Gestaltung von Chatbot-Interfaces und Avataren

Jan Herbst
Hochschule Aalen
Aalen, Deutschland
82784@studmail.htw-aalen.de

Gabriel Roth
Hochschule Aalen
Aalen, Deutschland
82798@studmail.htw-aalen.de

Florian Merlau
Hochschule Aalen
Aalen, Deutschland
81775@studmail.htw-aalen.de

Zusammenfassung

Chatbots haben sich in den letzten Jahren zu einer zunehmend relevanten Technologie in digitalen Interaktionsprozessen entwickelt, welche in ganz unterschiedlichen Altersgruppen relevanz findet und heutzutage kaum noch weg zu denken ist. Neben funktionalen Aspekten gewinnen die soziale Gestaltung, das visuelle Erscheinungsbild, sowie die Ausarbeitung von Avataren an Bedeutung, um hochwertige und nutzerorientierte Kommunikationsprozesse zu ermöglichen.

Dieses Paper untersucht theoretische Grundlagen sozialer und visueller Chatbot-Gestaltung und analysiert den Einfluss von Designentscheidungen auf Nutzererfahrung, Vertrauen und Interaktionsqualität. Dabei werden unterschiedliche Avatar-Typen, deren Realitätsgrad, Körpersprache, Emotionen sowie Personalisierungsoptionen betrachtet. Außerdem werden die Stimme und Interface-Designfaktoren wie Positionierung, Farbgebung und Layout untersucht. Zudem werden Präferenzen verschiedener Altersgruppen analysiert. Auf Basis dieser Analyse werden Gestaltungsprinzipien abgeleitet und in einer Evaluierung reflektiert. Ziel ist es, einen umfassenden Überblick über zentrale Designfaktoren von Chatbots zu geben und ihre Bedeutung für eine effektive Mensch–Computer-Interaktion herauszuarbeiten.

Keywords

Chatbots, Avatare, Interface-Design, Nutzerpräferenzen

ACM Reference Format:
Jan Herbst, Gabriel Roth, and Florian Merlau. 2026. Gestaltung von Chatbot-Interfaces und Avataren. In *Seminararbeit / Projektbericht, Hochschule Aalen*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Einleitung

Chatbots werden in digitalen Anwendungen zunehmend zu einer Grundlage, welche nicht mehr wegzudenken ist. Sie unterstützen Unternehmen, erleichtern Informationszugänge, begleiten Nutzer im Alltag und übernehmen zunehmend soziale Rollen in Systemen. Mit dem technologischen Fortschritt der künstlichen Inteligenz wandelt sich auch die Erwartungshaltung der Nutzer. Es wird erwartet, dass ein Chatbot nicht nur korrekt und effizient antwortet, sondern gleichzeitig auch sozial angemessen, visuell ansprechend und intuitiv bedienbar ist. Dadurch rücken Fragen der sozialen Interaktion, der visuellen Gestaltung sowie der Ausarbeitung von Avataren zunehmend in den Fokus der Forschung und Praxis.

Hochschule Aalen, Aalen, Deutschland
2026. ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Während frühe Chatbots hauptsächlich als textbasierte Dialogsysteme funktionierten, erlauben moderne Technologien die Integration multimodaler Elemente wie Sprache, grafische Avatare und komplexe Interface-Strukturen. Diese Entwicklungen eröffnen neue Gestaltungsmöglichkeiten, stellen Designer jedoch zugleich vor Herausforderungen: Welcher Avatar-Typ eignet sich für welche Anwendung? Wie realistisch sollte ein Avatar sein, um Vertrauen zu fördern, ohne den „Uncanny Valley"-Effekt zu erzeugen? Welche Rolle spielen Körpersprache, Emotionen und Personalisierung? Und wie beeinflussen Interface-Layout, Farbgestaltung oder die Platzierung des Avatars die Nutzerwahrnehmung? Auch die Frage, wie sich Präferenzen je nach Altersgruppe unterscheiden, gewinnt an Bedeutung, da Chatbots zunehmend für diverse Zielgruppen eingesetzt werden.

Das vorliegende Paper gibt einen systematischen Überblick über theoretische Grundlagen und wesentliche Gestaltungsfaktoren moderner Chatbots. Es untersucht die sozialen und visuellen Elemente, die das Nutzererlebnis prägen und beleuchtet etablierte Ansätze. Ziel ist es, ein grundlegendes Verständnis dafür zu schaffen, wie Chatbots so gestaltet werden können, dass sie sowohl funktional als auch authentisch für den Menschzen wirken. Die daraus gewonnen Kentnisse werden anschließend in einer Evaluierung zusammengefasst und leitet Implikationen für zukünftige Entwicklungen ab.

2 Theoretischer Hintergrund (Gabriel)

2.1 Generative Architektur und Pipeline (technischer Hintergrund)

Ein generativer Chatbot mit Avatar besteht aus einer mehrstufigen Pipeline vom Eingabesignal bis zur multimodalen Ausgabe. Text- oder Spracheingaben werden zunächst normalisiert; bei Sprache erfolgt eine automatische Spracherkennung (ASR), die heute meist auf selbstaufmerksamen Encoder-Architekturen wie dem *Conformer* basiert, um Robustheit und Latenz auszubalancieren [5]. Optionale Vorfilter (PII-Redaktion, Sicherheits-/Policy-Checks) laufen vor der eigentlichen Antwortfindung.

Die Antwortentstehung beginnt mit dem *Kontextaufbau*: System-/Rolleninstruktionen, Gesprächsverlauf (ggf. zusammengefasst), Benutzereinstellungen (z. B. Sprache, Vorlesen) und situative Parameter werden zu einem Prompt verbunden. Um Wissenslücken und Halluzinationen zu reduzieren, wird der Prompt häufig mit *Retrieval-Augmented Generation* (RAG) angereichert: Eine semantische Suche liefert Top-k Belegdokumente, die der Sprachmodell-Generierung als Beweisgrundlage dienen [6]. Für aufgabenorientierte Interaktionen werden zudem *Tool-/Function-Calling*-Mechanismen

eingesetzt: Das Sprachmodell plant und ruft externe Funktionen (z. B. Wissensdatenbanken, Kalender, Rechnungen, Websuche) auf und integriert die Ergebnisse wieder in den Dialog. Planungs- und Handlungsansätze wie *ReAct* oder schwach überwachtes Tool-Lernen zeigen, wie Modelle rationale Schritte und API-Aufrufe verzahnen [13**?**].

Die eigentliche Textgenerierung (*decoding*) wird über Temperatur, Top-p und Stop-Sequenzen kontrolliert; bei strukturierter Übergabe an die Oberfläche (UI) sind JSON-Schemata üblich (z. B. neben „message“ auch „ui_directives“ wie „tts_style“ oder „avatar_stage=true“). Qualität und Nützlichkeit profitieren von zielgerichtetem Alignment (z. B. RLHF) und regelbasierten Sicherheitsfiltern; verfassungsbasierte Ansätze ergänzen diese Ebene um konsistente Prinzipien [**? ?**]. Für Dialogfluss kommt *Streaming* zum Einsatz: Token werden sukzessive an das UI gesendet, sodass Tippen-Indikatoren kurz bleiben und die Antwort früh sichtbar wird.

