

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 gehören heute zum Alltag und sind kaum noch wegzudenken. Dabei ist nicht mehr nur wichtig, wie gut sie technisch funktionieren. Viel wichtiger ist, wie sie sich gegenüber dem Menschen verhalten und wie sie aussehen, zum Beispiel durch Avatare. Diese Arbeit gibt eine gute Übersicht darüber, welche Merkmale wichtig sind, damit Nutzer einem Chatbot vertrauen und gerne mit ihm arbeiten.

Es wird erläutert, warum kleine soziale Hinweise so wichtig sind und wie Avatare gestaltet sein sollten. Dabei wird auch darauf geachtet, ab wann ein zu menschliches Aussehen eher unheimliche Gefühle auslöst. Außerdem wird betrachtet, welche Rolle die Stimme, die Position auf dem Bildschirm und die Farben im Hintergrund spielen. Ein weiterer wichtiger Punkt sind die verschiedenen Vorlieben von jüngeren und älteren Menschen. Das Ziel der Arbeit ist es zu zeigen, wie man Chatbots so gestaltet, dass sie für den Menschen natürlich wirken und gut zu bedienen sind.

Keywords

Chatbots, Avatare, Soziotechnische Gestaltung, Interface-Design, Uncanny Valley, Nutzerpräferenzen

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 Intelligenz 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.

Während frühe Chatbots hauptsächlich als textbasierte Dialogsysteme funktionierten, erlauben moderne Technologien die Integration von Sprache, grafischen 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? Die Frage, wie sich Präferenzen je nach

Altersgruppe unterscheiden nimmt zunehmend Relevanz, da Chatbots für verschiedene Zielgruppen eingesetzt werden.

Das vorliegende Paper gibt einen systematischen Überblick über theoretische Grundlagen und wesentliche Gestaltungsfaktoren von Chatbots. Es untersucht die sozialen und visuellen Elemente, die das Nutzererlebnis prägen und hebt bestehende Ansätze hervor.

Ziel ist es ein grundlegendes Verständnis dafür zu schaffen, wie Chatbots so gestaltet werden können, dass sie sowohl funktional als auch authentisch für den Menschen wirken. Die daraus gewonnen Kenntnisse werden anschließend in einer Evaluierung zusammengefasst und leiten Schlussfolgerungen für zukünftige Entwicklungen ab.

2 Theoretischer Hintergrund

2.1 Definition und Funktionsweise eines Chatbots

Bei Chatbots handelt es sich um ein Computerprogramm, welche Dialogsysteme mit natürlichsprachlichen Fähigkeiten textueller oder auditiver Art führen. Dabei wird zwischen Chatbots ohne KI, konversationellen und generativen KI-Chatbots, sowie 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-Chatbots 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 Weiterentwicklung von KI-Chatbots, welche neben Deep Learning, um sich im Laufe der Zeit selbst zu verbessern, auch oft mit robotischer Prozessautomatisierung (RPA) in einer einzigen Schnittstelle kombinieren. Dadurch kann ohne weitere zusätzliche Eingriffe eines Menschen direkt auf die Absicht des Nutzers reagiert werden. [1, 4]

Die Funktionsweise von Chatbots war zu Beginn wie einfache FAQ-Systeme mit vordefinierten 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. [1]

2.2 Soziotechnische Gestaltung eines Chatbots

Einige Studien zeigen, dass bei der Entwicklung von Chatbots nicht nur der Einsatz von modernen 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. [7]

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. [7]

“In der zwischenmenschlichen Kommunikation spielen soziale Signale eine große Rolle und sind für den Aufbau und Erhalt erfolgreicher sozialer Beziehungen unerlässlich”, laut Burgoon et al. (2010) [7].

Soziale Signale sind für den Menschen implizite Hinweise 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. [7]

Laut Forschungen von Nass et al. (1994) hat sich gezeigt, dass auch in der Interaktion mit Computern und anderen Technologien soziale Signale unterbewusst wahrgenommen und automatisch darauf reagiert wird [7].

