WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

BACHELORARBEIT

Suchverhalten im Wandel: Ein experimenteller Vergleich klassischer und KI-basierter Informationssuche

verfasst von

Clemens Höller

angestrebter akademischer Grad

Bachelor of Science (WU), BSc (WU)

Matrikelnummer:	**12214487**
Studienrichtung:	Wirtschafts- und Sozialwissenschaften
Betreut von:	Univ.-Prof. Dr. Siham EL KIHAL

Wien, Juni 2025

Inhaltsverzeichnis

Abstract

In einer zunehmend digitalisierten Welt ist die Online-Informationssuche ein entscheidender Bestandteil des Konsumentenverhaltens und beeinflusst maßgeblich, wie Kaufentscheidungen vorbereitet und getroffen werden. Diese Bachelorarbeit untersucht, wie Konsument:innen Google und ChatGPT bei Low- und High-Involvement in der Informationssuche wahrnehmen und bewerten. Ziel ist es, Unterschiede in Vertrauen, Nutzererfahrung, Entscheidungssicherheit und wahrgenommener Personalisierung zwischen beiden Suchsystemen aufzuzeigen. Für Unternehmen und Forschende ergeben sich daraus wichtige Erkenntnisse über das digitale Suchverhalten und die Akzeptanz neuer Technologien.

Die Untersuchung basiert auf einem Online-Experiment mit 212 Teilnehmenden, die zufällig einer von vier Suchaufgaben (Google/ChatGPT x Low/High-Involvement) zugeteilt wurden. Bewertet wurden die jeweiligen Sucherfahrungen mithilfe standardisierter Skalen. Die zugrunde liegenden Hypothesen wurden auf Basis einer umfassenden Literaturauswertung zu Konsumentenverhalten, Involvement und digitalen Suchsystemen entwickelt.

Die Ergebnisse zeigen, dass Google bei High-Involvement-Suchen als vertrauenswürdiger und aktueller eingeschätzt wird, während ChatGPT bei Low-Involvement-Aufgaben eine angenehmere und effizientere Nutzung bietet. ChatGPT wird zudem durchgehend als stärker personalisiert erlebt. In Bezug auf Entscheidungssicherheit zeigt sich ein leichter Vorteil für ChatGPT, während die Wiederverwendungsbereitschaft bei beiden Plattformen ähnlich ausfällt.

Diese Ergebnisse verdeutlichen, dass die Wahrnehmung digitaler Suchsysteme stark vom Entscheidungskontext abhängig ist. Für die Entwicklung nutzerzentrierter Suchlösungen und den gezielten Einsatz von KI im digitalen Marketing können sich daraus konkrete Implikationen ergeben.

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

Ob es um die Wahl eines Arztes geht oder um die Suche nach einem passenden Café zum dortigen Arbeiten, immer mehr Konsument:innen greifen auf digitale Kanäle zurück, um vor einer Entscheidung Informationen zu sammeln. Was früher über persönliche Empfehlungen, Fachzeitschriften oder gedruckte Medien geschah, wird heute durch eine schnelle Suchanfrage im Internet ersetzt. Die zunehmende Digitalisierung hat unser alltägliches Verhalten grundlegend verändert, insbesondere in Hinblick auf das Konsumverhalten und die Art, wie Informationen eingeholt und Entscheidungen getroffen werden.

Lange Zeit galt Google als der zentrale Zugangspunkt zur Informationssuche. Die einfache Bedienbarkeit, die gewohnte Ergebnisdarstellung in Form von Linklisten und das breite Vertrauen in die Suchmaschine haben dazu geführt, dass sich der Begriff "googeln" fest in der Alltagssprache verankert hat. Doch in den letzten Jahren haben sich die technischen Möglichkeiten weiterentwickelt. Mit der Verbreitung von Künstlicher Intelligenz, insbesondere sogenannter Large Language Models (LLMs), ist eine neue Form der Informationssuche entstanden. Systeme wie ChatGPT ermöglichen es Nutzer:innen bereits, komplexe Fragen in natürlicher Sprache zu stellen und direkt strukturierte, ausführliche Antworten zu erhalten. Die Interaktion erfolgt nicht mehr über einzelne Schlagwörter, sondern über ganze Sätze und Dialoge, die kontextbezogen weitergeführt werden können.

Diese Veränderung wirft zentrale Fragen auf: Wie nehmen Konsument:innen solche Systeme im Vergleich zur klassischen Google-Suche wahr? Welche Methode wird als hilfreicher oder vertrauenswürdiger empfunden? Und wie beeinflusst der Kontext einer Entscheidung, also ob es sich um eine eher unwichtige Alltagsentscheidung oder eine risikobehaftete, persönlich bedeutsame Entscheidung handelt, die Bewertung des Suchprozesses?

Außerdem zeigt sich in der Literatur, dass das sogenannte Involvement-Level, also das Ausmaß an persönlicher Relevanz und wahrgenommenem Risiko einer Entscheidung, einen starken Einfluss auf die Komplexität der Informationsverarbeitung hat. In Situationen mit hohem Involvement suchen Nutzerinnen und Nutzer gezielt, vergleichen verschiedene Quellen kritisch und investieren mehr Zeit. In Low-Involvement-Situationen hingegen steht oft der schnelle Zugriff auf eine zufriedenstellende Antwort im Vordergrund, was den Erfolg eines Suchsystems an Effizienz und Komfort knüpft. Genau an dieser Stelle setzt diese Arbeit an, indem sie

untersucht, ob Google oder ChatGPT in den jeweiligen Entscheidungssituationen unterschiedlich bewertet werden und welche Faktoren diese Wahrnehmung beeinflussen.

Die Relevanz der Thematik ergibt sich nicht nur aus der zunehmenden Nutzung von KI im Alltag, sondern auch aus einem wissenschaftlichen Forschungsinteresse. Bisher gibt es nur vereinzelt empirische Studien, die den direkten Vergleich zwischen klassischen Suchmaschinen und LLM-basierten Systemen im Alltag der Konsument:innen analysieren. Besonders in Kombination mit psychologischen Konstrukten wie dem Involvement besteht eine Forschungslücke, die durch diese Arbeit adressiert wird.

Das Ziel dieser Bachelorarbeit ist es daher, Unterschiede in der Nutzerwahrnehmung und Bewertung von Google und ChatGPT bei der digitalen Informationssuche herauszuarbeiten. Im Fokus stehen Aspekte wie Vertrauen, Informationsqualität, Effizienz, wahrgenommene Entscheidungssicherheit und Personalisierung. Dabei wird der Einfluss des Involvements als moderierende Variable betrachtet, um ein differenziertes Bild zu erhalten, wie Konsument:innen Suchtechnologien je nach Entscheidungskontext erleben und bewerten.

Die Arbeit beginnt mit einer theoretischen Grundlage, in der zentrale Konzepte des Konsumentenverhaltens, des Involvements und der digitalen Informationssuche systematisch aufgearbeitet werden. Darauf aufbauend wird das methodische Vorgehen erläutert, bei dem ein Online-Experiment mit vier Szenarien durchgeführt wurde. Anschließend werden die empirischen Ergebnisse dargestellt und zur Überprüfung der entwickelten Hypothesen herangezogen. Abschließend folgt eine Interpretation der Resultate im Lichte der Theorie, ergänzt durch einen Vergleich mit bestehenden Studien, eine Reflexion über Limitationen sowie einen Ausblick auf zukünftige Forschung. Dabei werden fundierte wissenschaftliche Erkenntnisse ebenso wie praxisnahe Einblicke in die Nutzung von Google und ChatGPT in verschiedenen Entscheidungssituationen aufgezeigt.

2 Theoretische und konzeptionelle Grundlagen

Dieses Kapitel stellt die theoretischen und konzeptionellen Grundlagen dieser Arbeit dar. Ziel ist es, das Suchverhalten von Konsument:innen im digitalen Raum im Kontext unterschiedlicher Involvement-Stufen sowie verschiedener Suchsysteme, klassischer Suchmaschinen (Google) und Large Language Models (LLMs, konkret ChatGPT), einzuordnen. Dazu werden zentrale Begriffe, Modelle und bisherige empirische Erkenntnisse aus dem Forschungsfeld des Konsumentenverhaltens, der Involvement-Theorie sowie der digitalen Informationssuche aufgearbeitet. Dieses Kapitel dient der Einbettung der Forschungsfrage in den aktuellen wissenschaftlichen Diskurs und liefert die konzeptuelle Basis zur Entwicklung der Hypothesen. Durch die strukturierte Gegenüberstellung der beiden Suchmethoden soll ein vertiefendes Verständnis dafür geschaffen werden, wie sich digitale Entscheidungssituationen durch neue Technologien wie ChatGPT verändern.

2.1 Konsumentenverhalten in digitalen Entscheidungssituationen

Die fortschreitende Digitalisierung hat viele Lebensbereiche nachhaltig verändert. Davon betroffen ist auch das Verhalten von Konsument:innen, das sich zunehmend in den digitalen Raum verlagert. Digitale Informationssuche ermöglicht es Konsument:innen, Produkte und Dienstleistungen jederzeit und ortsunabhängig zu suchen, zu vergleichen und zu erwerben. Diese Entwicklung bringt veränderte Anforderungen an Entscheidungsprozesse und Informationsverhalten mit sich. Zentrale Fragen sind dabei, wie Konsument:innen digitale Informationen wahrnehmen, Entscheidungen treffen und mit den kognitiven sowie emotionalen Anforderungen digitaler Umgebungen umgehen (Hoffmann & Akbar, 2024, S. 196-197).

Der klassische Kaufentscheidungsprozess, bestehend aus der Problemerkennung, der eigentlichen Kaufentscheidung und der anschließenden Nutzung des Produkts oder der Dienstleistung, bleibt zwar bestehen, wird im digitalen Umfeld jedoch durch neue Rahmenbedingungen geprägt. Einflussfaktoren wie das individuelle Involvement, die Art der Entscheidung, die Wahl des Informationskanals sowie die Absicht der Konsument:innen spielen hierbei eine zentrale Rolle. Digitale Entscheidungssituationen zeichnen sich dadurch aus, dass Nutzer:innen flexibel zwischen reiner Informationssuche und konkreter Kaufabsicht wechseln können. Dadurch kommt es häufig zu Überschneidungen zwischen den einzelnen Entscheidungsphasen, sodass Informationsaufnahme, Bewertung und Entscheidung

eng miteinander verknüpft sind (Hoffmann & Akbar, 2024, S. 9; Swoboda & Schramm-Klein, 2025, S. 188-191).

Eine zentrale Herausforderung der digitalen Informationssuche ist die enorme Datenmenge, mit der Konsument:innen konfrontiert werden. Eine hohe visuelle Komplexität erschwert es, relevante von irrelevanten Informationen zu unterscheiden. Dies kann zu einer kognitiven Überforderung führen, die den wahrgenommenen Aufwand steigert und letztlich die Zufriedenheit mit dem Suchergebnis mindert. Dieses Phänomen ist als *Information Overload* bekannt und stellt eine häufige Barriere im digitalen Entscheidungsprozess dar (Abrardi et al., 2022, S. 986; Swoboda & Schramm-Klein, 2025, S. 115). Gleichzeitig können emotionale Reize im digitalen Raum entscheidungsrelevant sein. Konsument:innen interpretieren Gefühle häufig als informative Signale, was zu einer beschleunigten Entscheidung führen kann. Gestaltungselemente wie Farben, Bilder, Texte oder der Tonfall digitaler Inhalte tragen wesentlich dazu bei, Emotionen wie Vertrauen, Sicherheit oder Dringlichkeit zu erzeugen und somit das Verhalten zu beeinflussen (Swoboda & Schramm-Klein, 2025, S. 49, 174).

In diesen komplexen und dynamischen Umgebungen greifen Konsument:innen häufig auf sogenannte Heuristiken zurück. Diese vereinfachten Entscheidungsregeln helfen dabei, trotz Unsicherheit oder Zeitdruck zu einer Entscheidung zu gelangen. Besonders verbreitet sind die Ankerheuristik, bei der die erste erhaltene Information eine starke Wirkung auf die Entscheidung hat, sowie die Verfügbarkeitsheuristik, bei der leicht erinnerbare Informationen überbewertet werden. Im digitalen Raum äußern sich diese Muster beispielsweise in Form von Preisankern, Nutzerbewertungen oder Empfehlungen, die zur Orientierung dienen und den Auswahlprozess erleichtern (Hoffmann & Akbar, 2024, S. 117-119).