Die Audio-/Videoausgabe setzt auf neuronale TTS-Stacks: ein Text-Encoder mit Prosodie-/Stilkontrolle, ein Akustikmodell (z. B. Mel-Spektrogramm-Generator) und ein schneller Neural-Vocoder (etwa HiFi-GAN) erzeugen natürlich klingendes Audio in niedriger Latenz [12**?**]. Für Avatare werden *Viseme*-Ereignisse aus der Synthese genutzt oder per Audio-zu-Lippen-Netz (z. B. Wav2Lip) geschätzt und anschließend auf 2D/3D-Rigs bzw. Blend-Shapes gemappt, damit Mundbilder und Kopfbewegungen synchron laufen [7**?**]. Bei kurzzeitiger Asynchronität hat Audio Priorität; die Animation wird geglättet oder temporär reduziert, um Verständlichkeit zu sichern.

Im Betrieb sichern Caches und Teilkontext-Speicher (z. B. Zusammenfassungen, Dokument-Embedding-Cache) die Latenz. Beobachtbarkeit (Prompt-/Tool-Logs, Fehlerraten, Korrekturpfade) und regelmäßige Evaluation (z. B. Wissensfragen mit Ground Truth, Nutzungsmetriken, MOS für TTS) ermöglichen iterative Verbesserungen. Insgesamt verbindet die generative Architektur Wissenseinbindung (RAG), Handlungsfähigkeit (Tool-Calls) und kontrollierte Ausgabe (Schema, Stil, Sicherheit) mit einer durchgehenden Echtzeitkette bis zur TTS-/Avatar-Synchronisation.

2.2 Definition und Funktionsweise eines Chatbots

Bei Chatbots handelt es sich um ein Computergprogramm, welches Dialogsysteme mit natürlichsprachlichen Fähigkeiten textueller oder auditiver Art, führt. Dabei wird zwischen Chatbots ohne KI, konversationellen und generativen KI-Chatbots, so wie virtuellen Agenten unterschieden. Ein Chatbot ohne KI kann sich auch schon um eine starre Menü-Navigation im Entscheidungsbaum-Stil handeln. Konversationelle KI-Chatbots können Fragen oder Kommentare von Benutzern verarbeiten. Generative KI-Chatbos nutzen Sprachmodelle wie bspw. LLMs (Large Language Models). Durch LLMs können auch Inhalte, wie Bilder und Audio erzeugt werden. Bei den virtuellen Agenten handelt es sich um eine weitere Weiterentwicklung von KI-Chatbod-Software, welche neben Deep Learning, um sich im Laufe der Zeit selbst zu verbessern, auch oft mit robotischer Prozessautomatisierung (RPA) in einer einzigen Schnittstelle kombinieren. Daurch kann ohne weitere zusätzliche Eingriffe eines Menschen direkt auf die Absicht des Nutzers reagiert werden.

Die Funktionsweise von Chatbots war zu Beginn wie einfache FAQ-Systeme mit vordefiniterten Antworten und konnten kaum die reine natürliche Sprache verstehen. Über die Zeit entwickelten sich regelbasierte Chatbots und daraufhin auch Chatbots mit natürlicher Sprachverarbeitung und maschinellem Lernen, sodass Anfragen des Benutzers kontextabhängig erkannt und besser vorhergesagt werden können. Die modernen KI-Chatbots heutzutage nutzen das natürliche Sprachverständnis, auf Englisch: natural language understanding (NLU). Dadurch kann die Bedeutung offener Benutzereingaben erkannt und sogar Tippfehlern bis Fehlern in der sprachlichen Übersetzung überwunden werden, was auch dazu führt, dass die Chatbots allgemein vielseitiger und leistungsfähiger als in der vergangenen Zeit geworden sind.

2.3 Soziotechnische Gestaltung eines Chatbots

Einige Studien zeigen, dass bei der Entwicklung von Chatbots nicht nur der Einsatz modernster Technologien im Mittelpunkt steht, sondern auch die sozialen Faktoren einen sehr großen Faktor auf den Erfolg eines Chatbots haben können. Die soziotechnische Gestaltung basiert auf sowohl technischen, als auch sozialen Faktoren und Erkenntnissen. Sie zielt darauf ab diese möglichst gleichberechtigt in den Gestaltungsprozess einzubringen. Eine konkrete Richtlinie, die vorschreibt, wie die Gestaltung eines Chatbots auszusehen hat, gibt es bisher allerdings noch nicht. [4]

Um die Interaktion mit dem Chatbot so natürlich wie möglich gestalten zu können eignet es sich dem Chatbot eine Persönlichkeit zu geben und soziale Signale einzusetzen. Beispielhafte soziale Signale sind Smalltalk oder auch der Einsatz von Emojis.[4]

In der zwischenmenschlichen Kommunikation spielen soziale Signale eine grosse Rolle und sind für den Aufbau und Erhalt erfolgreicher sozialer Beziehungen unerlässlich. [4]

Soziale Signale sind für den Menschen implizite Hinweisreize wie Blickkontakt, Lächeln oder Stimmlage, aus denen sich soziale Informationen ableiten lassen. Unterbewusst werden diese Signale oft verwendet, um die Gefühle und Gedanken anderer Personen interpretieren zu können.[4]

Laut Forschungen hat sich gezeigt, dass auch in der Interaktion mit Computern und anderen Technologien soziale Signale unterbewusst wahrgenommen und automatisch darauf reagiert wird. [4]

In Abbildung X-TODO: Entstehung von sozialen Reaktionen in der Interaktion mit Chatbots wird veranschaulicht, wie soziale Reaktionen in der Interaktion mit Chatbots entstehen (basierend auf Krämer 2005 sowie Nass und Moon 2000). Nutzer nehmen zunächst unbewusst Merkmale des Chatbots wahr, wie sie aus der Kommunikation mit Menschen auch bereits bekannt sind, wie bspw. einem Namen. Diese Merkmale wirken als soziale Signale. Durch sie schreiben Nutzer dem Chatbot automatisch menschliche Eigenschaften zu und

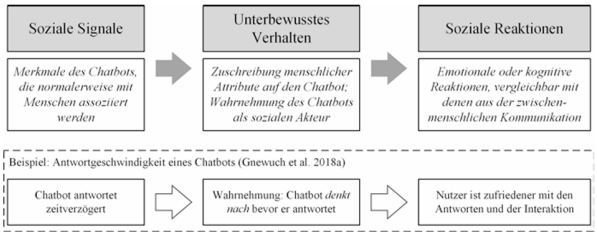

Abbildung 1: soziale Reaktionen [4]

übertragen vertraute Verhaltensmuster aus zwischenmenschlichen Gesprächen auf die Interaktion. Der Chatbot wird dadurch als sozialer Akteur wahrgenommen, was emotionale und kognitive Reaktionen auslöst, die denen in realen sozialen Interaktionen ähneln. [4]

Es wird zwischen verbalen, auditiven, unsichtbaren und visuellen Signalen unterschieden.

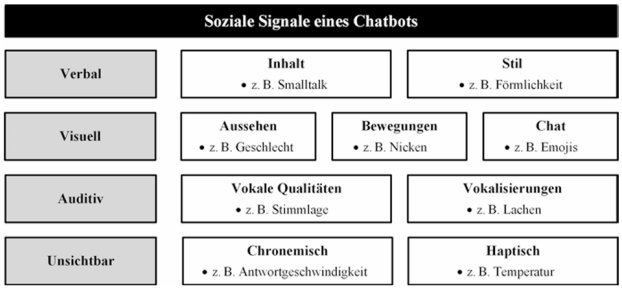

Abbildung 2: Klassifikation sozialer Signale von Chatbots [4]

Verbale Signale sind alle sozialen Signale, welche durch die Verwendung von Worten erzeugt werden.

Auditive Signale sind alle hörbaren sozialen Signale, welche durch nonverbale Laute erzeugt werden.

