufe in einer eigens entwickelten HTML-Oberfläche präsentiert wurden.

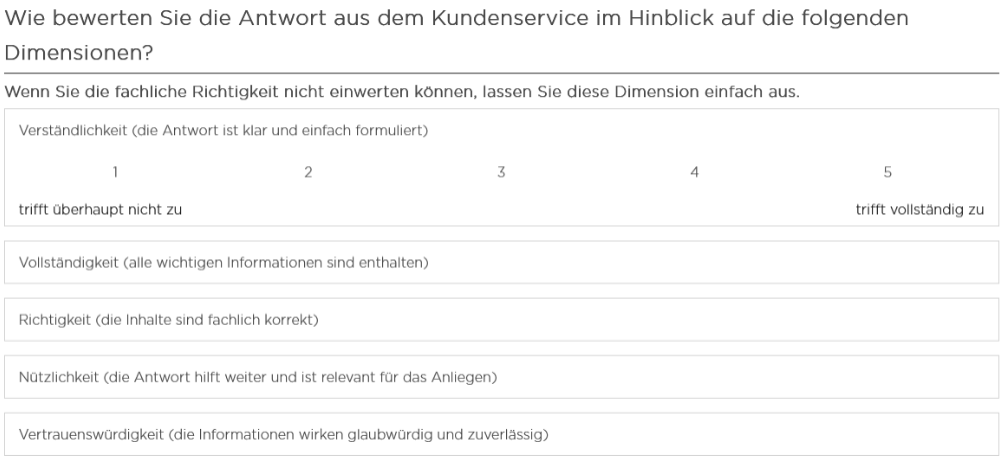
Abbildung 2: Serviceinteraktion im Chatformat



Quelle: Eigene Darstellung

Drei der gezeigten Antworten waren von einem aktuellen Large Language Model (GPT-4o) erzeugt, drei von menschlichen Autoren formuliert. Die jeweilige Quelle wurde den Teilnehmenden nicht offengelegt. Die Bewertung erfolgte für jede Interaktion in fünf Servicequalitätsdimensionen.

Abbildung 3: Bewertungsmatrix zur jeweiligen Serviceinteraktion



Quelle: Eigene Darstellung (erstellt mit der Umfragesoftware Tivian EFS)

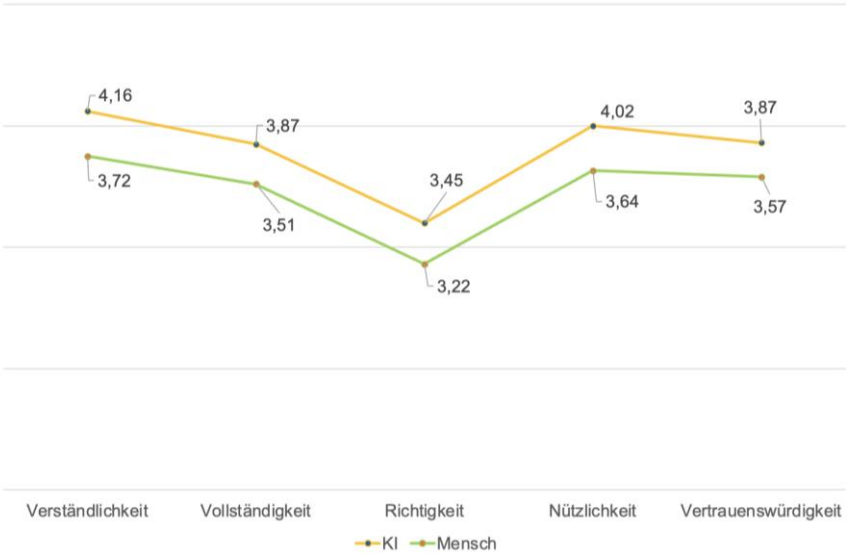
Ergänzend wurden Finanzerfahrung, Technologieaffinität, Kanalpräferenzen, Nutzungsszenarien und soziodemografische Merkmale erhoben sowie offene Antworten zu Anforderungen an KI-gestützte Services qualitativ ausgewertet. Die Stichprobe wurde mittels k-Means-Clusteranalyse zusätzlich in Kundengruppen eingeteilt. Aufgrund des explorativen, nicht zufallsbasierten Designs erlauben die Ergebnisse Aussagen über Tendenzen und Zusammenhänge innerhalb der Stichprobe, jedoch keine kausalen oder uneingeschränkt generalisierbaren Schlussfolgerungen (Boßow-Thies & Krol, 2022, S. 16). Die Auswertung kombiniert deskriptive Kennwerte mit gezielter inferenzstatistischer Absicherung. Berichtet werden durchgängig Effektstärken neben p-Werten, multiple Vergleiche werden Bonferroni-korrigiert. Die angewandten Verfahren dienen der Prüfung stichprobeninterner Muster. Eine klassische Populationsinferenz wird nicht beansprucht.

Ergebnisse der Studie

KI-GENERIERTE ANTWORTEN GEWINNEN DEN SERVICEQUALIÄTSVERGLEICH

Der aggregierte Servicequalitäts-Score (SQS), der als arithmetisches Mittel der Bewertungen über die fünf Qualitätsdimensionen gebildet wurde, fällt für KI-generierte Antworten mit M = 3,81 systematisch höher aus als für menschlich erstellte mit M = 3,48 (n = 273). Dies entspricht einem Unterschied von rund 10% zugunsten der künstlichen Intelligenz. Bemerkenswert ist die Konsistenz dieses Vorsprungs über alle fünf Einzeldimensionen hinweg. Die höchste Differenz weisen die Dimensionen Verständlichkeit (Δ M = 0,44) und Nützlichkeit (Δ M = 0,38) auf, die geringste die Richtigkeit (Δ M = 0,23). Auch die Vertrauenswürdigkeit fällt zugunsten der KI aus (M = 3,87 vs. 3,57).

Abbildung 4: SQS nach Qualitätsdimensionen



Quelle: Eigene Darstellung

Der Unterschied erscheint robust ($p < .001$) und erreicht mit $r = 0,56$ eine hohe Effektstärke. Auch in der dimensionsspezifischen Betrachtung erreichen alle fünf Qualitätsdimensionen ein konservatives Signifikanzniveau ($\alpha = 0,01$) und die Effektstärken bewegen sich zwischen $r = 0,37$ für die Richtigkeit und $r =$

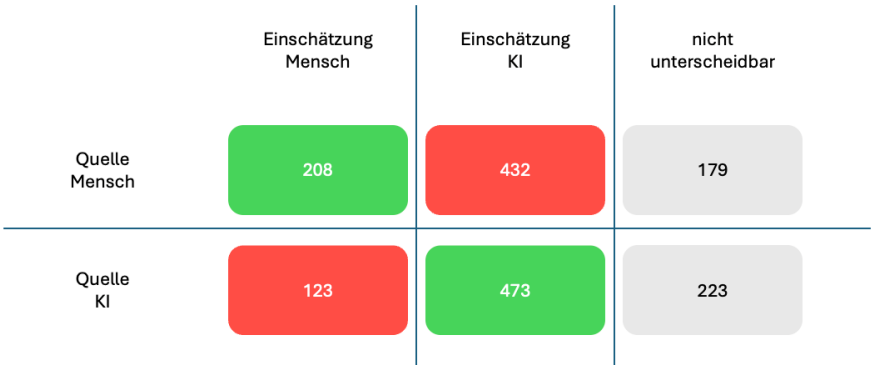
0,58 für die Verständlichkeit. Der Bewertungsvorteil zugunsten der KI-Antworten ist damit weniger ein punktuelles als ein dimensionsübergreifend Muster.