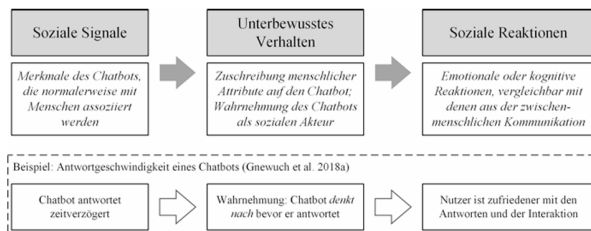


Abbildung 1: soziale Reaktionen [7]

In Abbildung 1: 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 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. [7]

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

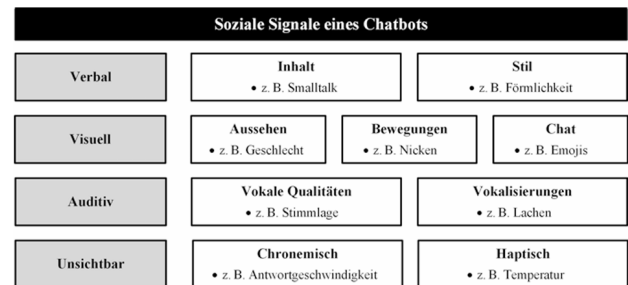


Abbildung 2: Klassifikation sozialer Signale von Chatbots [7]

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 der 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. [7]

2.3 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. [7]

Die Theorie des Uncanny Valley, wie in 3.2.2 beschreiben 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 sowieso Assoziationen zum menschlichem Verhalten erzeugt, gibt es keine allgemeingültige Antwort auf den optimalen Grad an der Ähnlichkeit zum echten Menschen. 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. [7]

3 Gestaltung des Chatbot-Avatars

Die visuelle Gestaltung von Chatbot-Avataren ist weit mehr als das reine Aussehen des Avatars. Sie bestimmt, wie sehr ein System als sozialer Interaktionspartner wahrgenommen wird, welche Erwartungen an seine „Kompetenz“ und „Menschlichkeit“ geknüpft werden und ob die Nutzer sich grundsätzlich wohl damit fühlen, mit ihm zu interagieren. Thaler et al. zeigt, dass bereits relativ kleine Variationen in der visuellen Gestaltung verkörperter Konversationsagenten (Embodied Conversational Agents, ECAs) messbare Effekte auf wahrgenommene Menschlichkeit, Unheimlichkeit und Attraktivität haben. [18]

Rashik et al. kategorisiert das Design von Avataren in Dimensionen, wie Aussehen, Geschlecht, Bewegung, Interaktion, Dimensionalität, Eingabemodalität, Gesprächstyp, Gerätetyp, Anzahl der Agenten und Anwendung [13] ein. 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 untersucht, richtet Karimova den Blick stärker auf Kontextabhängigkeiten. Im Design von Experten-Chatbots können hochgradig anthropomorphe (3.2) Gestaltungen bei vielen Nutzenden eher Misstrauen als Vertrauen hervorrufen, da sie als ablenkendes „psychologisches Decoy“ wahrgenommen werden [9]. Insbesondere in Bereichen wie Rechts- oder Personalwesen müssen Transparenz, Geschwindigkeit und die klare Erkennbarkeit als nicht menschliche 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. [13] beschreibt die Dimension Appearance als das äußere Erscheinungsbild eines Conversational Agents und unterscheidet 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 Arai [8], die verschiedene konkrete Avatar-Designstile für ein Chatbot-System miteinander vergleichen. Untersucht wurden 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 [18].

In der Gestaltung eines Chatbots wird Anthropomorphismus oft als Designstrategie eingesetzt, um das Vertrauen und die soziale Präsenz eines Chatbots zu erhöhen. Cao et al. beschreibt, dass anthropomorphe Avatare eine wichtige Rolle für den Ersteindruck der Benutzer spielen. Insbesondere menschlich und cartoonhaft dargestellte Chatbots haben einen positiven Einfluss auf den Ersteindruck [5].