Unsichtbare Signale sind alle sozialen Signale, welche nicht gesehen oder gehört werden können.

Visuelle Signale sind alle nonverbalen sozialen Signale, welche visuell wahrnehmbar sind und durch das Aussehen, Bewegungen und die Chat-Umgebung erzeugt werden können.

Die Vielzahl möglicher sozialer Signale eröffnet zahlreiche Gestaltungsoptionen, welche Wahrnehmung und Verhalten der Nutzer maßgeblich beeinflussen können. Dementsprechend haben Designentscheidungen einen großen Einfluss auf den Erfolg eines Chatbots. Eine gezielte Ausgestaltung dieser Signale kann Vertrauen fördern, die Nutzerzufriedenheit steigern und die Grundlage für eine langfristige Beziehung zwischen Chatbot und Nutzer schaffen.[4]

2.4 Einfluss visueller Gestaltung

Forschungen zeigen, dass visuelle soziale Signale, wie bspw. ein lächelnder Avatar, das Verhalten der Nutzer beeinflussen. Menschen lächeln häufiger zurück und interagieren länger, wenn ein Avatar sie anlächelt. Auch andere Signalarten, wie verbale Hinweise, wirken sich aus: Smalltalk beeinflusst die wahrgenommene Persönlichkeit eines Chatbots, und humorvolle Beiträge wie Witze lassen ihn für Nutzer humorvoller erscheinen.[4]

Die Theorie des Uncanny Valley nimmt auch Einfluss auf die visuelle Gestaltung.

Gleichzeitig zeigen Studien, dass Nutzer trotz dieses Risikos bestimmte soziale Signale erwarten, weil sie diese aus menschlicher Interaktion gewohnt sind. Da natürliche Sprache ohnehin Assoziationen zu menschlichem Verhalten erzeugt, gibt es keine allgemeingültige Antwort auf den optimalen Grad an Menschenähnlichkeit. Daher sollte der Gestaltungsprozess kontinuierlich prüfen, wie soziale und visuelle Signale wirken, um den passenden Grad der Menschenähnlichkeit situativ an den Nutzer anpassen zu können.[4]

3 Gestaltung des Chatbot-Avatars

Die visuelle Gestaltung von Chatbot-Avataren ist weit mehr als eine ästhetische Randentscheidung. Sie bestimmt, wie sehr ein System als sozialer Interaktionspartner wahrgenommen wird, welche Erwartungen an seine „Kompetenz“ und „Menschlichkeit“ geknüpft werden und ob Nutzende sich grundsätzlich wohl damit fühlen, mit ihm zu interagieren.Thaler et al. zeigen, dass bereits relativ kleine Variationen in der visuellen gestaltung verkörperter Konversationsagenten (Embodied Conversational Agents, ECAs) messbare Effekte auf wahrgenommene Menschlichkeit, Unheimlichkeit und Attraktivität haben.

Rashik et al. kategorisieren das Design von Avataren in die folgenden Dimensionen: Aussehen, Geschlecht, Bewegung, Interaktion, Dimensionalität, Eingabemodalität, Gesprächstyp, Gerätetyp, Anzahl der Agenten und Anwendung [**?**]. Dabei denken sie Avatar-Design nicht entlang eines einzigen Realismus-Kontinuums, sondern als Kombination verschiedener Eigenschaften. Unter diesen Dimensionen beschreiben insbesondere Aussehen, Geschlecht und Dimensionalität die grafische Darstellung des Chatbots. Gleichzeitig beziehen die Autoren aber auch weitere zentrale Eigenschaften, die nicht unmittelbar die visuelle Gestaltung betreffen.

Während Thaler et al. vor allem Unterschiede in der visuellen Ausgestaltung untersuchen, richtet Karimova den Blick stärker auf Kontextabhängigkeiten. Im Design von Expert:innen-Chatbots können hochgradig anthropomorphe Gestaltungen bei vielen Nutzenden eher Misstrauen als Vertrauen hervorrufen, da sie als ablenkendes „psychologisches Decoy“ wahrgenommen werden [**?**]. Insbesondere in Bereichen wie Rechts- oder Personalwesen müssen Transparenz, Geschwindigkeit und die klare Erkennbarkeit als nichtmenschliche Agenten priorisiert werden. In solchen Kontexten werden eher minimalistisch gestaltete, nur leicht anthropomorphisierte Chatbots als vertrauenswürdiger bewertet, da sie die Aufgabenerfüllung und Reaktionszeit weniger beeinträchtigen. Zudem spielt die Nutzungshäufigkeit eine Rolle: Tägliche Nutzende neigen dazu, anthropomorphe Merkmale,

die ineffizient oder fehleranfällig erscheinen, abzulehnen, während unregelmäßige Nutzende diese humanen Eigenschaften eher schätzen.

3.1 Erscheinungsformen von Chatbot-Avataren

Rashik et al. [**?**] beschreiben die Dimension Appearance als das äußere Erscheinungsbild eines Conversational Agents und unterscheiden vier grundlegende Formen: Human-like, Robotic, Object und Animal-like. Human-like Avatare ähneln dabei einer menschlichen Person, etwa durch Gesichtszüge, Frisur und Kleidung. Robotic Avatare sind klar als Maschine gestaltet (z. B. rechteckiges „Robotergesicht“ mit Augen und Antennen), können aber einzelne menschliche Merkmale tragen. Object Avatare repräsentieren unbelebte Gegenstände wie Logos, Früchte oder Symbole, während Animal-like Avatare an Tiere angelehnt sind und oft cartoonhaft gezeichnet werden.

Eine weiterführende Perspektive bieten Hadjiev und Araki [**?**], die verschiedene konkrete Avatar-Designstile für ein Chatbot-System miteinander vergleichen. Untersucht werden unter anderem menschliche und robotische Figuren sowie unterschiedliche Darstellungsstile wie Fotobilder, Animationen oder niedrig aufgelöste Pixelgrafiken. Sie zeigen, dass sich selbst innerhalb einer Erscheinungsform (z. B. Human-like oder Robotic) die Wirkung der Avatare deutlich unterscheidet, je nachdem wie realistisch oder stilisiert sie umgesetzt sind.

3.2 Anthropomorphismus und Realitätsgrad

Anthropomorphismus beschreibt das zuordnen von Menschlichen Eigenschaften, Motivationen, Absichten und Emotionen zu nicht Menschlichen Entitäten [**?**]. In der Gestaltung eines Chatbots wird Anthropomorphismus oft als Designstrategie eingesetz um das vertrauen und die soziele Präsent zu einem Chatbot zu erhöhen. Cao et al. beschreibt ‚dass anthropomorphe Avatare eine wichtiger role für den ersten eindruck eine benutzer sind. Insbesondere menschliche und karikierte Dahrgestellte Chatbots habe einen Positiven einfluss auf den ersten eindruck [**?**].

Gleichzeitig wird davon gewarnt, Anthropomorphismus ausschließlich als vorteill zu sehen. Dies wäre ein Missverständnis, da Thaler et al. empirisch zeigen, dass mit zunehmender wahrgenommener Menschlichkeit auch das Empfinden von Unheimlichkeit steigt und sich dadurch der bekannte Uncanny-Valley-Effekt abzeichnet [**?**].Hadjiev et al. kommen in einer Evaluation verschiedener Avatar-Designs zu einem ähnlichen Schluss. Stärker anthropomorphe Chatbot-Avatare waren nicht immer besser für die Nutzererfahrung als weniger menschenähnliche Varianten[**?**].