Insgesamt ist das digitale Konsumentenverhalten durch eine hohe Komplexität, Flexibilität und Relevanz geprägt. Die Beschaffung von Informationen über digitale Kanäle hat mittlerweile einen größeren Stellenwert als klassische Massenmedien. Unternehmen reagieren auf diese Veränderungen, indem sie ihre Angebote über mehrere Kanäle hinweg zugänglich machen. Ziel ist es, Konsument:innen entlang ihrer *Customer Journey* mit relevanten Informationen zu versorgen und dadurch möglichst viele Kontaktpunkte im Kaufentscheidungsprozess zu schaffen (Swoboda & Schramm-Klein, 2025, S. 92-94).

2.2 Involvement als moderierende Variable

Das Konzept des Involvements beschreibt im Wesentlichen das Maß an innerer Beteiligung, mit dem sich Konsument:innen einem Produkt, einer Entscheidung oder einer Situation widmen. Dabei spielt insbesondere die persönliche Relevanz eine zentrale Rolle. Celsi und Olson (1988) definieren „erlebtes Involvement" als subjektives Gesamtgefühl der persönlichen Bedeutung. Dieses beeinflusst nicht nur kognitive Prozesse wie Aufmerksamkeit und Verständnis, sondern auch beobachtbare Verhaltensweisen wie Informationssuche und Kaufhandlungen (Celsi & Olson, 1988, S. 210-211). Auch Swoboda und Schramm-Klein (2025) betonen, dass Involvement aus kognitiven, emotionalen und reaktiven Komponenten bestehen kann. Das kognitive Involvement fokussiert sich stärker auf die intensive Informationsaufnahme und -verarbeitung und das emotionale Involvement hingegen auf emotionale Reaktionen (Swoboda & Schramm-Klein, 2025, S. 198-199). Eine differenziertere Messung des Involvements schlägt das Modell von Laurent und Kapferer (1985) vor, welches fünf Einflussfaktoren berücksichtigt: Interesse am Produkt, Freude beim Kauf, wahrgenommener Nutzen, Risiko eines Fehlkaufs und die Kosten einer Fehlentscheidung. Dabei zeigen Konsument:innen nicht immer konsistent hohe oder niedrige Involvement-Werte. Diese können je nach Produkt oder Kontext stärker oder schwächer variieren (Laurent & Kapferer, 1985, S. 42-48).

In der Forschung hat sich zudem die grundlegende Unterscheidung zwischen Low- und High-Involvement etabliert. Bei High-Involvement handelt es sich um Situationen, in denen Konsument:innen eine hohe persönliche Relevanz oder wahrgenommenes Risiko empfinden, was mit intensiver Informationssuche und systematischer Bewertung einhergeht. Dabei halten sie höhere Ansprüche an die Qualität der Informationen und an die Glaubwürdigkeit sowie Aktualität der Quellen. Das Ausmaß der kognitiven Aktivitäten ist groß und grobe Änderungen in der Sichtweise oder Wahrnehmung aufwendiger und schwieriger. Bei Low-Involvement hingegen ist die persönliche Relevanz geringer und Konsument:innen zeigen ein weniger starkes kognitives Engagement und greifen daher eher auf vereinfachte Entscheidungsstrategien zurück. Das wahrgenommene Risiko bei einer Fehlentscheidung wird nicht zu hoch bewertet und so sind Kriterien wie Qualität, Aktualität und Glaubwürdigkeit nicht Mittelpunkt oder ausschlaggebend ihrer Entscheidung. Eine bekannte Entscheidungsstrategie davon ist die Low-Involvement-Lernhierarchie, die besagt, dass Kaufentscheidungen durch geringe Aufmerksamkeit, eine affektive Haltung und eine vordefinierte Einstellung geprägt sind, die wiederum

durch repetitive Werbebotschaften eines Produkts entstehen (Moorthy et al., 1997, S. 263-265; Swoboda & Schramm-Klein, 2025, S. 138-140).

Die wichtigsten Faktoren bei der Unterscheidung von Low- und High-Involvement-Situationen sind das wahrgenommene Kaufrisiko, das Ausmaß der benötigten Informationen sowie deren Qualität zur Entscheidungsfindung (Swoboda & Schramm-Klein, 2025, S. 92). Klassische Beispiele für diese beiden Kategorien sind daher:

- High Involvement – Autokauf: Der Kauf eines Autos erfordert eine sorgfältige Recherche zu Produkteigenschaften, Vergleichskriterien und möglichen Alternativen. Die Entscheidung ist durch einen hohen finanziellen Einsatz, eine langfristige Nutzung sowie potenzielle Risiken im Alltag geprägt, was ein intensives Informationsbedürfnis zur Folge hat.
- Low Involvement – Kauf von Haushaltsreinigern: Der Kauf von Reinigungsmitteln erfolgt häufig routinemäßig und basiert auf Gewohnheiten oder Markenvertrautheit, die durch wiederholte Werbekontakte entsteht. Das geringe wahrgenommene Risiko bei einem Fehlkauf führt zu einer eher oberflächlichen Auseinandersetzung mit der Entscheidung (Swoboda & Schramm-Klein, 2025, S. 139-140).

2.3 Suchverhalten bei Nutzung klassischer Suchmaschinen (Google)

Suchmaschinen, allen voran Google, sind fest im Alltag der Konsument:innen verankert und gelten heute als unverzichtbares Werkzeug zur digitalen Informationsbeschaffung. Die verbreitete Wahrnehmung lautet: Nahezu jede Information lässt sich über Google schnell und unkompliziert finden. Diese Selbstverständlichkeit zeigt sich auch im alltäglichen Sprachgebrauch, in dem das Verb „googeln" längst etabliert ist und sogar Eingang in den Duden gefunden hat. Laut Eberspächer und Holtel (2007) hat sich diese Form der Informationssuche zu einer regelrechten „Kulturtechnik" entwickelt, die das Informationsverhalten der Menschen nachhaltig prägt (Eberspächer & Holtel, 2007, S. 117, 135).

Der Ablauf einer typischen Google-Suche ist für die meisten Nutzer:innen bereits intuitiv. Nach der Eingabe weniger Keywords werden innerhalb von Millisekunden zahlreiche Ergebnisse angezeigt. Diese bestehen aus einer Liste anklickbarer Links zu Webseiten, die inhaltlich zu den Suchbegriffen passen. Die Ergebnisanzeige ist nach Relevanz sortiert, wobei auch bezahlte Anzeigen einen sichtbaren Platz einnehmen (Eberspächer & Holtel, 2007, S. 135). In der Praxis beschränken sich Nutzer:innen

häufig auf die erste Ergebnisseite und klicken bevorzugt auf die obersten Einträge. Dabei handelt es sich meist um Quellen, die entweder algorithmisch als besonders relevant bewertet wurden oder durch bezahlte Werbung hervorgehoben sind (*EN Datos - Consumer Search Behavior in the Age of AI*, o. J., S. 6; *Online Search After ChatGPT*, o. J.; S. 5). In ihrer Entscheidungsfindung verlassen sich Nutzer:innen dabei nicht nur auf den Algorithmus, sondern auch auf bekannte Marken oder auffällige Titel, denn das vorrangige Ziel besteht darin, schnell zu einer passenden Information zu gelangen. Dieses Bedürfnis nach Effizienz wird von Google durch ein vertrautes Interface und eine klar strukturierte Ergebnisdarstellung bedient, auch wenn dies gleichzeitig als Einschränkung der eigenen Informationskontrolle empfunden werden kann (*EN Datos - Consumer Search Behavior in the Age of AI*, o. J., S. 7).

Ein zentrales Merkmal des Suchverhaltens bei Google ist die Nutzung einfacher Heuristiken, die Konsument:innen dabei helfen, trotz der Fülle an Informationen zügig Entscheidungen zu treffen (Hoffmann & Akbar, 2024, S. 117). Besonders augenfällig ist der sogenannte *Position Bias*: Die höchsten Klickraten entfallen auf das erste Ergebnis, unabhängig vom tatsächlichen Informationsgehalt. Dieser Effekt lässt sich dadurch erklären, dass viele Nutzer:innen der Sortierung durch die Suchmaschine vertrauen, ohne die Auswahlmechanismen zu hinterfragen. Diese Art der Informationssuche erfolgt oberflächlich, wird jedoch in vielen Fällen als ausreichend empfunden, vor allem bei niedrig involvierenden Aufgabenstellungen (Jansen & Spink, 2006, S. 15).

Das Vertrauen in Google ist dabei tief verankert, gestützt auf die jahrelange Nutzung, alltägliche Routine und die als zuverlässig wahrgenommenen Ergebnisse. Nicht selten wird Google als objektive, neutrale Instanz wahrgenommen, obwohl die genauen Kriterien der Ergebnisreihung nicht transparent sind und wirtschaftliche Interessen eine Rolle spielen können (Eberspächer & Holtel, 2007, S. 119-120). Xu et al. (2023) zeigen, dass viele Nutzer:innen den angezeigten Ergebnissen grundsätzlich Vertrauen schenken, dabei jedoch kaum Motivation verspüren, deren Quelle oder Richtigkeit kritisch zu hinterfragen. Gleichzeitig weist ihre Studie darauf hin, dass die Suche über Google einen gewissen Aufwand erfordert. Denn um relevante Informationen zu finden, müssen mehrere Ergebnisse verglichen und manuell bewertet werden und dieser zusätzliche Aufwand kann die wahrgenommene Informationsqualität mindern (Xu et al., 2023, S. 6).

2.4 Suchverhalten bei Nutzung von LLMs (wie ChatGPT)

Mit dem Aufkommen dialogbasierter Systeme wie ChatGPT hat sich die Art und Weise, wie Konsument:innen online nach Informationen suchen, grundlegend gewandelt. Statt wie bei Google Schlagwörter einzugeben, formulieren Nutzer:innen nun vollständige Fragen, Sätze oder Aufgabenstellungen, sogenannte *Prompts*, in natürlicher Sprache. Auf diese erhalten sie eine ausführlich formulierte Antwort, die in einem verständlichen Fließtext dargestellt wird. Dieser Wandel kann als dialogorientierter Suchprozess beschrieben werden, der es erlaubt, Rückfragen zu stellen und individuelle, kontextbezogene Informationen zu erhalten, anstelle einer rein listenbasierten Ergebnisdarstellung (Xu et al., 2023, S. 19).

Die wachsende Bedeutung dieser Technologie zeigt sich auch in aktuellen Nutzungszahlen. Eine Analyse von Online Search After ChatGPT in Kooperation mit Statista verdeutlicht, dass bereits im Jahr 2023 rund 13 Millionen Erwachsene in den Vereinigten Staaten generative KI vorrangig zur Online-Recherche eingesetzt haben. Für das Jahr 2027 wird ein Anstieg auf etwa 90 Millionen prognostiziert (*Online Search After ChatGPT*, o. J., S. 15). So zeigt auch eine weitere Studie der Statista vom April 2023, dass generative künstliche Intelligenz von 11% der weltweit Befragten, bereits als Suchmaschinen-Ersatz verwendet wird (*Generative KI - Verwendungszweck 2023*, o. J.).

Wie stark sich LLMs im Alltag etablieren, zeigt auch die Studie von Wolf und Maier (2024). Dort gaben 13,3 Prozent der Befragten an, ChatGPT täglich im privaten Kontext zu nutzen. Im beruflichen Umfeld liegt dieser Anteil sogar bei 46,67 Prozent. Als Hauptgründe für die Nutzung nennen die Befragten insbesondere die wahrgenommene Nützlichkeit, die einfache Bedienbarkeit sowie die inspirierende Wirkung der KI-basierten Antworten (Wolf & Maier, 2024, S. 6-12).