Gleichzeitig wird davon gewarnt den Anthropomorphismus ausschließlich als Vorteil zu sehen. Dies wäre ein Missverständnis, da Thaler et al. empirisch zeigt, dass mit zunehmender wahrgenommener Menschlichkeit auch das Empfinden von Unheimlichkeit steigt und sich dadurch der bekannte Uncanny Valley Effekt abzeichnet [18]. Hadjiev et al. kommt 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 [8].

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

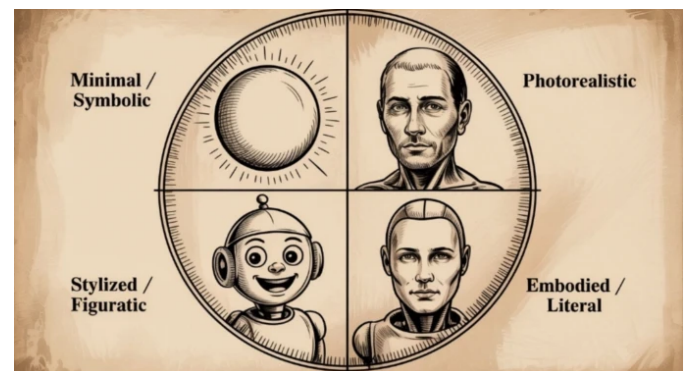


Abbildung 3: visuelle Avatar Typen [9]

Minimalistische Avatare Die symbolischen Darstellungen eines Avatars stehen am unteren Ende der Skala. Diese zeigen keine 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 [9].

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 [8]. Neuere Studien zeigen jedoch, dass innerhalb dieser Kategorie ein wichtiger Unterschied zwischen menschlichen Avataren und Avataren mit tierischen Charakteristiken besteht. Menschliche und cartoonhafte Figuren werden deutlich besser bewertet als Avatare mit tierischen Charakteristiken in Bezug auf wahrgenommene Wärme, Kompetenz, soziale Präsenz und Vertrauen [5].

Fotorealistische Avatare Fotorealistische oder High-Fidelity-Avatare stellen detaillierte menschliche Gesichter dar. Diese Avatare erzeugen eine deutlich höhere soziale Präsenz und Nähe zum menschlichen Erscheinungsbild [9]. Die hohe Detailtiefe und realistische Mimik steigert Vertrauen und Glaubwürdigkeit, kann aber gleichzeitig zu Irritationen führen, wenn Verhalten, Gestik oder Reaktionen nicht konsistent sind [17, 18]. Fotorealistische Avatare befinden sich somit in einem Bereich, in dem das Risiko des Uncanny Valley Effekts 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 [9]. Gleichzeitig steigt die Anfälligkeit für Uncanny Valley Effekte, wenn Animation oder Verhalten nicht mit der visuellen Glaubwürdigkeit übereinstimmen [17, 18].

3.2.2 Der Uncanny Valley Effekt. Die Uncanny Valley Hypothese von Mori et al [11] 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 “uncanny valley”, in dem kleinste Abweichungen im Aussehen oder Verhalten als besonders irritierend, unheimlich oder sogar abstoßend erlebt werden.

Für Chatbot-Avatare ist die Menschenähnlichkeit deshalb relevant, weil steigender Anthropomorphismus zunächst positive Effekte auf Sympathie, Wahrnehmung 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 [9].

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

4 Stimmen von Chatbots

Bei der Entwicklung künstlicher Stimmen stellt sich die Frage, wie menschlich sie klingen sollten, damit Nutzer sie akzeptieren. Stimmen sind bei Chatbots nicht nur dazu da, Informationen weiterzugeben. Sie vermitteln auch soziale Eindrücke wie Freundlichkeit, Kompetenz oder Empathie. Während man beim Aussehen von Avataren oft vom „Uncanny Valley“ (siehe Abschnitt 3.2.2) spricht, ist die Lage bei Stimmen weniger eindeutig.