Während sich in der Stichprobe für 73 % der Teilnehmenden ein SQS von ≥ 4 für die KI-Antworten ergibt, ist dies bei den menschlichen Antworten nur bei 48 % der Fall. KI-Antworten werden nicht nur im Mittel höher bewertet, sondern auch deutlich konsistenter im oberen Bewertungsbereich verortet.

KI-VERMUTUNGS BIAS IN DER STICHPROBE

Betrachtet man ausschließlich die eindeutigen Quellenzuordnungen (exklusive „kann ich nicht unterscheiden“) wurden 55,1 % der Antworten korrekt zugeordnet. Die Erkennungsleistung liegt damit nur knapp über dem Zufallsniveau und ist asymmetrisch verteilt. KI-Antworten wurden zu 79,4 % korrekt als KI-generiert erkannt, die Antworten menschlicher Autoren dagegen nur zu 32,5 %. Wenn die Befragten sich eine eindeutige Zuordnung zutrauten, tippten sie also systematisch zu häufig auf KI. Berücksichtigt man auch die unentschiedenen Antworten als Falscherkennung und betrachtet die individuelle Correct-Guess-Quote (M = 0,41), fällt diese sogar unter das Zufallsniveau. Diese Kombination von Befunden scheint ein Ausdruck einer systematischen Verzerrung bei der Zuordnung der Autorenschaft.

Abbildung 5: Matrix zur Erkennbarkeit der Antwortquelle in drei Kategorien



Quelle: Eigene Darstellung

Entscheidend für deren Bewertung ist, ob die unterstellte Quelle die abgegebenen Qualitätsurteile geprägt hat. Eine multivariate Analyse zeigt, dass innerhalb der Stichprobe nur die tatsächliche Quelle der Antwort signifikant auf die Bewertung wirkt ($\beta = 0{,}26$, $p < .001$), während die zugeschriebene Quelle keinen eigenständigen Effekt zeigt ($\beta = -0{,}09$, $p = 0{,}12$). Der Bewertungsvorteil der KI lässt sich damit als der tatsächlichen Inhaltlichen Qualität geschuldet interpretieren. Er entsteht nicht aus einer systematischen Bevorzugung vermeintlich KI-generierter Antworten und scheint auch nicht durch weitere Verzerrung wie die Algorithm Appreciation überlagert zu werden (Logg et al., 2019, S. 93–99).

EINBRUCH DER NUTZUNGSBEREITSCHAFT MIT STEIGENDER DATENSENSITIVITÄT

Einen besonders aufschlussreichen Befund liefert die Auswertung der zukünftigen Nutzungsabsichten in unterschiedlichen Anwendungsszenarien. Während 87 % der Befragten KI-Chatbots für unpersönliche Standardanfragen wie die Erklärung von Finanzbegriffen oder die Bereitstellung allgemeiner Produktinformationen akzeptieren würden, sinkt die Bereitschaft auf 57 % bei produktbezogenen Fragen,

auf 44 % bei Statusabfragen zu Aufträgen und Transaktionen und schließlich auf nur noch 19 % bei individualisierten Anfragen zu persönlichen Depotdaten.

Die Unterschiede zwischen den vier Szenarien sind belastbar ($p < .001$), und auch in den paarweisen Vergleichen erweisen sich sämtliche Übergänge nach Korrektur als signifikant. Die Akzeptanzkurve folgt keinem schwellenartigen Muster mit einer einzelnen Bruchkante. Vielmehr geht jeder Anstieg der Datensensitivität mit einem Akzeptanzverlust einher. Am stärksten fällt der Sprung von allgemeinen zu produktbezogenen Informationen aus (–29,7 Prozentpunkte), gefolgt vom Übergang von Statusabfragen zu Depotdaten (–24,9 Prozentpunkte).

Parallel dazu äußern 56 % der Befragten explizite Datenschutzbedenken bei der Nutzung von KI im Finanzkontext, und ein Drittel der Teilnehmenden führt Sicherheit und Datenschutz in den offenen Antworten als entscheidendes Akzeptanzkriterium auf. Ein Einstichproben-Wilcoxon-Test gegen den Skalen-Neutralwert zeigt, dass die Datenschutzbedenken signifikant über dem Neutralwert liegen ($p < .001$, $r = 0,38$, $n = 204$). Die Zahlungsbereitschaft für KI-gestützte Services ist mit 24 % gering, ein Chi²-Test bestätigt, dass sie innerhalb der erreichten Stichprobe nicht vom Haushaltseinkommen abhängt ($\chi^2(5) = 1,62$, $p = 0,90$).

NUTZERSEGMENTE INNERHALB DER STICHPROBE

Eine k-Means-Clusteranalyse über Finanzerfahrung, Technologieaffinität und Kanalpräferenzen hinweg identifiziert drei Segmente innerhalb der Stichprobe dieser Studie.

Tabelle 1: Ergebnisse der Clusteranalyse

Merkmal	Cluster 1	Cluster 2	Cluster 3
n (Anteil)	32 (11,7 %)	129 (47,3 %)	112 (41,0 %)
Finanzerfahrung (Ø, 1–5)	mittel	mittel	hoch
Tech- Affinität (Ø, 0–6)	gering	mittel	mittel bis hoch
Präferenz menschlicher Kontakt (Ø, 0–7)	mittel	mittel bis hoch	gering
Präferenz digitaler Self-Service (Ø, 0–7)	sehr gering	sehr gering	mittel bis hoch
Alter	35–44 Jahre (Tendenz älter)	25–34 Jahre	25–34 Jahre (Tendenz älter)