Ein zentraler Unterschied zwischen der Suchlogik von ChatGPT und klassischen Suchmaschinen zeigt sich in der Art, wie Informationen verarbeitet und präsentiert werden. Das zugrunde liegende Modell bezieht bei jeder neuen Eingabe durch die Nutzer:innen den gesamten bisherigen Gesprächsverlauf mit ein, um konsistente und kontextbezogene Antworten zu erzeugen (Capra & Arguello, 2023, S. 2). Dies ermöglicht es dem System, auch komplexere Suchanfragen oder Nachfragen im Laufe des Dialogs der Such- oder Informationsanfrage konsistent zu beantworten. Gleichzeitig unterscheidet sich die Eingabestruktur deutlich: In einer Vergleichsstudie beobachteten Xu et al. (2023), dass Nutzer:innen bei ChatGPT wesentlich längere und

vollständiger formulierte Prompts verwendeten konnten als bei Google, was auf ein stärker gesprächsorientiertes Suchverhalten hinweist (Xu et al., 2023, S. 18).

Auch die Darstellung der Ergebnisse hebt sich deutlich ab. Statt einer Liste anklickbarer Links liefert ChatGPT direkt formulierte und strukturierte Antworten in Form eines Fließtexts zurück (Brinkmann et al., 2023, S. 2-3). Laut Xu et al. (2023) erhöht diese Form der Darstellung die wahrgenommene Informationsqualität, da die Antworten kohärent und leichter verständlich sind. Nutzer:innen berichten zudem, dass sie sich schneller zurechtfinden und seltener auf andere Quellen zurückgreifen müssen (Xu et al., 2023, S. 5, 19). Dieses Antwortformat scheint laut Ebersprächer und Holtel (2007) auch den mentalen Suchmustern vieler Menschen zu entsprechen, da Nutzer:innen abhängig je nach Situation unterschiedliche Formen der Präsentation von Informationen erwarten (Eberspächer & Holtel, 2007, S. 17-20).

Das Vertrauen in ChatGPT basiert nicht nur auf der sachlichen Qualität der Antworten, sondern auch auf der Art und Weise, wie das System kommuniziert (Kim et al., 2021, S. 1142). Huynh und Aichner (2025) beschreiben dieses Phänomen als „*human-like trust*". Dieses ist ein Vertrauensverhältnis, das sich aus der wahrgenommenen Kompetenz, Integrität und dem Wohlwollen gegenüber diesem System äußert (Huynh & Aichner, 2025, S. 6-7). Nutzer:innen erleben ChatGPT durch dessen dialogische Struktur und natürliche Sprache nicht nur als technisches Werkzeug, sondern zunehmend auch als interaktiven Gesprächspartner. Diese menschenähnliche Kommunikation kann allerdings dazu führen, dass Inhalte unkritisch übernommen werden. Spatharioti et al. (2023) zeigen in ihrer Studie, dass Nutzer:innen LLM-basierte Suchtools zwar als besonders hilfreich und angenehm empfinden, dabei jedoch eine Tendenz zum blinden Vertrauen des Ergebnisses beobachten. Insbesondere dann, wenn die bereitgestellten Informationen fehlerhaft oder unvollständig sind (Spatharioti et al., 2023, S. 2-3). Das Vertrauen in ChatGPT basiert offenbar nicht nur auf sachlichen Einschätzungen, sondern auch auf emotionalen Eindrücken, was die kritische Auseinandersetzung mit den erhaltenen Informationen abschwächen kann.

Ein wesentliches Potenzial von ChatGPT liegt in der natürlichen und barrierefreien Kommunikation. Nutzer:innen können ihre Informationsbedürfnisse direkt und ohne technische Hürden formulieren, was besonders im Vergleich zur starren Keyword-Eingabe klassischer Suchmaschinen als Erleichterung empfunden wird. Wie Werner et al. (2024) festhalten, erlaubt ChatGPT es den Nutzenden, „ihre Präferenzen direkt

auszudrücken – statt auf Keyword-basierte Empfehlungen oder Webanzeigen angewiesen zu sein" (Werner et al., 2024, S. 6). Dies reduziert nicht nur die mentale Belastung während des Suchprozesses, sondern führt auch dazu, dass Produkte, Dienstleistungen oder Informationen einfacher und zielgerichteter auffindbar sind (Werner et al., 2024, S. 6). Vor allem bei wenig komplexen oder offenen Suchanfragen empfinden viele Nutzer:innen diesen Zugang als angenehm und effizient (Xu et al., 2023, S. 5-6).

Gleichzeitig zeigt sich jedoch, dass die Qualität der gelieferten Antworten stark vom Input der Nutzer:innen abhängt. Ungenaue oder vage formulierte Prompts führen häufig zu allgemeinen, ungenauen oder sogar fehlerhaften Informationen. Dieser Zusammenhang ist nicht immer allen Nutzer:innen bewusst. Zudem besteht weiterhin Skepsis gegenüber der Verlässlichkeit der Antworten. Gerade in der Anfangsphase wurde ChatGPT vielfach für faktische Fehler und veraltete Informationen kritisiert, was dazu geführt hat, dass viele Konsument:innen dem Tool inzwischen mit mehr Zurückhaltung begegnen. Diese Skepsis wird verstärkt durch die bekannte Problematik von LLMs, bestimmte *Bias* (Verzerrung der Ergebnisse) zu reproduzieren oder Empfehlungen ohne nachvollziehbare Quelle zu geben (Xu et al., 2023, S. 19-20). Wie Werner et al. (2024) zeigen, können Systeme wie ChatGPT Nutzerpräferenzen erkennen und in der Folge personalisierte Kaufargumente generieren, auch ohne dass die Konsument:innen dies aktiv bemerken (Werner et al., 2024, S. 7). Dies wirft ethische Fragen auf und stellt ein zentrales Hindernis für vollumfängliches Vertrauen in KI-gestützte Suchsysteme dar.

2.5 Vergleichende Betrachtung: Google vs. ChatGPT

Die theoretischen Ausführungen der vorherigen Abschnitte zeigen, dass sich Google und ChatGPT in zentralen Dimensionen der Informationssuche deutlich unterscheiden. Diese Unterschiede betreffen insbesondere die Art der Ergebnisdarstellung, die Interaktion mit dem System, das Vertrauen in die erhaltenen Informationen sowie die wahrgenommene Qualität und Personalisierung der Antworten.

Google stellt den Nutzer:innen eine Auswahl an Quellen zur Verfügung, die individuell geprüft, verglichen und bewertet werden müssen. Diese Suchlogik erfordert mehr Eigenleistung, bietet jedoch gleichzeitig Kontrolle und Transparenz über die Herkunft der Information. Nutzer:innen sind aktiver in die Bewertung involviert und neigen dazu, kritisch mit den Suchergebnissen umzugehen. In Studien wurde festgestellt, dass sich dadurch ein hohes Maß an Vertrauen etablieren kann, nicht zuletzt, weil die

Auswahl der Quellen als nachvollziehbar wahrgenommen wird (*Online Search After ChatGPT*, o. J., S. 74; Xu et al., 2023, S. 19).

ChatGPT hingegen basiert auf einer dialogbasierten Struktur und liefert vollständige, natürlich formulierte Antworten, ohne dass weitere Klicks notwendig sind. Dies senkt die kognitive Belastung und kann zu einer höheren Effizienz bei der Informationsbeschaffung führen, insbesondere bei einfachen oder eher offenen Aufgaben (Werner et al., 2024, S. 6; Xu et al., 2023, S. 19). Nutzer:innen empfinden die Interaktion häufig als intuitiver und angenehmer, was sich auch in einer insgesamt positiveren Bewertung der Nutzererfahrung niederschlägt (Wolf & Maier, 2024, S. 8-12). Gleichzeitig zeigen Studien, dass diese Bequemlichkeit zu einer geringeren kritischen Auseinandersetzung mit den gelieferten Informationen führen kann. So berichten Xu et al. (2023), dass viele Nutzer:innen die Inhalte von ChatGPT weitgehend ungeprüft übernehmen. Ein Effekt, der vor allem durch das vertrauenswürdige Auftreten der Antwortform begünstigt wird (Xu et al., 2023, S. 20).

Ein weiterer Aspekt betrifft die wahrgenommene Entscheidungssicherheit. Während Google-Nutzer:innen mehrere Quellen abwägen und dadurch ein differenzierteres Bild erhalten, vermittelt ChatGPT oft eine klare, zusammengefasste Antwort, die subjektiv als handlungsleitend empfunden wird. Diese direkte Form der Rückmeldung kann das Gefühl stärken, eine fundierte Entscheidung treffen zu können, auch wenn objektiv weniger überprüfbare Informationen vorliegen (Spatharioti et al., 2023, S. 2-3; Xu et al., 2023, S. 18-19).

Besonders auffällig sind die Unterschiede bei der wahrgenommenen Personalisierung. ChatGPT geht im Verlauf des Gesprächs auf Nutzerpräferenzen ein, greift den Kontext früherer Eingaben auf und formuliert Antworten individueller als klassische Suchmaschinen. Gkikas und Theodoridis (2022) argumentieren, dass LLMs aufgrund ihrer adaptiven Antwortstruktur eine personalisierte Konsumentenerfahrung ermöglichen, was sie zu besonders attraktiven Instrumenten im digitalen Marketing macht (Gkikas & Theodoridis, 2022, S. 165-166). Werner et al. (2024) beschreiben diesen Effekt als „barrierefreie Kommunikation", bei der Konsument:innen ihre Präferenzen direkt äußern können, ohne auf strukturierte Suchbegriffe oder vorgefertigte Links angewiesen zu sein (Werner et al., 2024, S. 6). Google hingegen bietet in dieser Hinsicht weniger Interaktionsspielraum und vermittelt tendenziell neutralere, standardisierte Ergebnisse (Xu et al., 2023, S. 19-20).

Ein weiterer entscheidender Einflussfaktor ist das persönliche Involvement der Nutzer:innen. Wie bereits in Kapitel 2.2 beschrieben, beeinflusst der Grad an Involvement, also das persönliche Interesse und wahrgenommene Risiko, maßgeblich die Art der Informationsverarbeitung. In High-Involvement-Situationen prüfen Nutzer:innen Inhalte systematisch und gründlich, was der transparenten Quellenauswahl von Google entgegenkommt. In Low-Involvement-Situationen hingegen stehen Effizienz, Einfachheit und intuitive Bedienung im Vordergrund, Aspekte, in denen ChatGPT wahrgenommen Vorteile aufweist (Celsi & Olson, 1988, S. 210-211; Swoboda & Schramm-Klein, 2025, S. 198-199).

Diese theoretische Gegenüberstellung bildet die Grundlage für die im nächsten Kapitel entwickelten Hypothesen, mit denen die wahrgenommenen Unterschiede zwischen Google und ChatGPT, unter Berücksichtigung des Involvements, empirisch überprüft werden.

3 Methodik

Das methodische Vorgehen dieser Bachelorarbeit basiert auf einem Feldexperiment in Form einer Onlinebefragung. Ziel war es, Unterschiede in der Wahrnehmung und Bewertung zweier Suchmethoden, der klassischen Suchmaschine Google und dem KI-basierten System ChatGPT, unter Berücksichtigung des Involvement-Levels (Low vs. High) zu erfassen. Das Kapitel umfasst die Beschreibung der Zielsetzung, die Hypothesenentwicklung auf Basis bestehender Literatur, das Untersuchungsdesign, die eingesetzten Messinstrumente sowie die Durchführung und Zusammensetzung der Stichprobe.

3.1 Ziel der empirischen Untersuchung

Ziel der empirischen Untersuchung war es, Erkenntnisse über das aktuelle Nutzerverhalten und die subjektive Bewertung von Suchsystemen zu gewinnen, konkret am Beispiel von Google und ChatGPT. Im Fokus standen Unterschiede in Nutzererfahrung, Vertrauen, wahrgenommener Informationsqualität, Entscheidungssicherheit und wahrgenommener Personalisierung, abhängig von der Involvement-Stufe einer Aufgabe. Die Ergebnisse sollen zeigen, wie Konsument:innen digitale Informationssuche heute wahrnehmen und bewerten, insbesondere im Spannungsfeld zwischen klassischen und KI-basierten Systemen.