3.2.1 Typen von Chatbot-Avataren entlang des Realitätsgrades. Entlang dieser Skala lassen sich mehrere Typen von Chatbot-Avataren unterscheiden:

Minimalistische Avatare Die symbolischen Darstellungen eines Avatars stehen am unteren Ende der Skala. Dieser zeigt keinerlei menschliche Eigenschaften. Diese Form besteht häufig aus Icons oder Logos und vermittelt eine klare technische Identität. Studien zeigen, dass stark reduzierte Avatare besonders in professionellen Anwendungen wie HR oder Recht als vertrauenswürdiger empfunden werden [**?**].

Stilisierte Avatare Die zweite Stufe umfasst stilisierte oder fantasiehaft-cartoonhafte menschliche Darstellungen, die bewusst von realistischer Menschenabbildung abweichen. Sie wirken freundlich und vermitteln soziale Präsenz, ohne dass realistische Mimik oder komplexe Bewegungen erforderlich sind [**?**]. Neuere Studien zeigen jedoch, dass innerhalb dieser Kategorie ein wichtiger Unterschied zwischen menschlichen Avataren und Avataren mit tierischen Charakteristiken besteht: Menschliche und karikierte Figuren werden deutlich besser bewertet als Avatare mit tierischen Charakteristiken in Bezug auf wahrgenommene Wärme, Kompetenz, soziale Präsenz und Vertrauen [**?**].

Photorealistische Avatare Photorealistische oder High-Fidelity-Avatare stellen detaillierte menschliche Gesichter dar. Diese Avatare erzeugen eine deutlich höhere soziale Präsenz und Nähe zum menschlichen Erscheinungsbild [**?**]. Die hohe Detailtiefe und realistische Mimik steigert Vertrauen und Glaubwürdigkeit, kann aber gleichzeitig Irritationen hervorrufen, wenn Verhalten, Gestik oder Reaktionen nicht konsistent sind [**? ?**]. Photorealistische Avatare befinden sich somit in einem Bereich, in dem das Risiko des Uncanny Valley besonders relevant ist.

Vollständig verkörperte Avatare Am Ende sind die Embodied- oder Literal-Avatare, die als vollständig verkörperte menschliche Figuren dargestellt werden, entweder in 2D oder 3D. Diese Avatare beinhalten Körperhaltung, Gestik und Bewegungen, wodurch die soziale Präsenz maximal erhöht wird [**?**]. Gleichzeitig steigt die Anfälligkeit für Uncanny-Valley-Effekte, wenn Animation oder Verhalten nicht mit der visuellen Glaubwürdigkeit übereinstimmen [**? ?**].

3.2.2 Der Uncanny-Valley-Effekt. Die Uncanny-Valley-Hypothese von Mori et al [**?**] beschreibt einen nichtlinearen Zusammenhang zwischen dem Grad menschlicher Ähnlichkeit und der affektiven Reaktion und verdeutlicht, dass übermäßiger Realismus eher Abneigung als Zuneigung hervorrufen kann. Insbesondere in dem Bereich knapp unterhalb der perfekten Menschenähnlichkeit entsteht ein üncanny valley", in dem kleinste Abweichungen im Aussehen oder Verhalten als besonders irritierend, unheimlich oder sogar abstoßend erlebt werden.

Für Chatbot-Avatare ist dieses Tal deshalb relevant, weil steigender Anthropomorphismus zunächst positive Effekte auf Sympathie, wahrenmung und soziale Präsenz haben kann, ab dies ab einem bestimmten Realitätsgrad jedoch kippt. Hochrealistische, aber nicht vollständig glaubwürdige Gesichter aktivieren beim Gegenüber Erwartungen an echte menschliche Interaktion, die das System technisch und sozial nicht einlösen kann. Neuere empirische Arbeiten legen nahe, dass das Erwartungs- und Inkonsistenzproblem nicht

nur die visuelle Dimension betrifft, sondern auch das Zusammenspiel von Erscheinungsbild, verbaler Kommunikation und Rollenverständnis des Systems [**?**].

Gestaltungsstrategisch folgt daraus, dass es in vielen Anwendungsfeldern sinnvoller sein kann, bewusst unterhalb der kritischen Zone zu bleiben. Etwa durch stilisierte oder karikierte Darstellungen anstatt so realistisch wie möglich zu designen.

3.3 Personalisierung von Avataren

4 Stimmen von Chatbots

Bei der Gestaltung synthetischer Stimmen stellt sich die Frage, wie realistisch eine Stimme klingen sollte, damit sie von Nutzenden akzeptiert wird. Stimmen übernehmen in der Interaktion mit Chatbots nicht nur die Rolle der Informationsübertragung, sondern transportieren auch soziale Signale wie Freundlichkeit, Kompetenz oder Empathie. Während in der visuellen Gestaltung häufig vom sogenannten „Uncanny Valley“ gesprochen wird, ist die Forschungslage für Stimmen weniger eindeutig.

Baird et al. untersuchten in einer Hörstudie, ob sich ein solcher Effekt auch bei synthetischer Sprache beobachten lässt. In der Studie bewerteten 25 Teilnehmende die wahrgenommene Menschlichkeit (*Human Likeness*) und die Beliebtheit (*Likeability*) von 13 synthetischen, männlichen deutschen Stimmen sowie einer menschlichen Vergleichsstimme [2]. Die Ergebnisse zeigen, dass Stimmen umso beliebter bewertet wurden, je menschlicher sie wahrgenommen wurden. Ein deutlicher Einbruch der Beliebtheit bei sehr hoher Menschlichkeit, wie er aus dem visuellen Uncanny-Valley-Modell bekannt ist, konnte in dieser Studie nicht festgestellt werden.

Gleichzeitig betonen die Autoren, dass diese Ergebnisse nicht als allgemeingültige Aussage verstanden werden dürfen. Die Wahrnehmung von Stimmen hängt stark vom verwendeten Stimulusmaterial, vom Anwendungskontext sowie von den Erwartungen der Nutzenden ab. Insbesondere in sensiblen oder professionellen Anwendungsfeldern können andere Kriterien wie Klarheit, Neutralität oder Vertrauenswürdigkeit wichtiger sein als maximale Natürlichkeit.

4.1 Zusammenhang von Menschlichkeit und Beliebtheit

Die Auswertung von Baird et al. zeigt einen klaren Zusammenhang zwischen wahrgenommener Menschlichkeit und Beliebtheit der Stimmen: Mit steigender Human Likeness nahm auch die Likeability zu [2]. Stimmen, die natürlicher klangen, wurden häufiger als angenehm, freundlich und vertrauenswürdig eingeschätzt. Für die Gestaltung von Chatbots deutet dies darauf hin, dass eine natürlich klingende Sprachsynthese die Akzeptanz erhöhen kann.

Gleichzeitig sollte berücksichtigt werden, dass eine hohe Menschlichkeit auch Erwartungen an Gesprächsverhalten, Reaktionsgeschwindigkeit und inhaltliche Kompetenz erzeugt. Werden diese Erwartungen nicht erfüllt, kann dies zu Enttäuschung oder Vertrauensverlust führen. Daher sollte der Grad der Natürlichkeit einer Stimme immer in Relation zur tatsächlichen Leistungsfähigkeit des Systems gewählt werden.

4.2 Einfluss des technischen Entwicklungsstands

Baird et al. verglichen Stimmen aus verschiedenen Entwicklungsstufen der Sprachsynthese, von älteren Verfahren bis hin zu neueren, datengetriebenen Ansätzen. Dabei zeigte sich, dass modernere Systeme im Durchschnitt höhere Bewertungen in Bezug auf Human Likeness und Likeability erzielten als ältere Syntheseverfahren [2]. Diese Ergebnisse sind jedoch relativ zu den verglichenen Systemen zu verstehen.