Quelle: Eigene Darstellung

Auffällig ist insbesondere das größte Segment der Mainstream-Nutzer: Diese Gruppe ist jung, finanziell erfahren und nutzt KI privat. Dennoch bevorzugt sie im Finanzkontext deutlich menschliche Servicekanäle. Die Selbstentscheider zeigen die erwartete Affinität für Self-Service-Angebote, während die kleine Gruppe traditioneller Bankkunden digitale Lösungen weitestgehend ablehnt. Die mit k = 3 durchgeführte k-Means-Clusteranalyse weist einen Silhouette-Koeffizienten von 0,27 auf. Eine Zwei-Cluster-Lösung (Silhouette = 0,38) erscheint damit formal etwas trennschärfer. Aus inhaltlichen Erwägungen wird jedoch an der Einteilung in drei Cluster festgehalten, weil sie das kleine, aber im Kontext der Fondsdienstleistungshäuser in Deutschland besonders relevante, Segment der traditionellen Bankkunden explizit sichtbar macht. Die Prüfung der Outcome-Variablen zeichnet allerdings ein weniger deutliches Bild. Bei der Nutzungsbereitschaft für allgemeine Informationen treten klare Cluster-Unterschiede zutage ($p < .001$), während bei vertrauenssensibleren Anwendungsfällen die Unterschiede nur tendenziell ausgeprägt sind. Aggregierte Größen wie der SQS, die Datenschutzbedenken oder die Zahlungsbereitschaft variieren nicht nennenswert zwischen den Segmenten. Die Cluster bilden also keine klar abgegrenzten Kundentypen ab, sondern helfen vor allem dabei unterschiedliche Ausgangsprofile übersichtlich zu sortieren. Über alle Cluster hinweg zeigt sich ein konsistentes Muster: Die Bewertung der KI-Antworten in der Vignette fällt segmentübergreifend ähnlich positiv aus, während die Nutzungsbereitschaft segmentspezifisch variiert. Dies stützt die zentrale Beobachtung, dass wahrgenommene Qualität und tatsächliche Nutzungsbereitschaft im Finanzkontext zwei weitgehend unabhängige Größen sind.

MODERIERENDE EFFEKTE

Der Bewertungsvorteil der KI hält auch dann stand, wenn man individuelle Merkmale der Befragten berücksichtigt. Finanzerfahrung wirkt sich leicht positiv auf die Qualitätsbewertung aus ($\beta = 0{,}11$, $p < .01$), Technologieaffinität tendenziell ($\beta = 0{,}07$, $p = 0{,}06$). Alter und Geschlecht zeigen hingegen keinen erkennbaren Einfluss. Soziodemografische Merkmale spielen damit eine eher untergeordnete Rolle, während Erfahrung und Technologieoffenheit die Bewertung leicht mitprägen.

Auch das Zusammenspiel von tatsächlicher Antwortquelle und Technologieaffinität liefert keinen nennenswerten Effekt ($\beta = 0{,}02$, $p = 0{,}59$). Der Bewertungsvorteil der KI zeigt sich also über alle Stufen der Technologieaffinität hinweg gleichmäßig und auch der Technologie gegenüber skeptische Befragte bewerten die KI-generierten Antworten besser als von Menschen erstellte. Damit verdichtet sich das Bild, dass der Bewertungsvorsprung der KI durch die Qualität der Antworten selbst entsteht, nicht durch eine Gruppe besonders technikaffiner Befragter.

Diskussion

WIE TRAGFÄHIG IST DER QUALITÄTSVORSPRUNG

Im Kontext der Servicekonstellationen dieser Studie generieren moderne Sprachmodelle damit Inhalte, die sich Qualitativ mindestens auf Augenhöhe mit menschlichen Servicemitarbeitenden befinden. Die Nichtunterscheidbarkeit zwischen den Quellen scheint damit zuzunehmen oder ist bereits gegeben. Die geringe Erkennungsleistung der Stichprobe stützt diese Interpretation.

Es ist allerdings möglich, dass diese durchgängig bessere Bewertung in Teilen kognitiven Verzerrungen zuzuschreiben ist. Die Forschung kennt unter den Begriffen Algorithm Appreciation und Automation Bias beispielsweise die Tendenz, algorithmischen Systemen eine erhöhte Objektivität und Konsistenz zuzuschreiben (Jones-Jang & Park, 2022, S. 2; Logg et al., 2019, S. 93–99). Auch wenn den Teilnehmenden die Quelle nicht offengelegt wurde, könnte das thematische Setting der gesamten Studie implizite Erwartungen aktiviert haben, die KI-typische Stilmerkmale wie Strukturiertheit und sprachliche Präzision tendenziell positiver gewichten ließen. In der Stichprobe selbst lässt sich allerdings kein nennenswerter Effekt der vermuteten Quelle auf die Bewertung nachweisen. Tendenziell zeigt er sogar in die umgekehrte Richtung ($\beta = -0{,}09$). Die Befragten scheinen die Antworten nicht besser bewertet zu haben, nur weil sie diese für KI halten.

Eine einfachere Erklärung ist daher nicht auszuschließen. Die KI könnte im Studienkontext schlicht bessere Inhalte zu den gestellten Fragen generiert haben. Das eingesetzte Modell repräsentierte den zum Erhebungszeitpunkt aktuellen Stand der Technik, welcher gerade in der standardisierten, schriftlichen Inhaltsgenerierung Stärken aufweist. Es kann nicht länger grundsätzlich davon ausgegangen werden, dass menschliche Antworten in der Servicekommunikation qualitativ überlegen sind. Bemerkenswert ist in diesem Zusammenhang vor allem die Dimension Richtigkeit. Sie wurde von der Stichprobe mit $M = 3{,}45$ für die KI- und $M = 3{,}22$ für die menschlichen Antworten in beiden Quellen am kritischsten beurteilt. Auch hier ist der KI-Vorsprung signifikant (Wilcoxon $p < .001$) bei mittlerer Effektstärke ($r = 0{,}37$). Das zeigt einerseits, dass die Befragten auch ohne Kenntnis der Quelle sensibel für inhaltliche Korrektheit sind und dieses Kriterium nicht mit sprachlicher Eloquenz verwechseln. Andererseits offenbart er eine wichtigste Stellschraube für die Inhaltsgenerierung. Gerade für regulierte Branchen, in denen fachliche Korrektheit essenziell ist, hat dies erhebliche Konsequenzen für den produktiven Einsatz. Verfahren wie RAG und systematische Verifikation ausgegebener Informationen sind nicht optional, sondern bilden das Fundament jeder Inbetriebnahme.

BEDEUTUNG DER HOHEN KI-ERKENNUNGS-QUOTE

Eine weitere interessante Erkenntnis dieser Studie liegt in der Erkennungsleistung selbst. Über alle Vignetten hinweg tippten die Befragten deutlich häufiger auf KI als auf einen menschlichen Autor. Daraus ergibt sich eine ausgeprägte Asymmetrie. Die Befragten erkennen KI-generierte Antworten überproportional gut (Sensitivität 79 %), menschliche dagegen schlecht (33 %). Die individuelle Trefferquote liegt unter dem Zufallsniveau. Dieser KI-Vermutungs-Bias könnte Ausdruck der Erwartungshaltung sein, mit der Endnutzer heute in eine digitale Servicekommunikation gehen. Die Vermutung liegt nahe, dass das

thematische Setting der Studie Erwartungshaltung verstärkt hat. Dahinter könnte sich eine breitere Verschiebung andeuten. Formal saubere, schriftliche Auskünfte werden zunehmend als KI-typisch wahrgenommen, weil Nutzer im Alltag bereits umfassend an KI-vermittelte Texte gewöhnt sind. Die Erkennungsleistung sinkt also weniger, weil die KI menschlicher wird als vielmehr, weil sich die Erwartung an menschliche Kommunikation in formellen Kontexten verschiebt.

Strategisch lassen sich daraus zwei Perspektiven ableiten. Zum einen können Unternehmen davon ausgehen, dass auch ohne die, regulatorisch ohnehin vorgesehene, explizite Kennzeichnung viele Nutzer einen KI-Einsatz vermuten. Zum anderen werden Antworten nicht per se abgewertet, nur weil sie als KI-Erzeugnis identifiziert oder vermutet werden.