3.2 Hypothesenentwicklung

Die Hypothesen dieser Arbeit basieren auf den theoretischen und empirischen Grundlagen, die in Kapitel 2 umfassend dargestellt wurden. Sie ergeben sich aus dem Zusammenspiel konsumentenpsychologischer Ansätze (Kapitel 2.1), der Involvement-Theorie (Kapitel 2.2), sowie spezifischer Merkmale des Suchverhaltens bei Google (Kapitel 2.3) und ChatGPT (Kapitel 2.4). Besonders in Kapitel 2.5 wurden zentrale Unterschiede zwischen beiden Systemen, etwa hinsichtlich Suchlogik, wahrgenommener Qualität, Vertrauen, Entscheidungserleben und Personalisierung, gegenübergestellt. In Kombination mit dem Einfluss des Involvements auf die Komplexität der Informationsverarbeitung ergibt sich ein differenziertes Bild, das als Grundlage für die Hypothesen dient (Celsi & Olson, 1988; Swoboda & Schramm-Klein, 2025). Auch aktuelle Studien zur Vertrauensbildung, Effizienz, Nutzungsakzeptanz und wahrgenommenen Qualität von KI-basierten Tools flossen in die Hypothesenformulierung ein (Huynh & Aichner, 2025; Werner et al., 2024; Wolf & Maier, 2024; Xu et al., 2023). Daraus ließen sich folgende Hypothesen ableiten:

Hypothese H1: Bei der Suche nach High-Involvement-Dienstleistungen bevorzugen Konsument:innen klassische Suchmaschinen wie Google gegenüber LLM-basierten Tools wie ChatGPT, da sie ihnen eine höhere Vertrauenswürdigkeit, Aktualität und Informationssicherheit zuschreiben.

Hypothese H2: Bei Low-Involvement-Suchaufgaben bevorzugen Konsument:innen ChatGPT gegenüber Google, da die Suche als effizienter und angenehmer empfunden wird.

Hypothese H3: Unabhängig vom Involvement-Level empfinden Konsument:innen die Nutzung von ChatGPT im Vergleich zur Google-Suche als hilfreicher bei der Entscheidungsfindung und zeigen eine höhere Bereitschaft, die Methode in Zukunft wieder zu verwenden oder weiterzuempfehlen.

Hypothese H4: In Informationssuch-Szenarien berichten Konsument:innen bei der Nutzung von ChatGPT von einer höheren subjektiven Entscheidungssicherheit als bei Google, da ChatGPT strukturierte und zugeschnittene Empfehlungen liefert.

Hypothese H5: Unabhängig vom Involvement-Level wird die Suche über ChatGPT von Konsument:innen als stärker personalisiert wahrgenommen als die Google-Suche, da sie ihre Anfrage freier formulieren können, die Ergebnisse individueller wirken und die Suchintention besser erkannt wird.

3.3 Studiendesign

Das Studiendesign dieser empirischen Arbeit basiert auf einem kontrollierten Onlineexperiment mit vier verschiedenen Versionen, die sich durch zwei Faktoren unterscheiden: die Art der Suchaufgabe (Low- vs. High-Involvement) und die verwendete Suchmethode (Google vs. ChatGPT). Jede Versuchsperson erhielt per Zufall eine dieser vier Versionen zugewiesen.

Ziel war es, zwei realitätsnahe Szenarien zu gestalten, die sich hinsichtlich ihres Involvements deutlich unterscheiden. Dabei sollten sich die Teilnehmenden intuitiv in die Entscheidungssituation hineinversetzen können.

Low-Involvement-Aufgabe: Die Teilnehmenden sollten ein passendes Café finden, das sich gut zum Arbeiten eignet, z. B. durch kostenloses WLAN, vorhandene Steckdosen und guten Kaffee. Die Entscheidung ist hierbei eher alltagsbezogen, wenig risikobehaftet und erfordert keine tiefgehende Recherche.

High-Involvement-Aufgabe: Die Teilnehmenden wurden gebeten, eine geeignete:n Allgemeinmediziner:in zu finden. Dabei spielten Kriterien wie gute Patientenbewertungen, lange Öffnungszeiten, gute Erreichbarkeit mit öffentlichen Verkehrsmitteln, angenehmes Klima und persönliche Betreuung eine Rolle. Diese Suche hat eine höhere persönliche Relevanz und ein größeres Informationsbedürfnis, ein klassisches Beispiel für ein High-Involvement-Szenario.

Die Suchaufgaben sollten mithilfe einer der beiden Plattformen gelöst werden:

- Google als klassische, weitverbreitete Suchmaschine
- ChatGPT als dialogbasiertes KI-System auf Basis eines Large Language Models (LLM)

Die Kombination aus den zwei Aufgaben (Low/High) und den zwei Plattformen (Google/ChatGPT) ergibt vier verschiedene Versionen des Fragebogens:

- Version 1: Low-Involvement x ChatGPT
- Version 2: Low-Involvement x Google
- Version 3: High-Involvement x ChatGPT
- Version 4: High-Involvement x Google

Die Zuteilung zu einer dieser vier Versionen erfolgte automatisiert über einen integrierten Zufallsgenerator in SoSci Survey. Dabei wurde ein rotierendes Verfahren gewählt (Ziehung ohne Zurücklegen), bei dem jede der vier Versionen nacheinander zugewiesen wurde, bevor der Zyklus erneut begann. Dadurch sollte eine gleichmäßige Verteilung der Stichprobe auf die vier Gruppen sichergestellt werden.

Die Versuchsanordnung verfolgt das Ziel, Unterschiede in der Nutzerbewertung von Suchergebnissen in realistischen, alltagsnahen Entscheidungssituationen zu untersuchen, sowohl abhängig von der Suchmethode als auch vom Involvement-Level. Durch die klar definierten, aber offen interpretierbaren Aufgaben konnten die Teilnehmenden ihre Sucherfahrung authentisch erleben und im Anschluss auf standardisierte Weise bewerten.

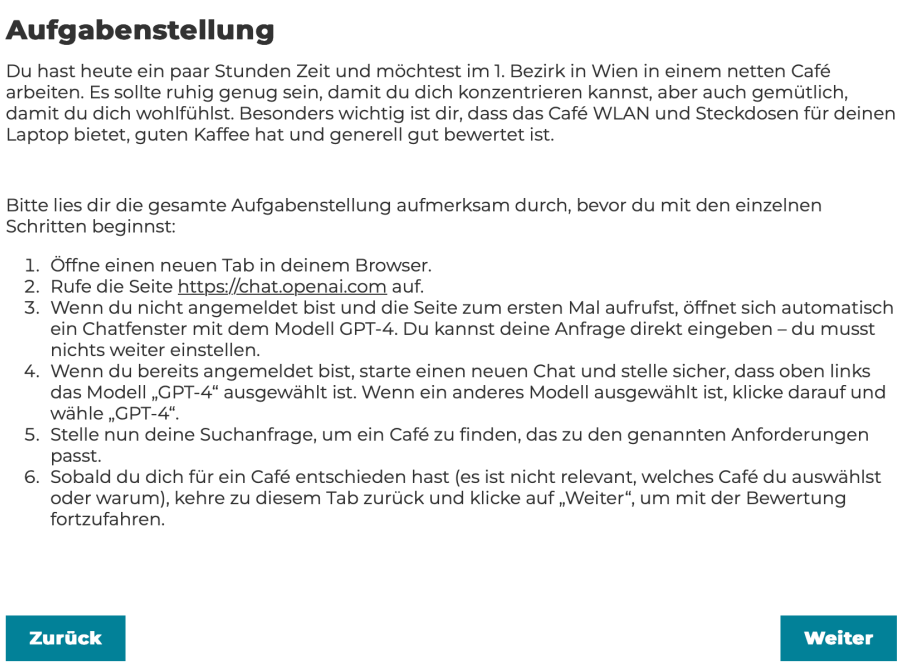

Abbildung 1: Beispielhafte Aufgabenbeschreibung – Version 1 (Low-Involvement x ChatGPT) Darstellung der Suchaufgabe für die Gruppe mit Low-Involvement-Szenario in Kombination mit der Suchmethode ChatGPT.

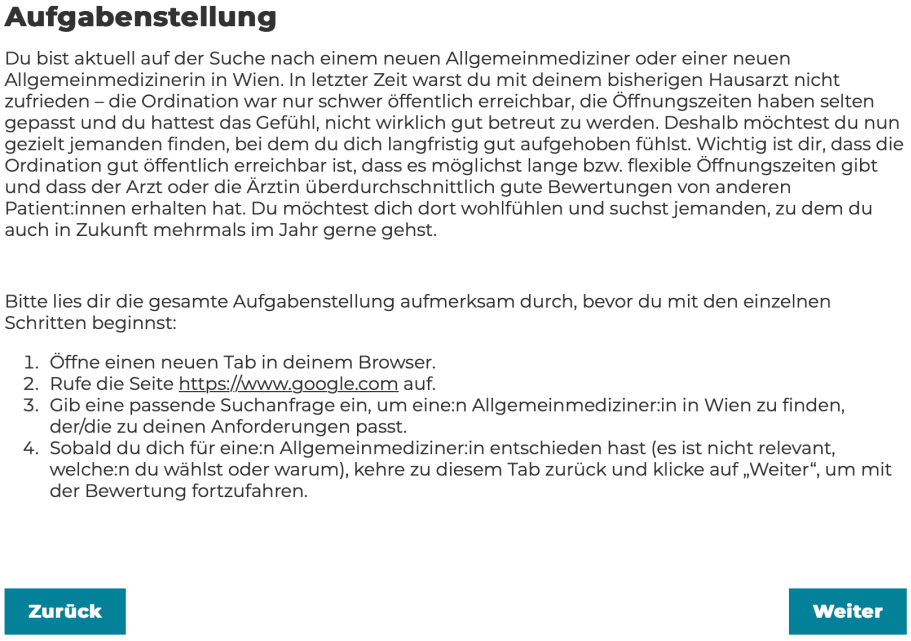

Abbildung 2: Beispielhafte Aufgabenbeschreibung – Version 4 (High-Involvement x Google) Darstellung der Suchaufgabe für die Gruppe mit High-Involvement-Szenario in Kombination mit der Suchmethode Google.

3.4 Messinstrumente

Zur Bewertung der wahrgenommenen Qualität der Suchmethoden wurden standardisierte Items eingesetzt, die sich auf zentrale Kriterien wie Nutzererfahrung, Informationsqualität, Entscheidungssicherheit und Personalisierung beziehen. Die Antworten erfolgten auf einer 7-stufigen Likert-Skala (1 = „stimme überhaupt nicht zu", 7 = „stimme voll und ganz zu"). Die Items wurden nach Durchführung der Suchaufgabe in drei aufeinanderfolgenden Bewertungsseiten präsentiert.

Die Zuordnung der Items zu Hypothesen erfolgt in Kapitel 4.2 (Hypothesenprüfung). Zur besseren Nachvollziehbarkeit ist hier bereits eine Übersicht über die verwendeten Rubriken und Itemformulierungen enthalten:

Tabelle 1: Rubriken und Bewertungskriterien des Fragebogens Übersicht über die drei Rubriken mit den zugehörigen Items, wie sie im Fragebogen zur Bewertung der Sucherfahrung eingesetzt wurden, sowie deren Item-Codes.

Nutzererfahrung & Suchprozess	Qualität der Informationen	Entscheidungs-sicherheit
EX01 – Die Nutzung der Suchmethode war angenehm.	**QU01 –** Ich hatte das Gefühl, umfassende Informationen erhalten zu haben.	**DE01 –** Ich habe das Gefühl, eine fundierte Entscheidung getroffen zu haben.
EX02 – Die Suchmethode war einfach zu verwenden.	**QU02 –** Die Suchergebnisse waren für meine Entscheidung hilfreich.	**DE02 –** Die Suchmethode hat mir die Entscheidung erleichtert.
EX03 – Ich konnte meine Suchanfrage so formulieren, wie ich wollte.	**QU03 –** Ich vertraue den gefundenen Informationen.	**DE03 –** Ich bin mir sicher, dass ich die beste Option gefunden habe.
EX04 – Die Suchergebnisse waren klar und verständlich.	**QU04 –** Ich hatte keinen Zweifel an der Richtigkeit der Ergebnisse.	**DE04 –** Ich habe genug Informationen erhalten, damit ich eine Entscheidung treffen konnte.
EX05 – Die Suche mit dieser Methode war leicht und unkompliziert.	**QU05 –** Ich habe die Informationen zur Entscheidungsfindung als aktuell wahrgenommen.	**DE05 –** Ich würde mich in einer ähnlichen Situation wieder auf diese Suchmethode verlassen.
EX06 – Die Suchergebnisse wirkten auf mich individuell und an meine Bedürfnisse angepasst.	**QU06 –** Die Quelle der Informationen war für mich nachvollziehbar.	**DE06 –** Ich würde anderen empfehlen, diese Suchmethode zu nutzen.
EX07 – Die Suchmethode hat meine Situation sowie Intention gut verstanden.		
EX08 – Ich hatte das Gefühl, schnell eine passende Entscheidung treffen zu können.		