Baird et al. haben in einer Studie untersucht, ob es diesen Effekt auch bei künstlicher Sprache gibt. 25 Teilnehmende bewerteten dabei, wie menschlich 13 verschiedene synthetische Männerstimmen wirkten und wie gut sie ihnen gefielen. Zum Vergleich wurde auch eine echte menschliche Stimme bewertet. Die Ergebnisse zeigen: Je menschlicher eine Stimme wirkte, desto beliebter war sie. Einen plötzlichen Abfall der Beliebtheit bei fast menschlichen Stimmen, wie man ihn von Grafiken kennt, gab es in dieser Studie nicht. [3]

Trotzdem betonen die Autoren, dass man diese Ergebnisse nicht blind verallgemeinern sollte. Wie wir eine Stimme wahrnehmen, hängt stark von den gezeigten Beispielen, dem Einsatzort und den Erwartungen der Nutzer ab. Besonders in sachlichen oder fachlichen Bereichen können Klarheit und Vertrauenswürdigkeit wichtiger sein als eine perfekte Natürlichkeit. [3]

4.1 Wie Menschlichkeit und Beliebtheit zusammenhängen

Die Studie von Baird et al. zeigt eine deutliche Verbindung: Je natürlicher die Stimmen klangen, desto besser kamen sie bei den Testpersonen an. Stimmen, die menschlicher wirkten, wurden oft als angenehmer, freundlicher und glaubwürdiger eingeschätzt. Für das Design von Chatbots bedeutet das, dass eine natürliche Sprachausgabe die Akzeptanz verbessern kann. [3]

Gleichzeitig muss man bedenken, dass eine sehr menschliche Stimme hohe Erwartungen weckt. Nutzer glauben dann oft, dass das System auch beim Antworten genauso schlau und schnell wie ein Mensch reagiert. Wenn die Technik das nicht halten kann, sind die Nutzer enttäuscht. Deshalb sollte die Stimme immer dazu passen, was das System tatsächlich leisten kann [10].

4.2 Fortschritt der Technik

Baird et al. verglichen Stimmen aus verschiedenen Phasen der Technikentwicklung, angefangen bei älteren Methoden bis hin zu modernen, computergestützten Verfahren. Dabei kam heraus, dass neuere Systeme meist als menschlicher und angenehmer bewertet wurden. [3]

Die Autoren weisen darauf hin, dass selbst die besten künstlichen Stimmen immer noch gut von der echten menschlichen Stimme zu unterscheiden waren. Die Studie zeigt also nicht, dass künstliche Stimmen schon perfekt sind. Sie macht vielmehr deutlich, dass sich moderne Systeme Schritt für Schritt den Merkmalen menschlicher Sprache annähern. [3]

4.3 Was eine Stimme natürlich klingen lässt

Neben den subjektiven Meinungen untersuchte die Studie auch technische Details der Stimmen. Ein wichtiger Punkt war die Tonhöhe. Stimmen, bei denen sich die Tonhöhe kaum änderte, klangen monoton und wurden als weniger menschlich und weniger angenehm empfunden. Mehr Abwechslung in der Tonhöhe sorgte dagegen für bessere Bewertungen. [3]

Zusätzlich fiel auf, dass modernere Systeme oft Wörter am Satzende ein wenig dehnen (das nennt man in der Fachsprache "phrase-final lengthening"). Solche Feinheiten helfen dabei, dass künstliche Sprache weniger roboterhaft wirkt. Da diese Analyse nur auf wenigen Beispielen beruht, sollte man sie eher als ersten Hinweis und nicht als endgültige Regel verstehen. [3]

Insgesamt scheint die Wahrnehmung von Stimmen weniger anfällig für den Uncanny Valley Effekt zu sein als das Aussehen von Avataren (siehe Abschnitt 3.2.2). Für Chatbots heißt das: Eine hochwertige Sprachausgabe lohnt sich, um die Zufriedenheit der Nutzer zu steigern. Wichtig bleibt aber, dass die Stimme zum Gesprächsinhalt und zum Kontext passt. [3]

5 Positionierung des Avatars im Interface

Die Stelle, an der ein Avatar im Bildschirmfenster (Interface) platziert wird, beeinflusst stark, wie wir ihn wahrnehmen und wie gut wir uns auf das Gespräch konzentrieren können. Avatare senden nicht nur gesprochene Worte, sondern auch körpersprachliche Zeichen wie Blickrichtung, Gesichtsausdruck oder Körperhaltung. Wenn zu viele Dinge gleichzeitig auf dem Bildschirm passieren, besteht die Gefahr, dass Nutzer überfordert sind. Ein wichtiges Ziel bei der Gestaltung ist es daher, Ablenkungen zu vermeiden und eine klare Ordnung zwischen dem Text, den Knöpfen und dem Avatar zu schaffen (siehe Abbildung 4).