WARUM UNTERSCHEIDEN SICH EXPERIMENT- UND PRAXISERFAHRUNG

Die im Vignettenexperiment gemessenen positiven Bewertungen stehen in Kontrast zu den von denselben Teilnehmenden geäußerten Erfahrungen mit real eingesetzten KI-Servicelösungen. Diese werden insbesondere in ihrer Nützlichkeit und Problemlösungsfähigkeit deutlich kritischer bewertet. Diese Diskrepanz ist erklärungsbedürftig und für die Praxis hochrelevant.

Ein Erklärungsansatz könnte in unterschiedlichen eingesetzten Modellen begründet liegen. Das im Experiment eingesetzte GPT-4o entsprach dem zum Zeitpunkt der Studie aktuellen Stand der Technik. Viele in der Praxis erlebte KI-Servicelösungen beruhen jedoch noch auf älteren, regelbasierten oder hybriden Architekturen, die nicht das Leistungsniveau aktueller LLMs erreichen oder sich zumindest nur schwer miteinander vergleichen lassen (Karanikolas et al., 2025, S. 28). Die Studie misst damit zumindest in Teilen das technologische Potenzial moderner generativer KI und nicht den durchschnittlichen Implementierungsstand am Markt.

Auch der Bewertungsgegenstand selbst ist differenziert zu betrachten. Im Vignettenexperiment wurden isolierte Einzelantworten beurteilt. In der Praxis werden Serviceinteraktionen jedoch ganzheitlicher erlebt. In diese realen Verläufe fließen Faktoren wie Fehlinterpretationen, Wiederholungsschleifen, missglückte Eskalationen an menschliche Mitarbeitende, situativer Zeitdruck und persönliche Betroffenheit ein. Es ist plausibel, dass sich negativen Praxiserfahrung aus dieser umfassenderen Wahrnehmung heraus erklären lassen.

VERTRAUEN ALS VORGELAGERTE AKZEPTANZBARRIERE

Trotz des hohen Qualitätsurteils bricht die Bereitschaft die Technologie im Finanzkontext auch tatsächlich zu nutzen, mit zunehmender Datensensitivität, deutlich ein, wenn man die Teilnehmenden nach möglichen zukünftigen Nutzungsszenarien befragt. Die Wahrnehmung der Servicequalität und die Nutzungsbereitschaft scheinen damit weitestgehend voneinander unabhängig. Dieser Akzeptanzeinbruch ist nicht nur grafisch visualisierbar, sondern auch in jedem paarweisen Vergleich nachweisbar (Cochran’s $Q = 264{,}7$, $p < .001$, alle sechs McNemar-Vergleiche signifikant). Hinzu kommt, dass die Datenschutzbedenken in der Stichprobe deutlich über dem neutralen Skalenmittelwert liegen und somit kein Randthema, sondern eine weitere mögliche Akzeptanzbarriere markieren. Diese Beobachtungen wirken auf die vorgestellten klassischen theoretischen Modelle.

TAM und seine Erweiterungen modellieren die Nutzungsabsicht primär als Funktion wahrgenommener Nützlichkeit und Benutzerfreundlichkeit, die sich aus einer Vielzahl an Dimensionen speisen. Die hier dokumentierten Werte erfüllen grundsätzlich beide Bedingungen. Die KI-generierten Antworten werden als nützlich und als verständlich beurteilt. Gleichwohl bleibt die Nutzungsbereitschaft im sensiblen Bereich aus. Daraus lässt sich der Schluss ableiten, dass im untersuchten Kontext eine weitere, in den klassischen Modellen nicht vorgesehene Bedingung wirksam wird, aus der sich der zentrale theoretischen Beitrag dieses Papers ergibt. Das Vertrauen in die Sicherheit und den verantwortungsvollen Umgang mit den eigenen Daten könnte im vertrauenssensiblen Finanzkontext nicht als eine Dimension neben anderen wirken, sondern als hierarchisch vorgelagerte Bedingung fungieren. Solange diese Bedingung nicht erfüllt ist, übersetzt sich auch eine objektiv hohe wahrgenommene Nützlichkeit auch nicht in eine tatsächliche Nutzungsbereitschaft. Diese Feststellung steht im Einklang, mit dem in der Literatur diskutierten Phänomen der Algorithm Aversion. Diese Verzerrung beschreibt die Tendenz, gegenüber algorithmischen Entscheidungen eine geringere Fehlertoleranz zu zeigen als gegenüber menschlichen Urteilen (Dietvorst et al., 2015, S. 10). Sie wird auch durch das verbreitete Muster gestützt, dass Nutzer Datenschutz und Sicherheit selbst dann fordern, wenn sie objektiv keine Möglichkeit haben, die Einhaltung dieser Standards eigenständig zu überprüfen.

Konzeptionell lässt sich daraus schließen, dass Vertrauen im Finanzkontext nicht über inkrementelle Verbesserungen einzelner Qualitätsmerkmale aufgebaut werden kann. Es ist vielmehr ein übergreifender Faktor im Zusammenspiel aus organisationalen Strukturen, regulatorischen Rahmenbedingungen und kommunikativen Signalen.

HETEROGENES KUNDENBILD

Die Clusteranalyse zerlegt die Stichprobe in drei Segmente mit unterschiedlichen Eingangsprofilen.

1. Die Mainstream-Nutzer (47,3 %) sind jung, finanziell erfahren und nutzen KI privat ausnahmslos. Dennoch bevorzugen sie im Finanzkontext eher menschliche Kontaktkanäle.

2. Die Digitalen Selbstentscheider (41,0 %) zeigen das gegenteilige Bild und weisen die höchste Finanzerfahrung und Technologieaffinität auf, sowie klare Präferenz für Self-Service.

3. Die kleinste Gruppe der traditionellen Bankkunden (11,7 %) ist älter, technologisch zurückhaltend bis ablehnend und auf persönliche Betreuung orientiert.

Die identifizierten Segmente lassen sich gut an etablierte Adoptionstypologien aus der Forschung anschließen. Die Digitalen Selbstentscheider entsprechen den Convenience Seekers, die zügige und autonome digitale Interaktionen suchen. Die Mainstream-Nutzer ähneln den Prospective Adopters, die digitale Lösungen grundsätzlich annehmen, dabei aber menschliche Rückversicherung schätzen. Die traditionellen Bankkunden wiederum weisen Merkmale der Persistent Non-Adopters auf, die digitale Finanzkanäle weitgehend meiden (Lee et al., 2005, S. 414, 431; van Thiel & van Raaij, 2017, S. 96–97). Diese Anschlussfähigkeit stützt die Plausibilität der hier vorgenommenen Segmentierung, ohne dass damit der Anspruch einer empirisch validierten Verhaltenstypologie erhoben würde.