Die statistische Auswertung der erhobenen Daten erfolgte mit der Software GraphPad Prism 10. Zur Überprüfung der Hypothesen wurden zwei zentrale Verfahren eingesetzt: Zum einen wurden unabhängige T-Tests auf Basis der aggregierten Mittelwerte durchgeführt, wobei eine Signifikanzschwelle von $p < .05$ zugrunde gelegt wurde. Zum anderen kamen Two-Way-ANOVAs zum Einsatz, um differenzierte Effekte

auf Itemebene zu untersuchen sowie Multiple-Comparisons-Tests nach Šídák durchzuführen.

Die Interpretation der Signifikanzniveaus orientierte sich an folgenden Schwellenwerten:

- $p < .05$ (*)
- $p < .01$ (**)
- $p < .001$ (***)
- $p < .0001$ (****)

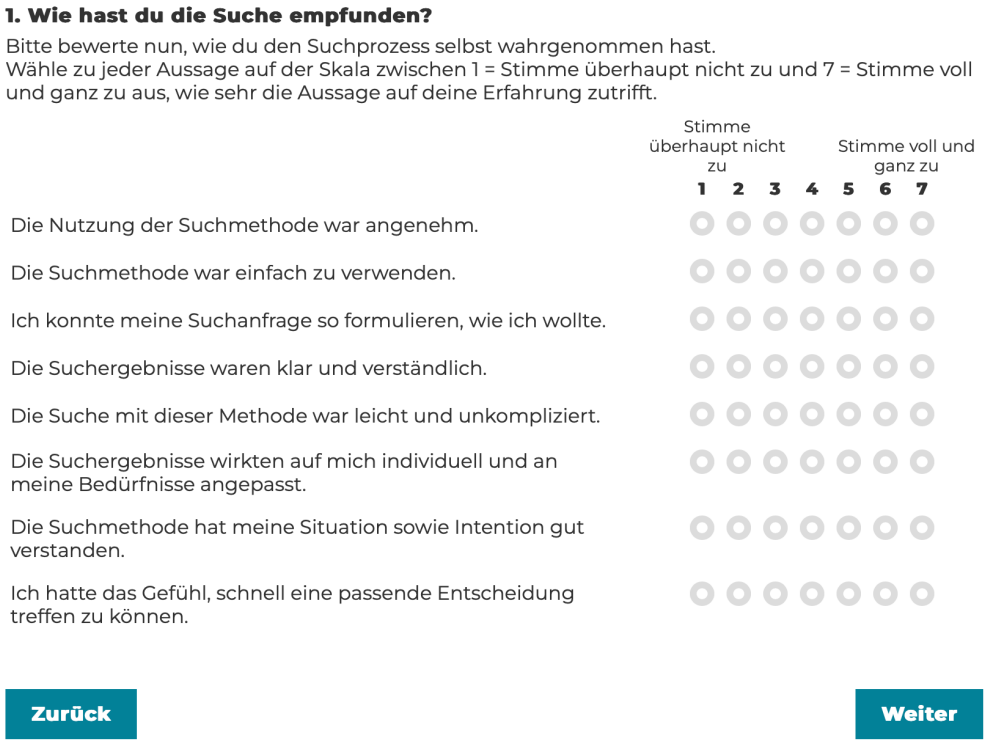

Abbildung 3: Bewertungsseite der Rubrik „Nutzererfahrung & Suchprozess" im Fragebogen Screenshot der Umfrageseite 3 zur Bewertung der ersten Rubrik nach Abschluss der gestellten Aufgabe.

3.5 Stichprobe & Durchführung

Die Umfrage wurde im Zeitraum vom 4. Mai bis zum 22. Mai 2025 über das Onlinebefragungstool SoSci Survey durchgeführt. Die Zielgruppe umfasste Personen, die über persönliche Kontakte sowie soziale Medien (Instagram, Facebook, WhatsApp) zur Teilnahme eingeladen wurden. Insgesamt klickten 795 Personen auf den Umfragelink, wovon 267 die Befragung begannen und 212 sie vollständig abschlossen.

Die Antwortquote beträgt damit 79,4 %. Der Mittelwert der Bearbeitungsdauer ohne Ausreißer lag bei 170,7 Sekunden (ca. 2 Minuten 51 Sekunden), die längste absolvierte Umfrage dauerte 488 Sekunden (ca. 8 Minuten 8 Sekunden).

Die Teilnehmenden wurden mittels Zufallsprinzips einer der vier Versionen des Fragebogens zugewiesen, wobei jede Version gleich häufig vergeben wurde. Die Zuteilung erfolgte über ein Ziehungssystem ohne Zurücklegen, um eine gleichmäßige Verteilung auf alle Gruppen sicherzustellen.

Die Umfrage wurde anonym durchgeführt. Es wurden keine personen- oder gerätebezogenen Daten gespeichert oder abgefragt. Die Teilnahmebedingungen (z. B. Deutschkenntnisse) wurden zu Beginn angegeben. Nach der Aufgabenseite bewerteten die Teilnehmenden ihre Suche auf mehreren Skalen. Ein kurzer Pretest mit Kommiliton:innen sowie Feedback der betreuenden Professorin stellte die Verständlichkeit und technische Funktionsfähigkeit des Fragebogens sicher.

4 Empirische Ergebnisse

Dieses Kapitel stellt die Ergebnisse der empirischen Untersuchung dar, welche das Ziel verfolgte, Unterschiede in der Wahrnehmung der beiden konkreten Suchmethoden Google und ChatGPT unter Berücksichtigung verschiedener Involvement-Stufen zu untersuchen. Zunächst werden die Rahmenbedingungen der Umfrage und ihre Durchführung besprochen, Kennzahlen der Stichprobe beschrieben und anschließend erfolgt die Hypothesenprüfung anhand von statistischen Tests, wie einer Two-Way-ANOVA und eines unabhängigen T-Test. Zur Visualisierung der Ergebnisse dieser Tests werden Boxplots mit Median beziehungsweise Mittelwert erstellt und hinsichtlich Signifikanz und Aussagekraft beschrieben.

4.1 Deskriptive Auswertung

Die folgende Übersicht zeigt die Verteilung der 212 abgeschlossenen Interviews auf die vier Aufgaben-Versionen:

- Version 1 (ChatGPT – Low-Involvement): 49 Teilnahmen
- Version 2 (Google – Low-Involvement): 56 Teilnahmen
- Version 3 (ChatGPT – High-Involvement): 44 Teilnahmen
- Version 4 (Google – High-Involvement): 63 Teilnahmen

Die leicht ungleichmäßige Verteilung lässt sich auf vorzeitige Abbrüche nach der zufälligen Zuteilung zu einer Aufgaben-Version zurückzuführen. Auffällig ist, dass die beiden ChatGPT-Aufgaben tendenziell etwas seltener durchgeführt und daher auch seltener vollständig abgeschlossen wurden.

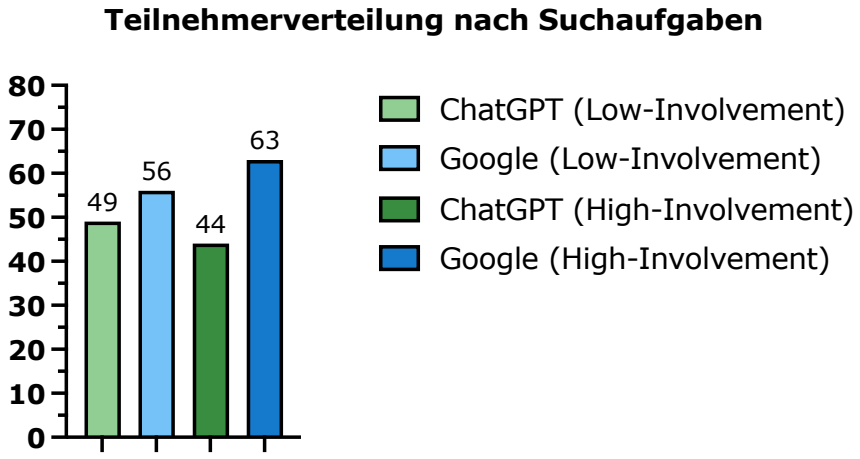

Abbildung 4: Verteilung der Teilnahmen auf die vier Aufgaben-Versionen
Veranschaulichung der Anzahl an vollständigen Teilnahmen je Version (Low/High-Involvement x Google/ChatGPT).

4.2 Hypothesenprüfung

Im folgenden Abschnitt werden die fünf in Kapitel 3.2 formulierten Hypothesen anhand der empirischen Ergebnisse überprüft. Für jede Hypothese wurde ein entsprechender T-Test auf Grundlage aggregierter Bewertungskriterien durchgeführt, um einen Gesamtunterschied zwischen den betrachteten Gruppen (z. B. ChatGPT vs. Google) festzustellen. Ergänzend wurde eine Two-Way-ANOVA durchgeführt, um die Unterschiede auf Item-Ebene zu validieren und die Aussagekraft einzelner Bewertungskriterien zu analysieren. Die Auswertungen der Hypothesentests wurden mit der Software GraphPad Prism 10 durchgeführt. In den Ergebnissen wird die Signifikanz standardisiert über Asterisk angegeben (* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$; **** $p < 0,0001$). Die grafischen Darstellungen in Form von Boxplots mit Median beziehungsweise Mittelwert unterstützen die Visualisierung der Ergebnisse.

Die im weiteren Verlauf analysierten Unterschiede zwischen den Bedingungen orientieren sich an den in Kapitel 2 dargestellten theoretischen Grundlagen und dienen der empirischen Überprüfung der daraus abgeleiteten Hypothesen.

4.2.1 Vertrauen & Informationsqualität bei High-Involvement-Suchen

In High-Involvement-Szenarien, wie etwa bei medizinischen oder finanziellen Entscheidungen, ist Vertrauen in die Informationsquelle von zentraler Bedeutung. Nutzer:innen in solchen Situationen legen besonderen Wert auf die Verlässlichkeit, Aktualität und Nachvollziehbarkeit der Informationen. Die zugrunde liegende Hypothese lautete daher:

Hypothese H1: Bei der Suche nach High-Involvement-Dienstleistungen bevorzugen Konsument:innen klassische Suchmaschinen wie Google gegenüber LLM-basierten Tools wie ChatGPT, da sie ihnen eine höhere Vertrauenswürdigkeit, Aktualität und Informationssicherheit zuschreiben.

Zur Überprüfung dieser Hypothese wurden die folgenden Bewertungskriterien herangezogen, welche inhaltlich das Thema Vertrauen, Informationsqualität und Aktualität abdecken:

- QU03 – Ich vertraue den gefundenen Informationen.
- QU04 – Ich hatte keinen Zweifel an der Richtigkeit der Ergebnisse.
- QU05 – Ich habe die Informationen zur Entscheidungsfindung als aktuell wahrgenommen.

- QU06 – Die Quelle der Informationen war für mich nachvollziehbar.

Ein T-Test der aggregierten Mittelwerte dieser vier Items ergab einen signifikanten Unterschied zwischen den Gruppen ChatGPT–High-Involvement und Google-High-Involvement ($p < 0.0001$). Die durchschnittliche Bewertung lag bei 4,26 Punkten für ChatGPT und 5,63 Punkten für Google, was einer Differenz von 1,37 Punkten (± 0,14) entspricht. Das 95 %-Konfidenzintervall reichte von 1,10 bis 1,65, mit einem eta²-Wert von 0,18, was auf eine starke Effektgröße hinweist. Dieses Ergebnis bestätigt deutlich, dass Google bei High-Involvement-Suchen als vertrauenswürdiger und qualitativ hochwertiger wahrgenommen wird als ChatGPT.

Zur weiteren Validierung wurde eine Two-Way-ANOVA durchgeführt. Dabei zeigte sich, dass der Faktor Suchmethode einen signifikanten Einfluss auf die Bewertung hatte ($p < 0.0001$, erklärter Varianzanteil: 18,33 %). Dies verdeutlicht, dass der Unterschied klar auf die Wahl der Plattform zurückzuführen ist und nicht auf bestimmte Items oder deren Wechselwirkung.