Frühere Studien zu digitalen Avataren zeigen, dass die Darstellung einer Figur vor allem den Ablauf des Gesprächs unterstützt. Sie hilft zum Beispiel dabei, die Aufmerksamkeit zu lenken oder anzuzeigen, wer gerade an der Reihe ist. Der Avatar dient also eher als Unterstützung, während der eigentliche Inhalt, also der geschriebene Text, im Mittelpunkt stehen sollte. Für das Design bedeutet das, dass Avatare zwar nützlich sind, aber nicht unbedingt mitten im Bild oder sehr groß dargestellt werden müssen, wenn gleichzeitig viele Informationen verarbeitet werden sollen. [6]

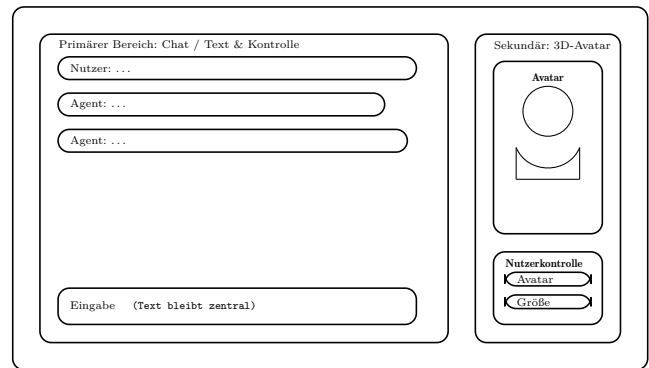


Abbildung 4: Konzeptionelles Layout eines Web-Interfaces mit 3D-Conversational-Avatar: Der Chatbereich bildet den primären Informations- und Kontrollbereich, der Avatar ist seitlich als sekundäres Element platziert. Zur Reduktion von Ablenkung und sozialem Druck wird eine zurückhaltende Blickrichtung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

5.1 Avatarplatzierung und Ordnung im Interface

Weil der Avatar eine unterstützende Rolle hat, sollte er so platziert werden, dass der Text und der Verlauf des Gesprächs klar im Vordergrund stehen. Eine Positionierung am Rand ermöglicht es den Nutzern, die Signale des Avatars wahrzunehmen, ohne den Blick ständig vom Text abwenden zu müssen. Diese klare Ordnung erleichtert die Informationsaufnahme und sorgt dafür, dass der Avatar das Gespräch ergänzt, anstatt davon abzulenken. [6]

Untersuchungen zeigen außerdem, dass Nutzer Avatare unterschiedlich verwenden. Einige wünschen sich einen schnellen Zugriff auf den Text (aufgabenorientiert), während andere den Avatar als motivierendes Element schätzen (interaktionsorientiert). Daher ist es sinnvoll, den Avatar so einzubauen, dass er für interessierte Personen da ist, aber andere nicht beim Lesen behindert. Eine Platzierung in einer Seitenspalte ist hierfür eine gute Lösung, um beiden Gruppen gerecht zu werden [17].

Auch die Blickrichtung des Avatars spielt eine wichtige Rolle. Studien belegen, dass der Blick eines Avatars die Reaktionen der Nutzer beeinflusst. Ein ständiger, direkter Blickkontakt kann als zu intensiv empfunden werden und sollte daher vorsichtig eingesetzt werden. Wenn ein Avatar zusätzlich sehr zentral und auffällig platziert ist, kann er aufdringlich wirken oder unheimlich erscheinen (siehe Abschnitt 3.2.2 sowie Abbildung 5). Eine Positionierung am Rand und ein zurückhaltender Blick machen das Gespräch angenehmer und weniger anstrengend [19].