Methodisch ist eine ohnehin eine zurückhaltendere Interpretation geboten. Die Cluster sind in den Eingangsvariablen klar unterscheidbar, in den Outcome-Variablensind diese weniger signifikant. Vor allem die wahrgenommene Servicequalität oder die Datenschutzbedenken variieren kaum. Die Cluster sind daher weniger als feste Verhaltenstypologie zu lesen, denn als analytisches Strukturierungsinstrument. In der Praxis bleibt eine differenzierte Servicearchitektur sinnvoll, weil sie unterschiedlichen Kundenerwartungen Raum gibt. Segmentübergreifend bleiben aber grundsätzlich ähnliche Vertrauensbarrieren zu überwinden.

LIMITATIONEN DER STUDIE

Die hier diskutierten Ergebnisse sind in ihrer Reichweite und Aussagekraft durch das Studiendesign begrenzt. Die Limitationen dieser Studie werden deshalb noch einmal ausdrücklich aufgeführt.

Die Stichprobe wurde nicht zufallsbasiert gezogen, sondern entstand vorrangig über die Veröffentlichung auf SurveyCircle, einer renommierte Online-Community für die Teilnahme an wissenschaftlichen Studien (SurveyCircle, 2026). Selbst- und Selektionseffekte können das Ergebnis deshalb beeinflussen. Insbesondere eine Überrepräsentation digital affiner Befragter ist möglich. Verstärkende zeigt die Stichprobe eine deutliche Unterrepräsentation älterer Umfrageteilnehmer und eine Überrepräsentation akademischer Abschlüsse. Die vorgenommenen inferenzstatistischen Tests sind in diesem Kontext zu lesen. Sie belegen die Robustheit beobachteter Muster gegenüber Zufallsschwankungen innerhalb der Stichprobe und liefern Effektstärken, die unabhängig von Stichprobenrepräsentativität interpretierbar bleiben. Eine Hochrechnung der konkreten Anteilswerte auf die Gesamtpopulation der Fondsdienstleistungskunden ist damit jedoch nicht möglich. Die KI-generierten-Antworten basieren auf einem einzigen Modell, das bereits heute durch Nachfolgemodelle ersetzt wurde (openAI, 2026). Vergleiche zwischen Modellen oder Architekturen sind im Rahmen dieser Studie deshalb nicht möglich und Aussagen über die KI als Gesamtphänomen wären eine unzulässige Verallgemeinerung. Schließlich erlaubt das quasi-experimentelle Design Aussagen über Zusammenhänge, nicht über Kausalitäten. Die in der Diskussion vorgeschlagenen Erklärungen, wie die vorgelagerte Rolle von Vertrauen oder möglicher KI-Vermutungs-Verzerrungen sind theoretisch plausibel und empirisch nahegelegt, aber kausal nicht abschließend belegbar.

Implikationen

IMPLIKATIONEN FÜR DIE FORSCHUNG

Aus den Studienergebnissen lassen sich Implikationen für die eingeführten Modelle der Servicequalität und der Technologieakzeptanz ableiten.

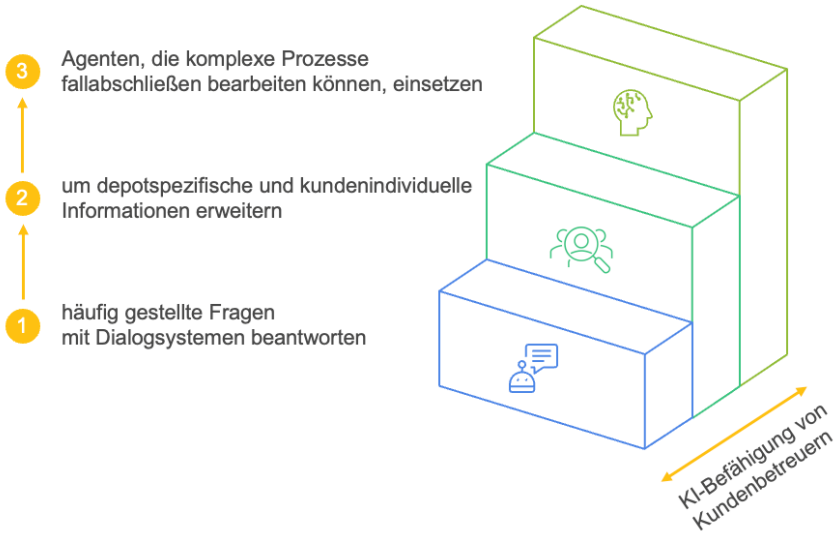
Die zentralen SERVQUAL-Dimensionen bleiben tragfähig, erfahren aber eine semantische Aktualisierung. Assurance kann als systemischen Kompetenz verstanden werden, hochwertige Informationen zu generieren. Ein Punkt, der durch die hohen Bewertungen in Richtigkeit, Vollständigkeit und Verständlichkeit sowie durch die qualitative Forderung nach „100 % richtigen Informationen" gestützt wird (Warjiyono et al., 2025, S. 129–130). Empathy verschiebt sich von emotionaler Nähe hin zur präzisen Erfassung von Kundenanliegen und kontextgerechter Individualisierung (Glikson & Woolley, 2020, S. 647–650). Reliability erweitert sich um inhaltliche Konsistenz und die Vermeidung von Halluzinationen. Die E-S-QUAL-Dimensionen werden ebenfalls bestätigt, allerdings legt die stark sinkende Nutzungsabsicht bei steigender Datensensitivität nahe, dass Privacy im Finanzkontext keine Dimension unter mehreren ist, sondern als hierarchisch vorgelagerte Akzeptanzbedingung wirkt. Eine empirisch fundierte Modellrevision sollte diese hierarchische Logik abbilden. Die Überführung wahrgenommener Nützlichkeit in tatsächliche Nutzung kann gemäß den Ergebnissen dieser Studie durch fehlende Vertrauenskomponenten blockiert werden. Diesen Mechanismus bilden die klassischen Modelle nur bedingt ab (Venkatesh & Davis, 2000). Die Bedeutung moderierender Variablen im Sinne des UTAUT2 (Venkatesh et al., 2012) bestätigt sich hingegen. Finanzerfahrung und Technologieoffenheit beeinflussen die Akzeptanz. Die Präferenz für menschliche Ansprechpartner bei komplexen Anliegen deutet im Sinne des KIAM zudem darauf hin, dass Nutzer KI im Fondsdienstleistungskontext aktuell primär als technisches Werkzeug wahrnehmen, mit der Folge, dass funktionale Kriterien wie Korrektheit und Sicherheit dominieren (Scheuer, 2020, S. 63). Das Extended TAM for Generative AI liefert schließlich den Erklärungsansatz für die beobachtete Risikoaversion. KI-spezifische Risikodimensionen und Datenschutzbedenkenwirken als direkte negative Treiber auf die Nutzungsabsicht (Lim et al., 2025, S. 186).

Daraus lässt sich weiterer Forschungsbedarf ableiten. Zunächst sollte die vorgeschlagene hierarchische Vertrauensstruktur in repräsentativen Stichproben empirisch geprüft werden. Weiterhin mangelt es an Längsschnittdaten dazu, ob sich der Akzeptanzeinbruch bei steigender Datensensitivität durch wiederholte positive Erfahrung abbauen lässt oder verfestigt. Schließlich dürften die hier beobachteten Vertrauensbarrieren, mit der Weiterentwicklung zu agentischen System weiter wachsen, woraus sich ein vollständig neues Forschungsfeld erschließt.