Auch die Multiple Comparison Analyse innerhalb der ANOVA bestätigt dieses Bild: Alle vier Bewertungskriterien zeigten signifikante Mittelwertunterschiede zwischen Google und ChatGPT. Die Items QU03, QU04 und QU05 wiesen einen hochsignifikanten Unterschied auf (****), QU06 hingegen war mit *** signifikant. Diese Ergebnisse unterstreichen, dass jedes einzelne Kriterium zur Bestätigung der Hypothese beiträgt und die Präferenz für Google in vertrauensbasierten Entscheidungssituationen statistisch fundiert ist.

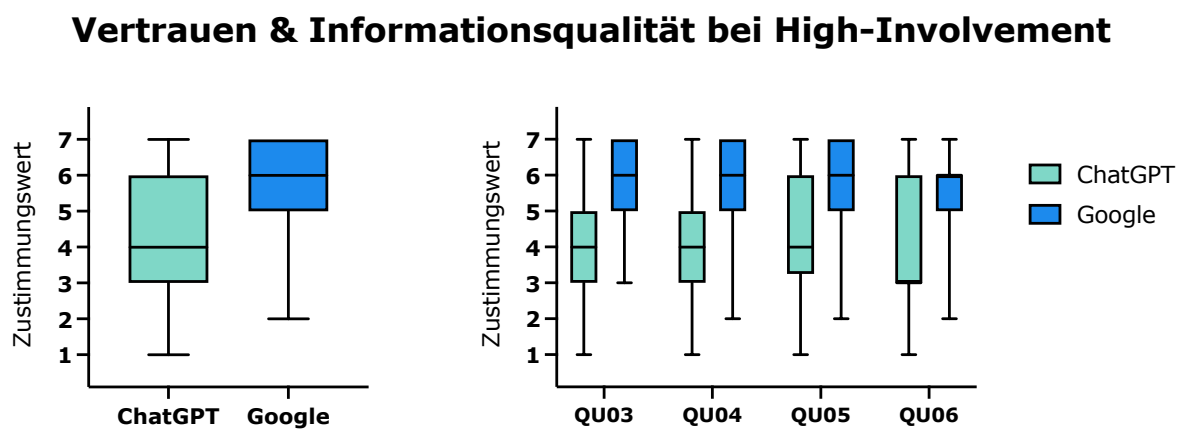

Abbildung 5: Ergebnisse Hypothese 1 – Vertrauen & Informationsqualität bei High-Involvement-Suchen
Links: T-Test der aggregierten Items (Boxplot, Median, Min, Max).
Rechts: Itemvergleich per Two-Way-ANOVA (Boxplots, Median, Min, Max).

4.2.2 Effizienz & Nutzungserlebnis bei Low-Involvement-Suchen

Low-Involvement-Suchaufgaben sind typischerweise durch ein geringes Risiko, wenig persönliche Relevanz und ein eher impulsives Entscheidungsverhalten geprägt. In solchen Fällen steht nicht die Informationsverlässlichkeit im Vordergrund, sondern vielmehr ein effizienter, unkomplizierter und angenehmer Suchprozess. Basierend auf diesen Überlegungen wurde folgende Hypothese formuliert:

Hypothese H2: Bei Low-Involvement-Suchaufgaben bevorzugen Konsument:innen ChatGPT gegenüber Google, da die Suche als effizienter und angenehmer empfunden wird.

Zur Überprüfung dieser Hypothese wurden vier Bewertungskriterien herangezogen, welche die wahrgenommene Einfachheit, Effizienz und Nutzerfreundlichkeit der Suchmethode abbilden:

- EX01 – Die Nutzung der Suchmethode war angenehm.
- EX02 – Die Suchmethode war einfach zu verwenden.
- EX05 – Die Suche war leicht und unkompliziert.
- EX08 – Ich hatte das Gefühl, schnell eine passende Entscheidung treffen zu können.

Ein T-Test der aggregierten Mittelwerte dieser vier Items zeigte einen signifikanten Unterschied zwischen ChatGPT-Low-Involvement und Google-Low-Involvement ($p < 0.0001$). Die durchschnittliche Bewertung lag bei 6,19 Punkten für ChatGPT und 5,30 Punkten für Google. Der Unterschied zwischen den Mittelwerten betrug 0,90 Punkte (± 0,13), mit einem 95 %-Konfidenzintervall von –1,16 bis –0,64. Der berechnete eta²-Wert von 0,099 weist auf eine mittlere Effektgröße hin. Diese Ergebnisse belegen klar, dass ChatGPT bei Low-Involvement-Suchen als angenehmer und effizienter empfunden wird als Google.

Die ergänzende Two-Way-ANOVA bestätigte ebenfalls die Signifikanz des Faktors Suchmethode ($p < 0.0001$, erklärter Varianzanteil: 9,95 %). Auch der Faktor Items war signifikant ($p < 0.0001$), während keine signifikante Interaktion festgestellt wurde ($p = 0,277$). Diese Ergebnisse legen nahe, dass der Unterschied in der Nutzerwahrnehmung auf die Plattformwahl zurückzuführen ist und nicht auf einzelne Items oder deren Interaktion.

Die Multiple Comparison Analyse der ANOVA zeigt zudem, dass drei der vier Bewertungskriterien signifikante Unterschiede zwischen den beiden Gruppen aufwiesen. Das Item EX01 war hochsignifikant (***), ebenso EX05 (****), während EX08 mit ** signifikant war. Lediglich EX02 war nicht signifikant. Insgesamt bestätigen diese Ergebnisse, dass ChatGPT bei Low-Involvement-Aufgaben tendenziell als benutzerfreundlicher wahrgenommen wird und die Hypothese damit weitgehend gestützt wird.

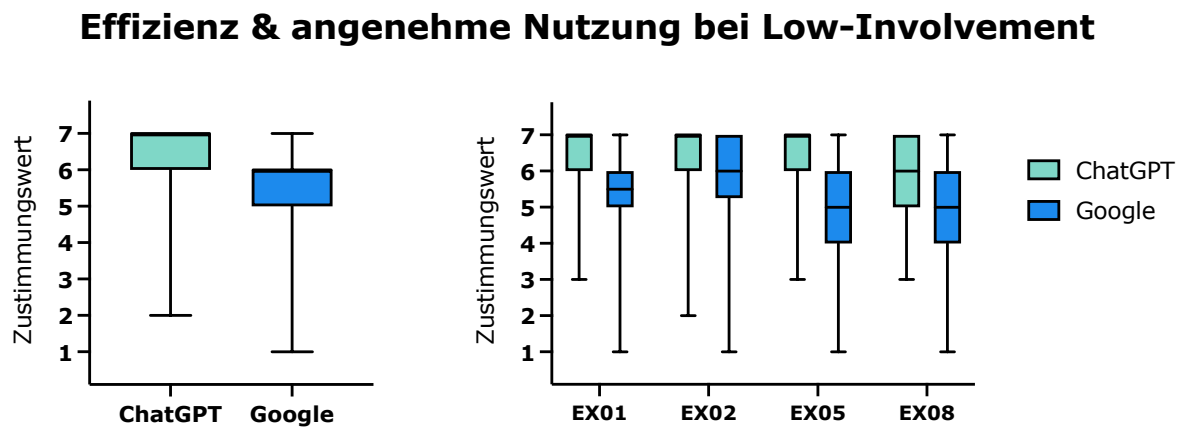

Abbildung 6: Ergebnisse Hypothese 2 – Effizienz & angenehme Nutzung bei Low-Involvement-Suchen
Links: T-Test der aggregierten Items (Boxplot, Median, Min, Max).
Rechts: Itemvergleich per Two-Way-ANOVA (Boxplots, Median, Min, Max).

4.2.3 Entscheidungsempfehlung & Wiederverwendbarkeit der Suchmethoden

Unabhängig vom Involvement-Level spielt die Frage eine Rolle, ob Nutzer:innen bereit wären, eine bestimmte Suchmethode erneut zu verwenden oder sie anderen weiterzuempfehlen. Diese Dimension kann als Indikator für die allgemeine Zufriedenheit mit der Nutzung sowie dem Vertrauen in die Methode gewertet werden. Daraus ergibt sich folgende Hypothese:

Hypothese H3: Unabhängig vom Involvement-Level empfinden Konsument:innen die Nutzung von ChatGPT im Vergleich zur Google-Suche als hilfreicher bei der Entscheidungsfindung und zeigen eine höhere Bereitschaft, die Methode in Zukunft wieder zu verwenden oder weiterzuempfehlen.

Zur Überprüfung dieser Hypothese wurden drei Bewertungskriterien ausgewählt, die die potenzielle Wiederverwendung und Weiterempfehlung der Suchmethode erfassen:

- DE02 – Die Suchmethode hat mir die Entscheidung erleichtert.
- DE05 – Ich würde mich in einer ähnlichen Situation wieder auf diese Methode verlassen.
- DE06 – Ich würde anderen empfehlen, diese Methode zu nutzen.

Die Ergebnisse des T-Tests beruhen auf einer Zusammenführung aller Bewertungen aus Low- und High-Involvement-Aufgaben, die jeweilige Entscheidungssituation wurde dabei nicht differenziert, sondern ausschließlich der genutzten Plattform (Google oder ChatGPT) betrachtet. Der T-Test zeigte keinen signifikanten Unterschied zwischen den beiden Plattformen ($p = 0{,}488$). Der Mittelwert bei ChatGPT lag bei 5,33 Punkten, jener bei Google bei 5,41 Punkten, eine minimale Differenz von 0,08 Punkten (± 0,12), mit einem 95 %-Konfidenzintervall von −0,15 bis 0,32. Der Effektgrößenwert (eta^2) betrug lediglich 0,0008 und ist somit vernachlässigbar.

Die Two-Way-ANOVA unterstreicht dieses Ergebnis. Auch hier wurden die Bewertungen beider Involvement-Stufen gemeinsam nach Suchmethode (Google vs. ChatGPT) analysiert. Der Faktor Suchmethode war nicht signifikant ($p = 0{,}347$), während die Items selbst ($p < 0.0001$) sowie die Interaktion zwischen Suchmethode und Item ($p = 0{,}014$) signifikant ausfielen. Die Unterschiede lassen sich demnach eher durch die Eigenheiten der Items oder deren Interaktion mit der Plattform erklären als durch die Plattform selbst.

Im Rahmen der Multiple Comparison Analyse wurde jedes der drei Items sowohl im Kontext der Low-Involvement- als auch der High-Involvement-Aufgabe analysiert. Deshalb erscheinen die Items DE02, DE05 und DE06 jeweils doppelt. Die Ergebnisse zeigen, dass fünf von sechs Vergleichen nicht signifikant waren. Lediglich DE05 (High-Involvement) zeigte einen signifikanten Unterschied in Richtung Google ($p = 0{,}038$, *). Alle anderen Vergleiche, etwa DE02 und DE06 für beide Involvement-Stufen, zeigten keine signifikanten Unterschiede zwischen Google und ChatGPT. Dieses Ergebnis lässt den Schluss zu, dass beide Plattformen hinsichtlich ihrer allgemeinen Akzeptanz, Wiederverwendung und Empfehlungsbereitschaft von Konsument:innen ähnlich bewertet werden.

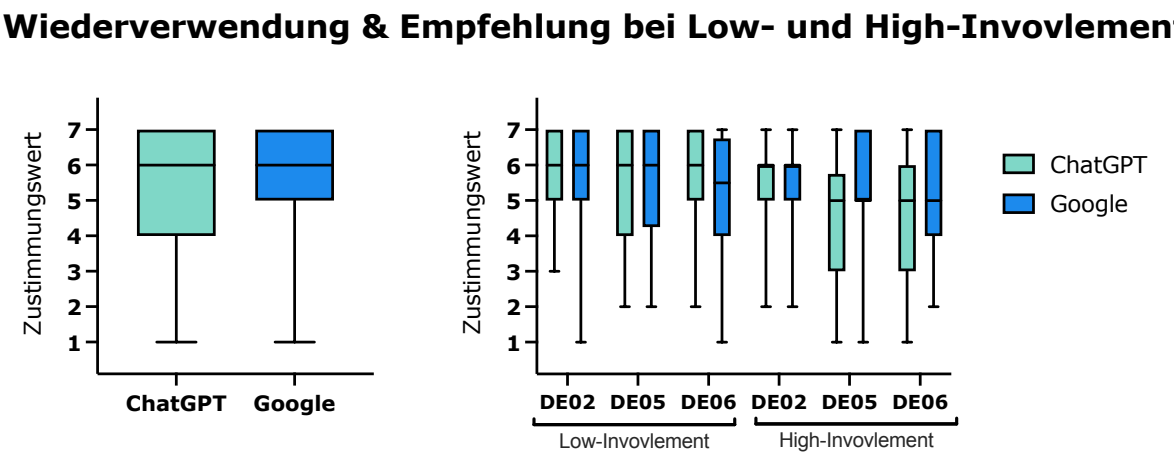

Abbildung 7: Ergebnisse Hypothese 3 – Unterstützung bei der Entscheidungsfindung & Wiederverwendung
Links: T-Test der aggregierten Items (Boxplot, Median, Min, Max).
Rechts: Itemvergleich per Two-Way-ANOVA (Boxplots, Median, Min, Max).