Die Autoren betonen, dass auch die bestbewerteten synthetischen Stimmen weiterhin klar von der menschlichen Referenzstimme unterscheidbar waren. Die Studie zeigt somit keine Gleichwertigkeit zwischen synthetischer und menschlicher Sprache, sondern vielmehr eine graduelle Annäherung moderner Systeme an bestimmte Merkmale menschlicher Sprachproduktion [2].

4.3 Akustische Merkmale natürlicher Stimmen

Neben subjektiven Bewertungen führten Baird et al. auch eine explorative akustische Analyse durch. Dabei wurde unter anderem die Varianz der Grundfrequenz ($F0$) untersucht. Stimmen mit geringer Tonhöhenvariation wirkten monoton und wurden tendenziell als weniger menschlich und weniger angenehm empfunden [2]. Eine größere Variation der Tonhöhe, wie sie in natürlicher menschlicher Sprache üblich ist, ging hingegen mit höheren Bewertungen einher.

Zusätzlich beschreiben die Autoren prosodische Merkmale wie das sogenannte *phrase-final lengthening*, also eine leichte Dehnung am Ende von Satzteilen, die bei moderneren Systemen häufiger beobachtet wurde. Solche Merkmale tragen dazu bei, synthetische Sprache weniger mechanisch wirken zu lassen. Da diese Analyse jedoch auf einer begrenzten Anzahl von Stimuli basiert, sollten die Ergebnisse vor allem als Hinweise verstanden werden und nicht als eindeutige Ursache-Wirkungs-Zusammenhänge.

Insgesamt zeigen die Ergebnisse, dass die Wahrnehmung synthetischer Stimmen weniger anfällig für starke Uncanny-Valley-Effekte zu sein scheint als die visuelle Gestaltung von Avataren. Für das Design von Chatbots bedeutet dies, dass Investitionen in eine hochwertige Sprachsynthese ein wichtiger Faktor für Akzeptanz und Nutzerzufriedenheit sein können, gleichzeitig jedoch immer im Zusammenspiel mit Dialogqualität, Kontext und Nutzererwartungen betrachtet werden sollten.

5 Positionierung des Avatars im Interface

Die Positionierung eines Avatars im Interface hat einen wesentlichen Einfluss auf Wahrnehmung, Aufmerksamkeit und Interaktionsqualität. Avatare transportieren neben sprachlichen Inhalten auch soziale und nonverbale Signale, etwa durch Blickrichtung, Mimik oder Körperhaltung. Dadurch besteht die Gefahr, dass Nutzende ihre Aufmerksamkeit zwischen mehreren visuellen Reizen aufteilen müssen. Ein zentrales Ziel der Interface-Gestaltung ist es daher, Ablenkung

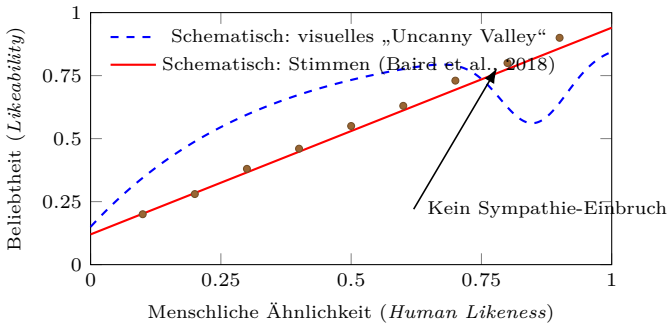

Abbildung 3: Konzeptionelle Gegenüberstellung zweier Wahrnehmungsmodelle: (1) das klassisch beschriebene „Uncanny Valley“ in der visuellen Domäne (nach Mori et al.) und (2) die in Baird et al. (2018) berichtete Beziehung zwischen wahrgenommener Menschlichkeit und Beliebtheit synthetischer Stimmen. Die Kurven dienen ausschließlich der Illustration qualitativer Trends; es werden keine empirischen Messdaten oder direkt vergleichbaren Skalen dargestellt.

zu reduzieren und eine klare visuelle Hierarchie zwischen Gesprächsinhalt, Interaktionskontrollen und Avatar zu etablieren.

Frühe Arbeiten zu Embodied Conversational Agents zeigen, dass die Verkörperung eines Agenten vor allem konversationelle Prozesse unterstützt, etwa durch soziale Hinweise, Aufmerksamkeitslenkung oder Turn-Taking, und nicht primär der Vermittlung inhaltlicher Information dient [3]. Aus dieser Perspektive ergibt sich, dass Avatare im Interface in erster Linie eine unterstützende Rolle einnehmen, während der eigentliche Gesprächsinhalt – insbesondere textuelle Informationen – im Fokus der Aufmerksamkeit verbleibt. Für das Interface-Design bedeutet dies nicht, dass Avatare grundsätzlich vermieden werden sollten, wohl aber, dass sie nicht zwangsläufig zentral oder visuell dominant platziert sein müssen, wenn gleichzeitig komplexe Inhalte verarbeitet werden.

5.1 Avatarplatzierung und visuelle Hierarchie

Aus der unterstützenden Funktion des Avatars folgt, dass seine Platzierung im Interface so gewählt werden sollte, dass Text, Dialogverlauf und Interaktionskontrollen klar im Vordergrund stehen. Eine seitliche oder periphere Positionierung ermöglicht es Nutzenden, soziale Signale des Avatars wahrzunehmen, ohne den Blick dauerhaft vom inhaltlichen Dialog abzuwenden. Eine solche visuelle Hierarchie kann die kognitive Verarbeitung erleichtern und unterstützt ein Interface-Design, bei dem der Avatar den Dialog ergänzt, anstatt mit ihm um Aufmerksamkeit zu konkurrieren [3].

Empirische Befunde weisen zudem darauf hin, dass Nutzende unterschiedliche Präferenzen im Umgang mit avatarbasierten Lern- und Assistenzsystemen aufweisen. Tan et al. unterscheiden zwischen eher *task-orientierten* Nutzenden, die einen effizienten Zugriff auf textuelle Informationen bevorzugen, und *engagement-orientierten* Nutzenden, die Avatare als motivierendes oder sozial unterstützendes Element wahrnehmen [11]. Diese Unterschiede sprechen dafür, Avatare so in das Interface zu integrieren, dass sie für interessierte Nutzende präsent sind, ohne für andere den Zugang zu Informationen zu erschweren. Eine visuell sekundäre Platzierung, etwa in einer Seitenspalte, stellt eine mögliche Designlösung dar, um diesen unterschiedlichen Nutzungsstilen gerecht zu werden.

Neben der räumlichen Platzierung beeinflusst auch das Blickverhalten des Avatars die Wahrnehmung der Interaktion. Aktuelle Studien zeigen, dass die Blickrichtung eines Avatars Nutzerreaktionen deutlich beeinflussen kann und mit psychologischen Effekten wie erhöhter Selbstaufmerksamkeit verbunden ist [14]. Ein dauerhaft direkter Blickkontakt kann – abhängig vom Nutzungskontext – als sozial intensiv wahrgenommen werden und sollte daher mit Bedacht eingesetzt werden. In Kombination mit einer zentralen und visuell dominanten Platzierung besteht die Gefahr, dass der Avatar als aufdringlich empfunden wird. Eine weniger dominante Positionierung, ergänzt durch zurückhaltendes oder situationsabhängiges Blickverhalten, kann dazu beitragen, die Interaktion angenehmer und weniger belastend zu gestalten.