IMPLIKATIONEN FÜR DIE PRAXIS

Aus diesen Erkenntnissen lassen sich drei zusammenhängende Handlungsfelder für Fondsdienstleistungsunternehmen ableiten, die den produktiven Einsatz generativer KI verfolgen. Ein iteratives Implementierungsmodell, eine segmentspezifische Ausgestaltung und ein systematischer Vertrauensaufbau.

Abbildung 6: Iterativer Ausbau von KI-gestützten Lösungen



Quelle: Eigene Darstellung

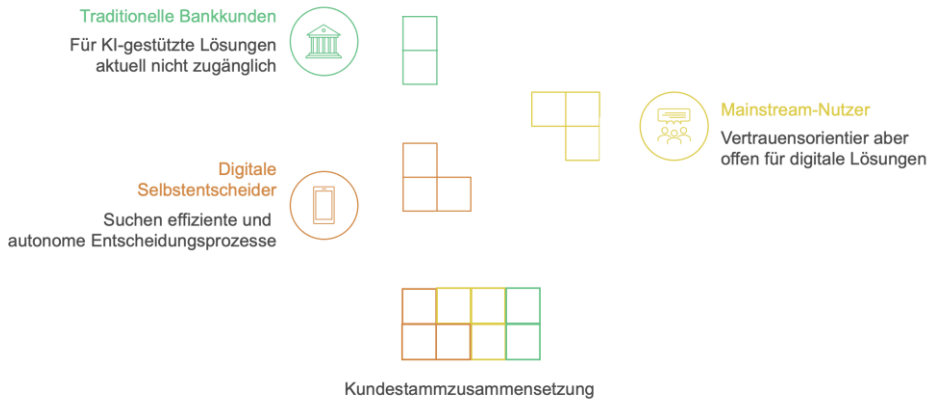
Es bietet sich an, im ersten Schritt Anfragen von geringer Komplexität, wie die Erklärungen von Finanzbegriffen oder allgemeine Produktinformationen zu automatisieren, die noch keinen persönlichen Kundenbezug aufweisen. Mit 87 % weisen diese Anwendungsfälle die höchste Akzeptanz innerhalb der Stichprobe auf. So lassen sich erste Effizienzgewinne erzielen, ohne das Kundenvertrauen zu gefährden. RAG-Verfahren, die das Sprachmodell auf verifizierte interne Wissensbasen stützen, sollten hier mitgedacht werden, um Halluzinationen zu minimieren (Yang et al., 2023, S. 5)

In der zweiten Stufe kann die prozessuale Komplexität durch standardisierte Vorgänge mit begrenztem Personenbezug erhöht werden. Darunter fallen etwa Statusabfragen oder Dokumentenanforderungen. Voraussetzung ist die Anbindung der KI an Backend-Systeme wie CRM- und Kernbankanwendungen. Legacy-Strukturen bilden hier die zentrale Integrationsbarriere (Yerra, 2025, S. 881–882). In dieser Phase wird bereits ein sogenannter „Human-in-the-Loop"-Eskalationsmechanismus relevant, der dem in der Stichprobe geäußerten Wunsch nach hybriden Lösungen entspricht und zugleich regulatorische Verantwortungszuordnung sicherstellt (Zetzsche et al., 2020, S. 38).

Schließlich können agentenbasierte Systeme größere Teile der Prozesslandschaft eigenständig orchestrieren und fallabschließend automatisieren (Kshetri, 2025, S. 20). Im Unterschied zu den hier untersuchten reaktiven Chatbots verfolgen sie eigenständig komplexe Ziele über mehrere Schritte hinweg und erhalten schreibende Systemrechte. KI wird damit vom Konsumenten von Daten zum aktiven Akteur in der Prozesslandschaft.

Phasenübergreifend kann KI, überall dort wo der menschliche Berater die Kundeninteraktion führt, als durch die Aufbereitung von Daten oder „Next-Best-Action"-Empfehlungen entlasten (Wohleb, 2025). Der Mensch bleibt verantwortliche Instanz, wird aber von repetitiven Aufgaben befreit (Bickel, 2025).

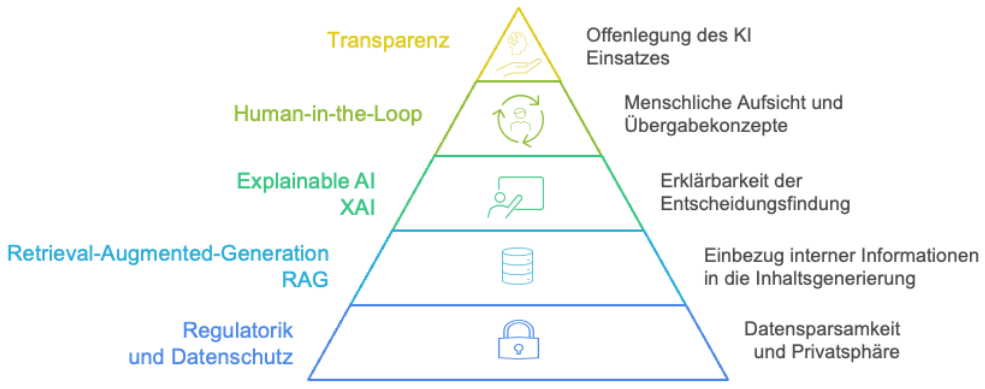
Abbildung 7: Kundensegmente innerhalb der Stichprobe



Quelle: Eigene Darstellung

Die identifizierten Nutzersegmente legen nahe, dass eine differenzierte Servicearchitektur den unterschiedlichen Kundenpräferenzen besser entgegenkommen kann, als eine einheitliche Ausbringungsstrategie. Für die Selbstentscheider kann ein Digital First-Angebot verfolgt werden, in dem KI-gestützte Self-Services werden proaktiv als primäre Anlaufstelle positioniert werden. Begleitet werden können diese mit Botschaften, die die Autonomiegewinne in den Vordergrund stellen. Für die Mainstream-Nutzer empfiehlt sich ein hybrides Modell. KI kann als erste Anlaufstelle mit niedrigschwelliger, jederzeit verfügbarer Eskalationsoption an menschliche Mitarbeitende fungieren. Für die traditionellen Bankkunden bleibt ein Human First-Angebot zentral, wobei KI hier zunächst auf Prozessebene wirken kann, ohne in der direkten Interaktion sichtbar zu werden. Wo eine segmentbasierte Steuerung mangels vorhandener Daten zur Kundensegmentierung nicht möglich ist, bietet sich eine modulare Architektur an, die alle drei Ausbaustufen parallel verfügbar macht. Diese überlässt es dem individuellen Nutzungsverhalten der Kunden den Automatisierungsgrad implizit zu wählen. Dieses Modell kann wahrgenommene Risiken reduzieren und ermöglicht eine schrittweise Gewöhnung an die neue Technologie

Abbildung 8: Maßnahmen zur Stärkung der Vertrauenswürdigkeit



Quelle: Eigene Darstellung

Wenn Vertrauen als vorgelagerte Akzeptanzbedingung wirkt, ist dessen systematischer Aufbau ein weiterer wichtiger Baustein der Implementierungsstrategie. Da LLMs keine eigene Verantwortung übernehmen können (Newen et al., 2025, S. 14) muss diese Lücke organisational, technisch und regulatorisch geschlossen werden.