4.2.4 Subjektive Entscheidungssicherheit bei der Informationssuche

Die Hypothese zu Entscheidungssicherheit geht davon aus, dass ChatGPT-Nutzer:innen bei der Informationssuche ein höheres Maß an subjektiv empfundener Sicherheit erleben, da die Antworten oft als klar, vollständig und handlungsleitend wahrgenommen werden. Im Gegensatz dazu fordert die Nutzung von Google häufig eine aktivere Auseinandersetzung mit unterschiedlichen Quellen, was zwar Transparenz schafft, jedoch auch Unsicherheit erzeugen kann. Daraus ergab sich folgende Annahme:

Hypothese H4: In Informationssuch-Szenarien berichten Konsument:innen bei der Nutzung von ChatGPT von einer höheren subjektiven Entscheidungssicherheit als bei Google, da ChatGPT strukturierte und zugeschnittene Empfehlungen liefert.

Zur Überprüfung wurden die folgenden Bewertungskriterien ausgewählt, die zentrale Aspekte der Entscheidungsfindung abbilden:

- DE01 – Ich habe das Gefühl, eine fundierte Entscheidung getroffen zu haben.
- DE03 – Ich bin mir sicher, dass ich die beste Option gefunden habe.
- DE04 – Ich habe genug Informationen erhalten, damit ich eine Entscheidung treffen konnte.

Ein T-Test über die zusammengeführten Bewertungen aller Teilnehmer:innen, unabhängig von Involvement-Stufe, zeigt einen signifikanten Unterschied zwischen

ChatGPT und Google (p = 0,0083). Die durchschnittliche Bewertung bei ChatGPT lag bei 5,14 Punkten, während sie bei Google 4,84 Punkte betrug. Die Differenz der Mittelwerte beläuft sich auf 0,30 Punkte (± 0,11), mit einem 95 %-Konfidenzintervall von −0,53 bis −0,08. Der Effekt ist somit schwach, aber statistisch signifikant.

Zur vertieften Analyse wurde eine Two-Way-ANOVA durchgeführt. Diese bestätigt einen signifikanten Haupteffekt der Variable Suchmethode (p = 0,0127), was die Ergebnisse des T-Tests untermauert. Weder die Interaktion zwischen Items und Suchmethode (p = 0,923) noch die einzelnen Items als Haupteffekt (p < 0,0001) liefern jedoch starke Unterschiede. Das deutet darauf hin, dass die Plattformwahl unabhängig vom konkreten Fragetypus einen Einfluss auf die wahrgenommene Entscheidungssicherheit hat.

Die Multiple Comparison Analyse differenziert zusätzlich zwischen den Involvement-Stufen. Dabei wurden die drei Bewertungskriterien jeweils für Low- und High-Involvement getrennt betrachtet. Keine der sechs Einzelvergleiche (DE01, DE03, DE04 jeweils für Low- und High-Involvement) erreichte ein signifikantes Niveau, was durch große Konfidenzintervalle mit Überschneidung bestätigt wird. Dies weist darauf hin, dass der Unterschied der Mittelwerte im Gesamttest (T-Test) nicht auf einem einzelnen Bewertungskriterium basiert, sondern aus der aggregierten Wahrnehmung der Nutzer:innen resultiert.

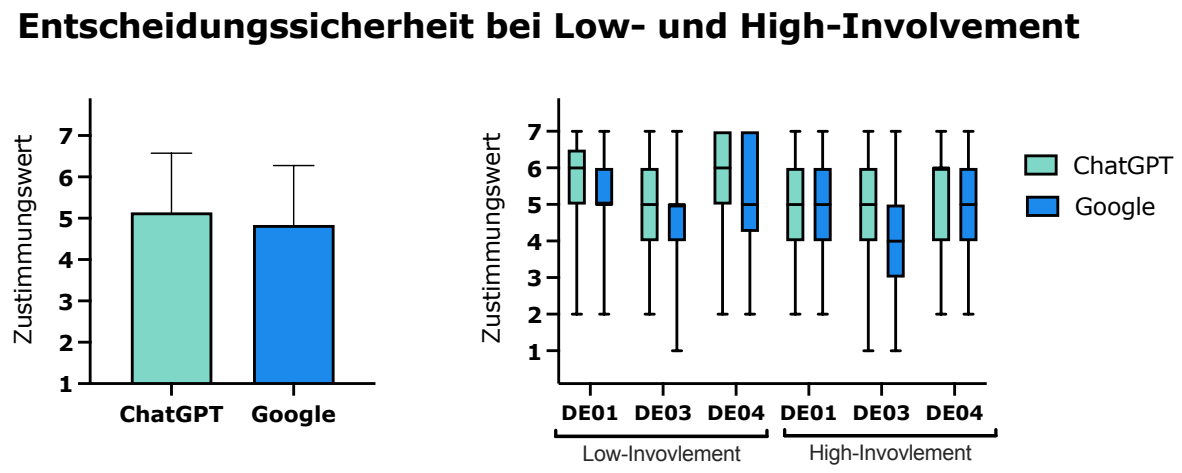

Abbildung 8: Ergebnisse Hypothese 4 – Entscheidungssicherheit bei der Nutzung von ChatGPT
Links: T-Test der aggregierten Items (Balkendiagramm, Mittelwert mit Standardabweichung).
Rechts: Itemvergleich per Two-Way-ANOVA (Boxplots, Median, Min, Max).

4.2.5 Wahrgenommene Personalisierung der Suchergebnisse

Personalisierung in digitalen Suchprozessen beschreibt das subjektive Empfinden, dass Informationen individuell auf die Nutzer:in zugeschnitten sind. ChatGPT erzeugt durch den dialogischen Charakter eine interaktive Suche, die viele Nutzer:innen als persönlicher wahrnehmen als das klassische Keyword-basierte Suchmodell von Google. Diese Wahrnehmung kann besonders bei offenen oder interpretationsbedürftigen Aufgaben verstärkt auftreten, unabhängig vom Involvement-Level. Die zugrunde liegende Hypothese lautete daher:

Hypothese H5: Unabhängig vom Involvement-Level wird die Suche über ChatGPT von Konsument:innen als stärker personalisiert wahrgenommen als die Google-Suche, da sie ihre Anfrage freier formulieren können, die Ergebnisse individueller wirken und die Suchintention besser erkannt wird.

Zur Überprüfung dieser Hypothese wurden die folgenden Bewertungskriterien verwendet, welche inhaltlich den Aspekt der wahrgenommenen Personalisierung und Nutzerzentrierung abbilden:

- EX03 – Ich konnte meine Suchanfrage so formulieren, wie ich wollte.
- EX06 – Die Suchergebnisse wirkten auf mich individuell und an meine Bedürfnisse angepasst.
- EX07 – Die Suchmethode hat meine Situation sowie Intention gut verstanden.

Ein T-Test der aggregierten Mittelwerte dieser drei Items ergab einen signifikanten Unterschied zwischen ChatGPT und Google ($p < 0.0001$). Die durchschnittliche Bewertung für ChatGPT lag bei 6,14 Punkten, während Google nur auf 4,07 Punkte kam. Die Differenz von 2,07 Punkten (± 0,14) bei einem 95 %-Konfidenzintervall von 1,80 bis 2,34 ist statistisch hoch signifikant ($eta^2 = 0,263$) und deutet auf einen starken Effekt hin. Damit wird deutlich, dass Nutzer:innen ChatGPT als deutlich personalisierter empfinden als Google.

Die Two-Way-ANOVA bestätigt diese Ergebnisse: Der Faktor Suchmethode erklärte 26,07 % der Gesamtvarianz und war hochsignifikant ($p < 0.0001$). Weder die Items ($p = 0.6809$) noch deren Interaktion ($p = 0.6809$) zeigten signifikante Effekte. Das deutet darauf hin, dass der Unterschied allein durch die Plattformwahl zustande kommt und konsistent über beide Involvement-Stufen hinweg beobachtet werden konnte, da High- und Low-Involvement-Aufgaben in dieser Analyse zusammengeführt wurden.

Besonders aufschlussreich war die Multiple Comparison Analyse, bei der jedes Bewertungskriterium getrennt nach Involvement-Level zwischen ChatGPT und Google verglichen wurde. Alle sechs getesteten Paare (jeweils drei pro Involvement-Stufe) zeigten hochsignifikante Unterschiede mit ****. So zeigten die Items EX03, EX06 und EX07 in jeder der beiden Versionen signifikante Mittelwertunterschiede (alle $p < 0.0001$). Diese Ergebnisse stützen die Hypothese eindeutig: ChatGPT wird über beide Aufgabenformate hinweg als personalisierter wahrgenommen als Google.

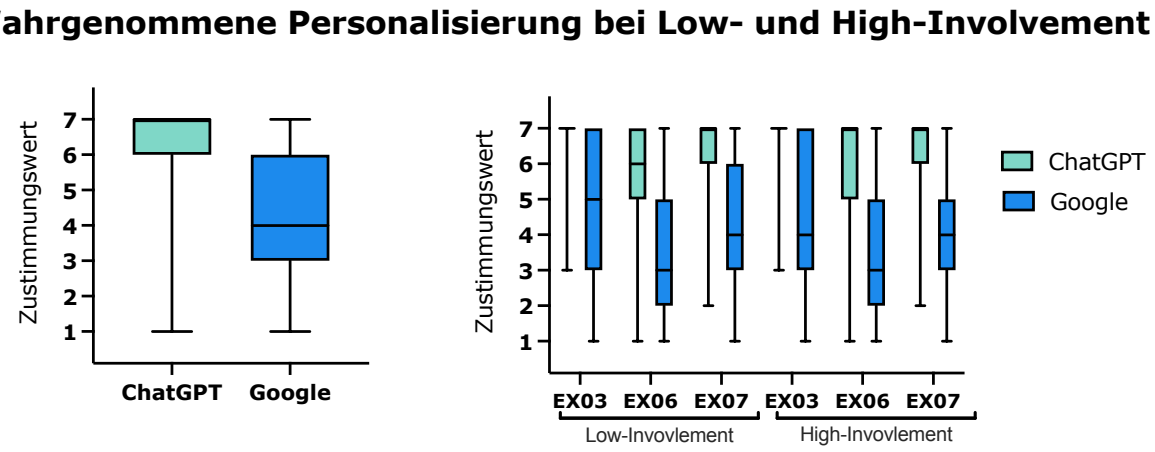

Abbildung 9: Ergebnisse Hypothese 5 – Wahrgenommene Personalisierung der Suchmethode Links: T-Test der aggregierten Items (Boxplot, Median). Rechts: Itemvergleich per Two-Way-ANOVA (Boxplots, Median, Min, Max).

5 Diskussion der Ergebnisse

Die Ergebnisse der empirischen Untersuchung lassen sich gut im Licht der theoretischen Grundlagen aus Kapitel 2 zuordnen. Besonders das Involvement-Modell (Celsi & Olson, 1988; Laurent & Kapferer, 1985) sowie bestehende Erkenntnisse zu digitalen Entscheidungssituationen (Hoffmann & Akbar, 2024; Swoboda & Schramm-Klein, 2025) liefern eine schlüssige Erklärung für die beobachteten Unterschiede zwischen Google und ChatGPT. Auch das spezifische Suchverhalten in digitalen Räumen, insbesondere in Bezug auf Heuristiken, Informationsverarbeitung und Systemvertrauen, wurde durch die Ergebnisse weitgehend bestätigt. Die theoretische Differenzierung zwischen High- und Low-Involvement-Situationen erwies sich als besonders nützlich, um unterschiedliche Bewertungsmuster der Suchmethoden zu erklären.