Auf Basis dieser theoretischen und empirischen Erkenntnisse wurde bewusst ein Interface-Layout gewählt, in dem der Avatar als sekundäres visuelles Element positioniert ist. Diese Layoutgestaltung stellt keine direkte Vorgabe aus der Literatur dar, sondern ist das Ergebnis einer Designentscheidung, die darauf abzielt, soziale Präsenz zu ermöglichen, ohne den inhaltlichen Dialog zu überlagern.

5.2 Nutzerkontrolle und Aufmerksamkeitssteuerung

Neben der reinen Positionierung des Avatars ist die Möglichkeit zur Nutzerkontrolle ein zentraler Faktor für eine positive Interaktion. Gestaltungsrichtlinien für Mensch-KI-Interaktion empfehlen, Nutzenden Kontrolle über das Verhalten und die Darstellung von KI-Systemen zu geben, um Überforderung zu vermeiden und Vertrauen zu fördern [1]. Diese Empfehlungen lassen sich auch auf avatarbasierte Interfaces übertragen.

Konkret bedeutet dies, dass Nutzende Einfluss auf die Sichtbarkeit und Prominenz des Avatars haben sollten. Optionen zum Reduzieren, Verkleinern oder zeitweisen Ausblenden des Avatars ermöglichen es, die Aufmerksamkeit gezielt auf den Gesprächsinhalt zu lenken. Gleichzeitig bleibt es den Nutzenden überlassen, den Avatar als soziales Element stärker einzubeziehen, wenn dies dem jeweiligen Nutzungskontext oder persönlichen Präferenzen entspricht. Eine solche flexible Gestaltung unterstützt unterschiedliche Nutzungssituationen und trägt zu einer insgesamt besser kontrollierbaren und akzeptableren Interaktion mit avatarbasierten KI-Systemen bei [1].

5.3 Farbgestaltung

Die Farbgestaltung eines Interfaces kann bei verkörperten Conversational Agents helfen, die Interaktion verständlicher

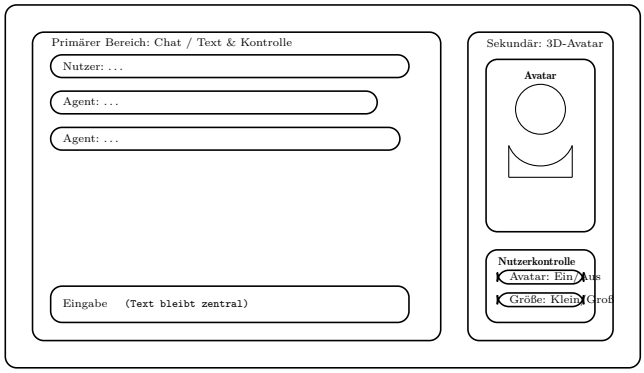

Abbildung 4: Konzeptionelles Layout eines Web-Interfaces mit 3D-Conversational-Avatar: Der Chatbereich bildet den primären Informations- und Kontrollbereich, der Avatar ist seitlich als sekundäres Element platziert. Zur Reduktion von Ablenkung und sozialem Druck wird eine zurückhaltende Blickrichtung sowie Nutzerkontrolle über Sichtbarkeit/Prominenz vorgesehen.

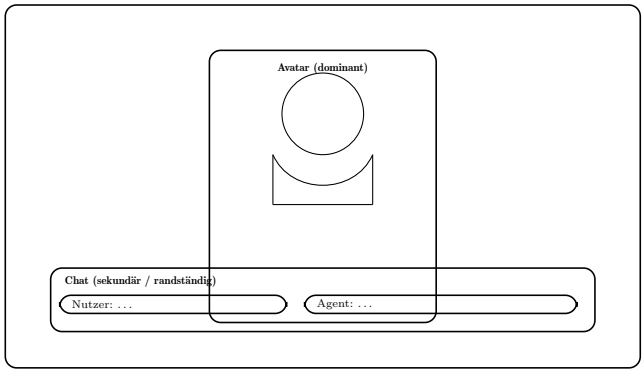

Abbildung 5: Beispiel für ein ungünstiges Layout eines avatarbasierten Conversational-Interfaces: Der Avatar ist zentral und visuell dominant platziert und richtet seinen Blick direkt auf die Nutzenden. Der Chatverlauf ist randständig angeordnet und untergeordnet. Diese Gestaltung kann zu erhöhter Ablenkung, sozialem Druck und reduzierter Kontrolle führen.

zu machen und den Kontext einer Antwort zu unterstützen. Dabei geht es nicht darum, dass Farben „Emotionen ersetzen“, sondern dass sie als zusätzliches, eher indirektes Signal im Interface wirken können.

Ein bekanntes Risiko bei stark menschenähnlichen Avataren ist der *Uncanny-Valley*-Effekt. Schwind et al. beschreiben dazu, dass negative Reaktionen besonders dann entstehen können, wenn verschiedene Wahrnehmungshinweise nicht zusammenpassen (“conflicting cues”), z. B. wenn ein Avatar sehr realistisch aussieht, aber Bewegung oder Ausdruck nicht dazu passen [9]. Solche Inkonsistenzen können durch Bewegung noch stärker auffallen [9].

Wichtig ist dabei: Nicht nur der Avatar selbst beeinflusst die Wahrnehmung, sondern auch die visuelle Umgebung, in der er gezeigt wird. Schneeberger et al. untersuchen dies in einem Szenario mit einem virtuellen Interviewer, der sich gleich verhält, während Hintergrund und Lichtbedingungen verändert werden. Die Ergebnisse zeigen, dass solche Änderungen die Bewertung des virtuellen Agents und der Interaktion beeinflussen können [8]. Damit ist es plausibel, Interface- und Umgebungsfaktoren (z. B. Farbstimmung, Helligkeit, Akzentfarben) als Teil des Gesamteindrucks mitzudenken.

Vor diesem Hintergrund kann es sinnvoll sein, emotionale oder soziale Signale nicht nur über Mimik oder Animationen des Avatars zu vermitteln, sondern auch über das Interface. Farben können hierbei als eine Art *Ambient Feedback* dienen: Sie geben einen Hinweis auf den Kontext (z. B. „sachlich“, „unterstützend“, „Warnung“), ohne dass dafür komplexe Gesichtsanimationen nötig sind. Allgemein wird in der HCI- und Affective-Computing-Literatur diskutiert, dass Licht und Farbe die emotionale Wahrnehmung beeinflussen können [10].

Daraus lassen sich folgende einfache Gestaltungsansätze ableiten:

- **Kontextunterstützung:** Hintergrund- oder Akzentfarben können den Charakter einer Systemantwort begleiten (z. B. kühl für sachliche Informationen, wärmer für unterstützende Antworten). Wichtig ist, dass die Farbänderungen dezent bleiben, damit sie nicht ablenken.
- **Fokussteuerung:** Wenn der Avatar wenig Aktivität zeigt (oder das System kurz lädt), können leichte Farbakzente die Aufmerksamkeit stärker auf den Textbereich lenken. Das kann helfen, dass Nutzende weniger auf mögliche Unstimmigkeiten im Avatar-Ausdruck achten.

Die genannten Punkte sind Gestaltungsentscheidungen. Sie sind nicht als „Beweis“ zu verstehen, dass Farbe das Uncanny Valley sicher verhindert, sondern als nachvollziehbare Designableitung aus (1) dem Problem inkonsistenter Hinweise bei Avataren [9] und (2) Befunden, dass auch Umgebung und visuelle Gestaltung die Bewertung virtueller Agents beeinflussen können [8].