Regulatorik und Datenschutz bilden das Fundament. Da rund die Hälfte der Befragten Datenschutzbedenken äußert, sind glaubwürdige Signale, wie Zertifizierungen, Datensparsamkeit oder klare Auskunftsrechte wichtig. Technisch lässt sich dieses Prinzip etwa über Federated Learning verankern, bei dem sensible Kundendaten das eigene System nicht verlassen (Rehman & Gaber, 2021, S. 2). Retrieval-Augmented Generation (RAG) stützt das Modell auf verifizierte interne Wissensquellen und minimiert damit Halluzinationen (Yang et al., 2023, S. 5). Gerade im regulierten Finanzkontext sollte dies eine zwingende Voraussetzung darstellen, keine optionale Ergänzung. Explainable AI (XAI) reduziert den Black-Box-Charakter der Modelle und macht Entscheidungsgrundlagen für Kunden und Aufsichtsbehörden nachvollziehbar (Barredo Arrieta et al., 2020, S. 108; Černevičienė & Kabašinskas, 2024, S. 32). Human-in-the-Loop sichert die menschliche Aufsicht an entscheidungsrelevanten Schnittstellen. Bei unzureichender Konfidenz sollte eine saubere Übergabe an menschliche Kundenbetreuer vorgesehen sein (Choung et al., 2023, S. 736; Newen et al., 2025, S. 5). Zugleich erfüllt der Mechanismus regulatorische Anforderungen an Verantwortungszuordnung (Zetzsche et al., 2020, S. 38). Schließlich verpflichtet der EU AI Act zur Offenlegung von KI-Interaktionen (Veale & Zuiderveen Borgesius, 2021, S. 106). Da viele Nutzer ohnehin eine KI-Autorenschaft vermuten, bestätigt die Kennzeichnung eine bereits vorhandene Erwartung und kann aktiv als vertrauensbildendes Signal eingesetzt werden.

Fazit

Das hier aufgezeigte Spannungsfeld ist für die strategische Diskussion über generative KI in der Fondsdienstleistungsbranche zentral. Auf der einen Seite zeigen die Daten, dass moderne Sprachmodelle in der standardisierten Servicekommunikation ein Qualitätsniveau erreicht haben, das von Menschen erstellten Antworten ebenbürtig ist oder diese übertrifft und die Ursprungsquelle von Kunden nicht mehr zuverlässig voneinander unterschieden werden kann. Auf der anderen Seite wird deutlich, dass dieses Qualitätsniveau allein nicht ausreicht, um Akzeptanz im vertrauenssensiblen Finanzkontext zu erzeugen. Die Nutzungsbereitschaft folgt einer eigenen Logik, die durch die Datensensitivität des Anliegens dominiert und von Datenschutzbedenken beeinflusst wird. Das objektive Potenzial generativer KI im Fondsdienstleistungskontext scheint die subjektive Akzeptanz der Nutzer aktuell also noch zu übersteigen. Die Herausforderung der Branche ist deshalb weniger eine technologische als eine vertrauens- und designorientierte. Die Frage lautet nicht mehr, ob die KI gut genug ist sondern wie Implementierungs-, Kommunikations- und Vertrauensarchitekturen gestaltet sein müssen, damit Kunden der vorhandenen Qualität auch die Bereitschaft entgegenbringen, sie für ihre sensiblen Anliegen zu nutzen.

Daraus ergeben sich Konsequenzen für Forschung und Praxis gleichermaßen. Etablierte Akzeptanzmodelle sollten um eine kontextuelle und hierarchisch vorgelagerte Vertrauensdimension erweitert werden. In der praktischen Umsetzung sollte ein reibungsloses Zusammenspiel aus iterativen Ausbaustufen, segmentspezifischen Strategien und systematischem Vertrauensaufbau etabliert werden. Mit Blick auf die technologische Weiterentwicklung ist davon auszugehen, dass sich diese Herausforderung weiter verschärfen wird. Mit dem Übergang von generativer zu agentischer KI, also Systemen, die nicht nur Antworten liefern, sondern eigenständig Handlungen ausführen und Prozesse fallabschließend bearbeiten (Kshetri, 2025, S. 19), erweitert sich der Vertrauensbedarf erheblich. Die hier identifizierten Barrieren können damit weiter an Bedeutung gewinnen. Die Branche sollte den Aufbau von Vertrauen in neue Technologien deshalb nicht als vorübergehende Begleiterscheinung der KI-Adaption behandeln, sondern als strukturelle Daueraufgabe ihrer Digitalisierungsstrategie verstehen und konsequent in eine stufenweise Implementierung sowie in die an den jeweiligen Nutzerclustern ausgerichtete Angebotsgestaltung integrieren.

Literaturverzeichnis:

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Baumann, L. (2025). Generative KI Chatbots: Die Zukunft der Kundenbetreuung. In K. Altenfelder, S. Kieffer-Radwan, & D. Schönfeld (Hrsg.), Services Management und Künstliche Intelligenz: Grundlagen und Anwendungsfelder für den Einsatz von KI-unterstützten Services (S. 55–68). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-46665-7_4

Bickel, M. (2025). KI in der Finanzberatung: Chancen & Veränderungen. Sachkundegurus. https://www.sachkundegurus.de/journal/der-einfluss-von-ki-auf-finanzberatung-und-kundenkommunikation

Blüthgen, C. (2025). Technische Grundlagen großer Sprachmodelle. Die Radiologie, 65(4), 227–234. https://doi.org/10.1007/s00117-025-01427-z

Boßow-Thies, S., & Krol, B. (Hrsg.). (2022). Quantitative Forschung in Masterarbeiten: Best-Practice-Beispiele wirtschaftswissenschaftlicher Studienrichtungen. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-35831-0

Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. Artificial Intelligence Review, 57(8), 216. https://doi.org/10.1007/s10462-024-10854-8

Choung, H., David, P., & Ross, A. (2023). Trust and ethics in AI. AI & SOCIETY, 38(2), 733–745. https://doi.org/10.1007/s00146-022-01473-4

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114–126. https://doi.org/10.1037/xge0000033

Ferraro, C., Demsar, V., Sands, S., Restrepo, M., & Campbell, C. (2024). The paradoxes of generative AI-enabled customer service: A guide for managers. Business Horizons, 67(5), 549–559. https://doi.org/10.1016/j.bushor.2024.04.013

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. Academy of Management Annals, 14(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Glock, S. (2024). KI im Kundenservice – Auslaufmodell Kundenhotline? BM: Bank Und Markt, (7), 34–36. Business Source Ultimate (178406999).

Hafner, N., & Hundertmark, S. (2024). Generative AI in Finance Studie 2024. HSLU Hochschule Luzern.