Die empirischen Ergebnisse bestätigen die theoretische Annahme, dass Konsument:innen je nach Involvement-Stufe unterschiedliche Anforderungen an Suchsysteme stellen (Celsi & Olson, 1988; Swoboda & Schramm-Klein, 2025).

Hypothese H1 wurde klar gestützt, da bei High-Involvement-Szenarien wie der Arztsuche Google hinsichtlich Vertrauens, Aktualität und Informationsnachvollziehbarkeit signifikant besser abschnitt als ChatGPT. Dieses Ergebnis steht im Einklang mit bisherigen Studien, die Google trotz algorithmischer Intransparenz eine höhere Glaubwürdigkeit zuschrieben (Xu et al., 2023).

Hypothese H2 konnte ebenfalls bestätigt werden. Bei Low-Involvement-Aufgaben, wie der Suche nach einem geeigneten Café, wurde ChatGPT als angenehmer und effizienter wahrgenommen. Dies stimmt mit der in der Literatur beschriebenen Tendenz überein, bei geringer persönlicher Relevanz auf einfache, intuitive Systeme zurückzugreifen (Capra & Arguello, 2023; Swoboda & Schramm-Klein, 2025).

Hypothese H3 wurde hingegen nicht signifikant bestätigt. Die Wiederverwendungsabsicht und Empfehlungsbereitschaft unterschieden sich zwischen Google und ChatGPT nicht statistisch. Dies könnte auf bestehende Gewohnheiten oder das fortbestehende Vertrauen in klassische Systeme zurückzuführen sein, trotz der wahrgenommenen Vorteile von ChatGPT im Nutzungserlebnis (Jansen & Spink, 2006).

Hypothese H4 wurde insgesamt bestätigt. Nutzer:innen gaben bei der Verwendung von ChatGPT ein höheres Maß an subjektiver Entscheidungssicherheit an als bei Google. Dies lässt sich auf die strukturierte Darstellung der Antworten und das direkte

Eingehen auf die Suchanfrage zurückführen, wodurch Entscheidungen leichter getroffen werden konnten (Spatharioti et al., 2023; Xu et al., 2023). Dennoch war der Effekt nur schwach ausgeprägt und auf Ebene der einzelnen Bewertungskriterien zeigte sich keine Signifikanz. Das Ergebnis weist somit auf einen konsistenten, aber begrenzten Effekt hin.

Die deutlichste Unterstützung erhielt Hypothese H5 zur wahrgenommenen Personalisierung: Über alle Involvement-Stufen hinweg wurde ChatGPT als individueller, verständnisvoller und besser an die eigene Intention angepasst wahrgenommen. Der gesprächsartiger Austausch und die freie Formulierung von Suchanfragen bieten offenbar eine personalisierte Nutzererfahrung, die sich deutlich von der klassischen Suchmaschinenstruktur abhebt (Werner et al., 2024; Xu et al., 2023).

Trotz der sorgfältigen Umsetzung weist diese Studie einige Einschränkungen auf. Die Stichprobe wurde nicht zufällig rekrutiert und bestand überwiegend aus digital affinen Personen, da die Umfrage auf elektronischen sozialen Medien verbreitet wurde, was die Übertragbarkeit auf die Gesamtbevölkerung einschränkt. Darüber hinaus wurde lediglich eine begrenzte Auswahl an Entscheidungsszenarien untersucht, nämlich die Suche nach einem Café sowie nach einer Allgemeinmediziner:in. Dadurch ist die Übertragbarkeit der Ergebnisse auf andere Produkte oder Dienstleistungen nur eingeschränkt möglich. Auch der experimentelle Aufbau mit nur einer einzigen Suchinteraktion pro Person bildet die realen, oft mehrstufigen Suchprozesse im Alltag nur begrenzt ab. Ein weiterer Aspekt betrifft die Nutzung von ChatGPT, bei der die Qualität der Antwort stark von der Formulierung des Prompts abhängt. Unterschiede im sprachlichen Ausdrucksvermögen der Teilnehmenden könnten somit die Bewertung der Plattform beeinflusst haben.

Die Ergebnisse zeigen jedoch, dass Large Language Models wie ChatGPT insbesondere bei einfachen und wenig bedeutsamen Informationssuchen als ernstzunehmende Alternative zu klassischen Suchmaschinen wahrgenommen werden. Für die Praxis im digitalen Marketing ergibt sich daraus die Empfehlung, Inhalte künftig nicht nur für Suchmaschinen wie Google, sondern auch für dialogbasierte Systeme zu optimieren. Unternehmen sollten prüfen, inwiefern der Einsatz solcher Systeme entlang der Customer Journey Mehrwert bringen kann. Auch aus wissenschaftlicher Sicht ergeben sich neue Perspektiven. Zukünftige Forschung sollte sich stärker mit den Kontextbedingungen und Nutzungsmotiven in Suchsituationen beschäftigen und

untersuchen, inwiefern Vertrauen, wahrgenommene Personalisierung und kognitive Entlastung das Nutzererlebnis prägen.

Zukünftige Studien könnten untersuchen, wie sich das Nutzerverhalten bei längeren und komplexeren Suchprozessen über mehrere Interaktionen hinweg entwickelt. Besonders relevant wäre zudem eine genauere Analyse der Prompt-Gestaltung. Dabei stellt sich die Frage, inwiefern die Form, Länge und Klarheit einer Eingabe die Qualität der generierten Antwort und die subjektive Bewertung beeinflusst. Auch die Weiterentwicklung von Chatbots und Suchsystemen auf Basis von Künstlicher Intelligenz eröffnet neue Forschungsfelder. Es bleibt offen, wie sich solche Systeme langfristig auf das Verhalten bei Informationsbeschaffungen von Konsument:innen auswirken und welche Rolle sie in zukünftigen Kaufentscheidungen einnehmen werden.

Die vorliegende Arbeit zeigt, dass Google und ChatGPT in zentralen Bereichen der digitalen Informationssuche unterschiedlich wahrgenommen werden. Gleichzeitig wird deutlich, dass das individuelle Involvement der Nutzer:innen maßgeblich beeinflusst, wie Suchsysteme bewertet und genutzt werden. Die empirischen Ergebnisse liefern somit einen differenzierten Beitrag zum Verständnis digitaler Entscheidungssituationen und machen deutlich, dass neue Technologien wie ChatGPT nicht nur technische Veränderungen darstellen, sondern auch zu neuen Erwartungen, Verhaltensmustern und Wahrnehmungen auf Seiten der Konsument:innen führen.

6 Literaturverzeichnis

Abrardi, L., Cambini, C., & Rondi, L. (2022). Artificial intelligence, firms and consumer behavior: A survey. *Journal of Economic Surveys*, *36*(4), 969–991. https://doi.org/10.1111/joes.12455

Brinkmann, A., Shraga, R., Der, R. C., & Bizer, C. (2023). *Product Information Extraction using ChatGPT* (No. arXiv:2306.14921). arXiv. https://doi.org/10.48550/arXiv.2306.14921

Capra, R., & Arguello, J. (2023). *How does AI chat change search behaviors?* (No. arXiv:2307.03826). arXiv. https://doi.org/10.48550/arXiv.2307.03826

Celsi, R. L., & Olson, J. C. (1988). The Role of Involvement in Attention and Comprehension Processes. *Journal of Consumer Research*, *15*(2), 210. https://doi.org/10.1086/209158

Eberspächer, J., & Holtel, S. (Hrsg.). (2007). Suchen und Finden im Internet. In *Suchen und Finden im Internet* (S. 138–155). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-38157-0_14

EN Datos—Consumer Search Behavior in the Age of AI. (o. J.).

Generative KI - Verwendungszweck 2023. (o. J.). Statista. Abgerufen 5. Juni 2025, von https://de.statista.com/statistik/daten/studie/1403840/umfrage/verwendungszweck-generativer-ki-tools/

Gkikas, D. C., & Theodoridis, P. K. (2022). AI in Consumer Behavior. In M. Virvou, G. A. Tsihrintzis, L. H. Tsoukalas, & L. C. Jain (Hrsg.), *Advances in Artificial Intelligence-based Technologies* (Bd. 22, S. 147–176). Springer International Publishing. https://doi.org/10.1007/978-3-030-80571-5_10

Hoffmann, S., & Akbar, P. (2024). *Konsumentenverhalten: Konsumenten verstehen – Marketingmaßnahmen gestalten*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-45648-1

Huynh, M.-T., & Aichner, T. (2025). In generative artificial intelligence we trust: Unpacking determinants and outcomes for cognitive trust. *AI & SOCIETY*. https://doi.org/10.1007/s00146-025-02378-8

Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management*, *42*(1), 248–263. https://doi.org/10.1016/j.ipm.2004.10.007

Kim, J., Giroux, M., & Lee, J. C. (2021). When do you trust AI? The effect of number presentation detail on consumer trust and acceptance of AI recommendations. *Psychology & Marketing*, *38*(7), 1140–1155. https://doi.org/10.1002/mar.21498

Laurent, G., & Kapferer, J.-N. (1985). Measuring Consumer Involvement Profiles. *JOURNAL OF MARKETING RESEARCH*.

Moorthy, S., Ratchford, B. T., & Talukdar, D. (1997). Consumer Information Search Revisited: Theory and Empirical Analysis. *Journal of Consumer Research*, *23*(4), 263. https://doi.org/10.1086/209482

Online Search After ChatGPT. (o. J.).

Spatharioti, S. E., Rothschild, D. M., Goldstein, D. G., & Hofman, J. M. (2023). *Comparing Traditional and LLM-based Search for Consumer Choice: A Randomized Experiment* (No. arXiv:2307.03744). arXiv. https://doi.org/10.48550/arXiv.2307.03744

Swoboda, B., & Schramm-Klein, H. (2025). *Käuferverhalten: Customer Insights – Customer Journey – Customer Relations*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-45121-9

Werner, T., Soraperra, I., Calvano, E., Parkes, D. C., & Rahwan, I. (2024). *Experimental Evidence That Conversational Artificial Intelligence Can Steer Consumer Behavior Without Detection* (No. arXiv:2409.12143). arXiv. https://doi.org/10.48550/arXiv.2409.12143

Wolf, V., & Maier, C. (2024). ChatGPT usage in everyday life: A motivation-theoretic mixed-methods study. *International Journal of Information Management*, *79*, 102821. https://doi.org/10.1016/j.ijinfomgt.2024.102821

Xu, R., Feng, Y. (Katherine), & Chen, H. (2023). ChatGPT vs. Google: A Comparative Study of Search Performance and User Experience. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4498671

Anhang

Zur Unterstützung bei der sprachlichen und stilistischen Ausarbeitung der Bachelorarbeit wurde das Sprachmodell ChatGPT (Modell 4o) von OpenAI eingesetzt. Die Nutzung erfolgte primär zur Verbesserung von Rechtschreibung, Grammatik und Ausdruck sowie zur stilistischen Verfeinerung und Erhöhung der Verständlichkeit einzelner Textabschnitte. In einigen Fällen wurden auch ganze Absätze neu formuliert, um inhaltliche Klarheit, Kohärenz und einen akademischeren Schreibstil sicherzustellen. Alle mithilfe von ChatGPT erarbeiteten oder überarbeiteten Inhalte wurden vom Verfasser kritisch geprüft, hinsichtlich ihrer inhaltlichen Richtigkeit und wissenschaftlichen Nachvollziehbarkeit reflektiert und eigenverantwortlich in die Arbeit integriert. Die konzeptionelle Ausarbeitung, die Struktur und die Argumentation der Arbeit basieren vollständig auf der Eigenleistung des Verfassers.

(Optional): Die folgende Übersicht zeigt die im Rahmen der empirischen Untersuchung genutzten Dokumente:

- BA_Umfrage_Interview_Ansicht: Ansicht der Fragebögen in allen vier Aufgabenversionen und Bewertungenskriterien
- BA_Umfrage_Interview_Data: Rohdaten der quantitativen Online-Erhebung
- BA_Umfrage_Interview_Auswertung: Aufbereitete Ergebnisse und grafische Darstellung zur Hypothesenprüfung