6 Präferenzen von Avataren in unterschiedlichen Altersgruppen

In Abhängigkeit des Alters bevorzugen Personen unterschiedliche Designs, dieses gilt auch für die Avatere von Chatbots.

6.1 Literatur zu altersabhängigen Designpräferenzen

Nurhas et al. (2022) untersuchte mittels Q-Methodologie subjektive Designpräferenzen für Chatbots in intergenerationalen Innovationssettings. Durch diese Studie stellten sich vier typische Nutzer-Typen mit unterschiedlichen Prioritäten heraus.

(1) **Typ 1: Explorative/spielerische Nutzer**: Ältere Erwachsene betonen respektvolle Kommunikation, unterstützendes Verhalten und die Möglichkeit zu explorativen, spielerischen Interaktionen. Technische oder emotionale Barrieren werden von ihnen als wenig relevant eingeschätzt.

(2) **Typ 2: Thoughtful/unterstützende Nutzer**: Diese gemischte Gruppe aus jüngeren und älteren Erwachsenen legt Wert auf Gleichbehandlung aller Generationen, gemeinsame Zielorientierung und eine klare Kommunikation der Fähigkeiten und Grenzen des Systems. Emotionale Barrieren spielen hier eine geringere Rolle, kulturelle Unterschiede werden jedoch stärker beachtet.
(3) **Typ 3: Zielorientierte jüngere Nutzer**: Jüngere Erwachsene priorisieren Ausdrucksfähigkeit, klare Zielorientierung und die Wertschätzung aufgrund ihrer Ausbildung. Sie lehnen kulturelle Unterschiede und negative Urteile deutlich ab und stellen kollektive Zielerreichung in den Vordergrund.
(4) **Typ 4: Schnell reagierende empathische Nutzer**: Ebenfalls jüngere Erwachsene, die vor allem Empathie, schnelle Reaktionsfähigkeit und unmittelbare Verfügbarkeit hervorheben. Emotionale Barrieren werden gering bewertet, soziale Entspannung ist für sie weniger relevant.

6.2 Relevante Designaspekte im Bezug auf Chatbot-Avatare

Aus den am höchsten bewerteten Chatbot-Eigenschaften lassen sich mehrere, auch für Avatare relevante Designanforderungen ableiten:

- **Empathie und Unterstützungsverhalten**: Der Chatbot sollte nicht wertend sein, sondern Empathie zeigen. Solche sozialen Signale fördern Vertrauen und Teilhabe, insbesondere in intergenerationalen Settings, daher ist dies sehr wichtiger Aspekt bei der Chatbot-Avataren.
- **Immediacy / Reaktionsverfügbarkeit**: Schnelle Verfügbarkeit und unmittelbare Rückmeldung wurden von Benutzern erwünscht. Speziell für einen Avatar wäre dies die Interaktionslatenz, sowie die Präsenzgestaltung.
- **Expressivität und Zurückhaltung**: Jüngere Gruppen priorisieren expressive, zielorientierte Eigenschaften, wie bspw. eindeutige Hinweise. Andere Gruppen hingegen fanden spielerische/explorative Elemente eher unwichtig. Diese Divergenz spricht für adaptive Avatare, die sich an die Nutzergruppe anpassen können.
- **Transparenz und Grenzen**: Ein Teil der Teilnehmenden wünscht, dass die Assistenz ihre Fähigkeiten kommuniziert, also bspw. keine eigene Meinung nennen. Für Avatare bedeutet dies, dass diese eine klare Rolle sichtbar machen.

6.3 Implikationen für die Gestaltung von Avataren in verschiedenen Altersgruppen

Die Ergebnisse der Studie zeigen, dass bei jüngeren Erwachsenen sich herausstellte, da diese eine viel Breitere Auswahl an Designpräferenzen haben, während ältere Erwachsene insgesamt ein eher einheitlicheres Präferenzmuster zeigten. Damit wird gezeigt, dass bei der Gestaltung der Avatare die Bedürfnisse älterer Nutzer besonders berücksichtigt werden sollten. Um hingegen die Präferenzen jüngerer Nutzergruppen abzudecken, werden adaptive Strategien benötigen.

7 Fazit

Moderne Chatbots mit Avataren sind das Ergebnis eines hochkomplexen Zusammenspiels aus generativer KI-Architektur und soziotechnischer Gestaltung sind. Auf technischer Ebene zeigen aktuelle Pipeline-Ansätze, dass leistungsfähige Dialogsysteme weit über reine Textverarbeitung hinausgehen. Durch den Einsatz großer Sprachmodelle, Retrieval-Augmented Generation, Tool-Calling-Mechanismen sowie multimodaler Ausgabeformen, wie TTS oder visuelle Avatare entstehen interaktive Systeme, welche kontextsensitiv, handlungsfähig und in Echtzeit reagieren können. Gleichzeitig stellen Alignment-Strategien, Sicherheitsmechanismen und performante Laufzeitarchitekturen zentrale Voraussetzungen für Zuverlässigkeit, Verständlichkeit und Nutzerakzeptanz dar.

Darüber hinaus wird deutlich, dass der Erfolg eines Chatbots nicht allein von seiner technischen Leistungsfähigkeit abhängt. Die soziotechnische Perspektive zeigt, dass Nutzer Chatbots häufig als soziale Akteure wahrnehmen und auf soziale Signale, bewusst oder unbewusst auch ähnlich reagieren wie in einer zwischenmenschlichen Kommunikation. Elemente wie Persönlichkeit, Sprache, visuelle Gestaltung oder ein Avatar beeinflussen Vertrauen, Zufriedenheit und Interaktionsbereitschaft. Insbesondere visuelle und auditive Signale können positive Effekte erzielen, welche jedoch auch Risiken mit sich bringen im Zusammenhang mit dem Uncanny-Valley-Effekt. Daraus ergibt sich die Notwendigkeit den Grad der Menschenähnlichkeit bewusst zu steuern und die künstliche Natur des Systems transparent zu kommunizieren.

In der Studie [**?**] wurde eine Lernumgebung an einer Universität getestet, bei der 45 Teilnehmende über einen Zeitraum von 14 Tagen mit drei unterschiedlichen Chatbot-Avataren interagieren konnten. Kritisch zu bewerten ist die geringe Stichprobengröße von nur 45 Personen, was die Aussagekraft der Ergebnisse einschränkt. Positiv hervorzuheben ist jedoch die längere Studiendauer, die Einblicke in die Nutzung über einen längeren Zeitraum ermöglicht.

Die Studie [**?**] untersucht die Bedeutung von anthropomorphen versus minimalistischen Chatbot-Interfaces in Expertensystem-Kontexten wie Recht oder HR, indem sie qualitative Interviews mit Fachpersonen kombiniert und Bayesianische Simulationen zur Verallgemeinerung kleiner Stichproben nutzt. Sie zeigt, dass Fachanwender in professionellen Kontexten tendenziell minimalistische, klare Darstellungen bevorzugen, während in eher kreativen oder UX-orientierten Domänen humanähnliche Merkmale positiver bewertet werden. Kritisch ist jedoch, dass die Basis der Interviews mit nur zehn Personen relativ klein ist. Trotz dieser begrenzten Aussagekraft halten wir die Ergebnisse für wertvoll, da sie die Bedeutung der Kontextabhängigkeit bei der Gestaltung von Chatbot-Avataren verdeutlichen.