Hartmann-Wendels, T., Pfingsten, A., & Weber, M. (2019). Bankbetriebslehre. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-58290-9

Höhne, M. M.-C. (2021). Nachvollziehbare Künstliche Intelligenz: Methoden, Chancen und Risiken: Black Box zu White Box. Datenschutz und Datensicherheit - DuD, 45(7), 453–456. https://doi.org/10.1007/s11623-021-1470-x

Jones-Jang, S. M., & Park, Y. J. (2022). How do people react to AI failure? Automation bias, algorithmic aversion, and perceived controllability. Journal of Computer-Mediated Communication, 28(1), zmac029. https://doi.org/10.1093/jcmc/zmac029

Kamath, U., Keenan, K., Somers, G., & Sorenson, S. (2024). Large Language Models: A Deep Dive: Bridging Theory and Practice. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-65647-7

Karanikolas, N. N., Manga, E., Samaridi, N., Stergiopoulos, V., Tousidou, E., & Vassilakopoulos, M. (2025). Strengths and Weaknesses of LLM-Based and Rule-Based NLP Technologies and Their Potential Synergies "2279.

Kshetri, N. (2025). The Rise of Agentic AI in Finance: Opportunities, Risks, and Human-Centric Integration. IT Professional, 27(4), 19–24. https://doi.org/10.1109/MITP.2025.3585227

Lee, E., Kwon, K., & Schumann, D. W. (2005). Segmenting the non-adopter category in the diffusion of internet banking. International Journal of Bank Marketing, 23(5), 414–437. https://doi.org/10.1108/02652320510612483

Lim, J. S., Shin, D., Lee, C., Kim, J., & Zhang, J. (2025). The Role of User Empowerment, AI Hallucination, and Privacy Concerns in Continued Use and Premium Subscription Intentions: An Extended Technology Acceptance Model for Generative AI. Journal of Broadcasting & Electronic Media, 69(3), 183–199. https://doi.org/10.1080/08838151.2025.2487679

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Malaquias, F. F., & Hwang, Y. (2016). Trust in mobile banking under conditions of information asymmetry: Empirical evidence from Brazil. Information Development, 32(5), 1600–1612. https://doi.org/10.1177/0266666915616164

Miko-Schefzig, K. (2022). Forschen mit Vignetten:Gruppen, Organisationen, Transformation. Beltz/Juventa. BASE (edsbas.F7EA2073). https://research.ebsco.com/linkprocessor/plink?id=fc2a25b9-4656-3b72-a7ae-481beda097b4

Newen, C., Müller, E., & Newen, A. (2025). Trust and Uncertainties: Characterizing Trustworthy AI Systems Within a Multidimensional Theory of Trust. Topoi. https://doi.org/10.1007/s11245-025-10287-0

openAI. (2026, März 10). Retiring GPT-4o, GPT-4.1, GPT-4.1 mini, and OpenAI o4-mini in ChatGPT. https://openai.com/index/retiring-gpt-4o-and-older-models/

Parasuraman, A. P., Zeithaml, V., & Berry, L. (1988). SERVQUAL A Multiple-item Scale for Measuring Consumer Perceptions of Service Quality. Journal of Retailing, 64, 12–40.

Parasuraman, A., Zeithaml, V. A., & Malhotra, A. (2005). E-S-QUAL: A Multiple-Item Scale for Assessing Electronic Service Quality. Journal of Service Research, 7(3), 213–233. https://doi.org/10.1177/1094670504271156

Peters, A. (2025). Herausforderungen für Banken im Rahmen von Digitalisierung, KI und Cybersecurity. Zeitschrift für das Gesamte Kreditwesen, 78(5), 24–27. Business Source Ultimate (183299761).

Raab, W. (Hrsg.). (2019). Grundlagen des Investmentfondsgeschäftes. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-24155-1

Raff, D. T., Billen, D. P., & Hofmann, J. (2025). KI-Nutzerakzeptanz am Beispiel von Chatbots im Online-Banking. Marketing Review St. Gallen.

Rehman, M. H. U., & Gaber, M. M. (Hrsg.). (2021). Federated Learning Systems: Towards Next-Generation AI (Bd. 965). Springer International Publishing. https://doi.org/10.1007/978-3-030-70604-3

Scheuer, D. (2020). Akzeptanz von Künstlicher Intelligenz: Grundlagen intelligenter KI-Assistenten und deren vertrauensvolle Nutzung. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-29526-4

Seemann, M. (2022). Künstliche Intelligenz, Large Language Models, ChatGPT und die Arbeitswelt der Zukunft. Working Paper Forschungsförderung, Nummer 304, September 2023.

Senger, R. (2025). Chatbots im Kundenservice: Selfservice-Lösungen. In K. Altenfelder, S. Kieffer-Radwan, & D. Schönfeld (Hrsg.), Services Management und Künstliche Intelligenz: Grundlagen und Anwendungsfelder für den Einsatz von KI-unterstützten Services (S. 39–54). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-46665-7_3

Supriyanto, A., Wiyono, B. B., & Burhanuddin, B. (2021). Effects of service quality and customer satisfaction on loyalty of bank customers. Cogent Business & Management, 8(1), 1937847. https://doi.org/10.1080/23311975.2021.1937847

SurveyCircle. (2026). SurveyCircle – Studienteilnehmer finden in der größten Community für Online-Forschung | Jetzt kostenlos mitmachen & Umfrageteilnehmer erhalten. SurveyCircle.com. https://www.surveycircle.com/de/

van Thiel, D., & van Raaij, F. (2017). Targeting the robo-advice customer: The development of a psychographic segmentation model for financial advice robots. Journal of Financial Transformation, 46, 88–104.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need.

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act. SocArXiv. https://doi.org/10.31235/osf.io/38p5f

Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. Decision Sciences, 39(2), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science, 46(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology1. MIS Quarterly, 36(1), 157–178. https://doi.org/10.2307/41410412

Voge, V. (2020). Künstliche Intelligenz und Ethik. Aktuelle Themen der KI WS 20/21, AM 3.

Warjiyono, W., Hadiyanto, H., & Nugraheni, D. M. K. (2025). AI chatbot quality in customer service: Extending measurement models with conversational capability. Edelweiss Applied Science and Technology, 9(10), 1416–1436. https://doi.org/10.55214/2576-8484.v9i10.10674

Wohleb, D. (2025, Juni 3). Finanzberatung: So profitieren Berater und Kunden von Künstlicher Intelligenz. handelsblatt.com. https://www.handelsblatt.com/finanzen/banken-versicherungen/banken/finanzberatung-so-profitieren-berater-und-kunden-von-kuenstlicher-intelligenz/100130187.html

Yang, H., Liu, X.-Y., & Wang, C. D. (2023). FinGPT: Open-Source Financial Large Language Models (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2306.06031

Yerra, P. V. K. (2025). Agentic AI Framework for Automating Legacy Core- Banking Operations and Regulatory Reporting Pipe- lines. Journal of Information Systems Engineering and Management, (10(2)).

Zetzsche, D. A., Douglas, A., Buckley, R., & Tang, B. W. (2020). Artificial Intelligence in Finance: Putting the Human in the Loop. CFTE Academic Paper Series, 1 / 2020.