Die Studie von Thaler et al. [**?**] untersucht Embodied Conversational Agents unter dem Gesichtspunkt des Uncanny Valley. Sechs unterschiedliche ECAs, darunter vier autonome und zwei menschlich gesteuerte Avatare, wurden in kurzen Videoclips präsentiert und von 215 Teilnehmenden hinsichtlich ihrer wahrgenommenen Menschlichkeit, Unheimlichkeit und

Attraktivität bewertet. Die Ergebnisse zeigen, dass menschlich aussehende Avatare als unheimlicher wahrgenommen werden, was den Uncanny Valley Effekt bestätigt. Im Vergleich zu anderen Studien ist es positiv zu bewerten, dass 215 Probanden teilgenommen haben. Kritisch anzumerken ist jedoch, dass den Teilnehmenden lediglich 8-sekündige Videos gezeigt wurden. Dadurch konnte kein fundierter Eindruck entstehen, und es fand auch keine echte Interaktion mit einem Chatbot statt.

Die Analyse altersabhängiger Designpräferenzen ergänzt die Erkenntnisse der soziotechnischen Gestaltung um eine differenzierte Nutzerperspektive. Die betrachteten Studien zeigen, dass jüngere Nutzergruppen deutlich heterogenere Erwartungen an Chatbots und Avatare haben, während ältere Erwachsene eher konsistente Präferenzmuster aufweisen, etwa im Hinblick auf respektvolle Kommunikation, Unterstützung und Klarheit. Daraus lassen sich konkrete Implikationen für die Gestaltung ableiten: Empathie, Transparenz und geringe Interaktionsbarrieren sind über alle Altersgruppen hinweg relevant. Hingegen sind die Expressivität, Geschwindigkeit und spielerische Elemente je nach Zielgruppe unterschiedlich wahrgenommen worden. Insbesondere adaptive Avatar-Konzepte erscheinen geeignet, um die doch sehr unterschiedlichen Erwartungen jüngerer Nutzer abzudecken, ohne dabei ältere Nutzer zu überfordern.

Zusammenfassend lässt sich festhalten, dass die Entwicklung erfolgreicher Chatbot-Avatare eine allgemeine Betrachtung erfordert, welche technologische Innovationen mit sozialwissenschaftlichen Erkenntnissen verbindet. Zukünftige Systeme sollten daher nicht nur auf maximale funktionale Leistungsfähigkeit ausgelegt sein, sondern flexibel, nutzerzentriert und kontextsensitiv gestaltet werden. Gerade im Hinblick auf demografisch vielfältige Nutzergruppen kommt adaptiven, transparenten und empathischen Avatar-Designs eine Schlüsselrolle zu, um nachhaltige Akzeptanz und eine positive Mensch-KI-Interaktion zu gewährleisten.

Daher sollte es zukünftig auch eine Art Richtlinie geben, welche zumindest die Richtung in Anbetracht der unterschiedlichen Nutzergruppen vorgeben sollte. Da es zum heutigen Zeitpunkt keine verbindlichne Vorgaben für Chatbot-Avatare hinsichtlich ihrer soziotechnischen Gestaltung oder deren Anpassung an alters- und kontextspezifische Präferenzen gibt. Solche Richtlinien könnten als gestalterischer Orientierungsrahmen dienen, welcher technische Möglichkeiten, soziale Signalwirkungen und ethische Aspekte miteinander verbindet, ohne dabei die notwendige Flexibilität für unterschiedliche Anwendungsszenarien einzuschränken. Insbesondere im Hinblick auf visuelle und soziale Signale, Transparenz über die künstliche Natur des Systems sowie adaptive Gestaltungsstrategien könnten einheitliche Empfehlungen dazu beitragen, Fehlwahrnehmungen, Überforderung oder Vertrauensverlust bei Nutzern zu vermeiden. Gleichzeitig würden derartige Leitlinien Entwicklern und Gestaltern eine sehr gute Grundlage bieten, um Chatbot-Avatare zielgruppenorientiert, verantwortungsvoll und nutzerzentriert zu entwickeln.

Literatur

[1] Saleema Amershi, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, Eric Horvitz, Daniel Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, and Paul N. Bennett. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13. doi:10.1145/3290605.3300233

[2] Alice Baird, Emilia Parada-Cabaleiro, Simone Hantke, Felix Burkhardt, Nicholas Cummins, and Björn W. Schuller. 2018. The Perception and Analysis of the Likeability and Human Likeness of Synthesized Speech. In *Proceedings of Interspeech 2018*. ISCA, 2863–2867. doi:10.21437/INTERSPEECH.2018-1093

[3] Justine Cassell. 1999. Embodiment in Conversational Interfaces: Rea. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 520–527. doi:10.1145/302979.303150

[4] Ulrich Gnewuch, Jasper Feine, Stefan Morana, and Alexander Maedche. 2020. *Soziotechnische Gestaltung von Chatbots*. Springer Fachmedien Wiesbaden, Wiesbaden, 169–189. doi:10.1007/978-3-658-27941-7_7

[5] Anmol Gulati, James Qin, Chung-Cheng Chiu, Niki Parmar, Yu Zhang, Jiahui Yu, Wei Han, Shibo Wang, Zhengdong Zhang, Yonghui Wu, and Ruoming Pang. 2020. Conformer: Convolution-augmented Transformer for Speech Recognition. In *Proc. Interspeech*. https://arxiv.org/abs/2005.08100

[6] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Vladimir Karpukhin, Naman Goyal, Heinrich Kučer, Sewon Min, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP. In *Advances in Neural Information Processing Systems (NeurIPS)*. https://arxiv.org/abs/2005.11401

[7] Microsoft Azure AI Speech. 2025. *Get facial position with viseme — Speech service*. https://learn.microsoft.com/en-us/azure/ai-services/speech-service/how-to-speech-synthesis-viseme Viseme-Events und Zeitmarken für Lippen-/Gesichtssynchronisation.

[8] Tobias Schneeberger, Florian Müller, and Wolfgang Minker. 2019. Impact of Virtual Environment Design on the Assessment of Virtual Agents. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents (IVA '19)*. ACM, 133–140. doi:10.1145/3308532.3329469

[9] Valentin Schwind, Katrin Wolf, and Niels Henze. 2018. Avoiding the Uncanny Valley in Virtual Character Design. *Interactions* 25, 5 (2018), 45–49. doi:10.1145/3236673

[10] Marina Sokolova and Antonio Fernández-Caballero. 2015. A Review on the Role of Color and Light in Affective Computing. *Applied Sciences* 5, 3 (2015), 275–293. doi:10.3390/app5030275

[11] Chek Tien Tan, Indriyati Atmosukarto, Budianto Tandianus, Songjia Shen, Steven Wong, et al. 2025. Exploring the Impact of Avatar Representations in AI Chatbot Tutors on Learning Experiences. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3706598.3713456

[12] Xu Tan, Tao Qin, Frank Soong, and Tie-Yan Liu. 2021. A Survey on Neural Speech Synthesis. *arXiv preprint arXiv:2106.15561* (2021). https://arxiv.org/abs/2106.15561

[13] Shunyu Yao, Jeffrey Yang, Nan Cui, Karthik Narayanan, Dragomir Radev, et al. 2022. ReAct: Synergizing Reasoning and Acting in Language Models. In *Proceedings of the 2023 International Conference on Machine Learning (ICML) Workshop*. https://arxiv.org/abs/2210.03629

[14] Jingyi Yuan, Xixian Peng, Yichen Liu, and Qiuzhen Wang. 2025. Don't look at me!: The role of avatars' presentation style and gaze direction in social chatbot design. *Computers in Human Behavior* 164 (2025), 108501. doi:10.1016/j.chb.2024.108501