Auf Basis dieser Erkenntnisse wurde das Layout so gewählt, dass der Avatar eher im Hintergrund steht. Diese Entscheidung sorgt dafür, dass der Avatar zwar präsent ist, aber das eigentliche Gespräch nicht überlagert (siehe Abbildung 4).

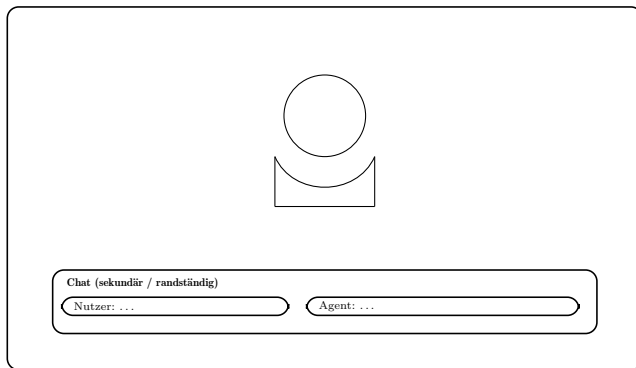


Abbildung 5: Beispiel für ein ungünstiges Layout eines avatarbasierten Conversational-Interfaces: Der Avatar ist zentral und visuell dominant platziert und richtet seinen Blick direkt auf die Nutzenden. Der Chatverlauf ist randständig angeordnet und untergeordnet. Diese Gestaltung kann zu erhöhter Ablenkung,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

5.2 Kontrolle durch den Nutzer

Damit die Zusammenarbeit mit einem Chatbot als angenehm empfunden wird, ist es wichtig, dass der Nutzer selbst entscheiden kann, wie das Programm aussieht. Fachleute empfehlen, den Nutzern die Kontrolle über das Verhalten und das Aussehen von KI-Systemen zu geben. Das gilt auch für Avatare. [2]

Konkret bedeutet das: Nutzer sollten selbst bestimmen können, wie sichtbar oder auffällig der Avatar ist. Wenn man ihn verkleinern oder kurzzeitig ausblenden kann, hilft das dabei, sich voll auf den Text zu konzentrieren. Gleichzeitig haben Nutzer die Freiheit, den Avatar wieder stärker einzublenden, wenn sie das möchten. Eine solche flexible Gestaltung hilft in verschiedenen Situationen und sorgt dafür, dass der Chatbot besser akzeptiert wird. [2]

5.3 Farbgestaltung

Farben im Interface können helfen, die Interaktion mit einem Chatbot verständlicher zu machen. Es geht dabei nicht darum, dass Farben Gefühle ersetzen. Sie wirken eher als zusätzliches Signal im Hintergrund.

Ein bekanntes Problem bei sehr menschlich aussehenden Avataren ist der Uncanny Valley Effekt. Schwind et al. beschreiben, dass negative Reaktionen entstehen können, wenn verschiedene Eindrücke nicht zusammenpassen. Das passiert zum Beispiel, wenn ein Avatar sehr echt aussieht, aber die Bewegung oder der Ausdruck nicht dazu passt. Solche Widersprüche fallen bei Bewegung noch stärker auf. [15]

Wichtig ist dabei: Nicht nur der Avatar selbst beeinflusst die Wahrnehmung, sondern auch die Umgebung, in der er gezeigt wird. Schneeberger et al. haben das mit einem virtuellen Interviewer untersucht. Der Interviewer verhielt sich immer gleich, aber Hintergrund und Licht wurden verändert. Die Ergebnisse zeigen, dass solche Änderungen die Bewertung

des Avatars und der gesamten Interaktion beeinflussen können. Daraus folgt, dass man Farben, Helligkeit und andere Elemente im Interface als Teil des Gesamteindrucks sehen sollte. [14]

Vor diesem Hintergrund kann es sinnvoll sein, Gefühle und soziale Signale nicht nur über den Gesichtsausdruck oder Bewegungen des Avatars zu zeigen, sondern auch über das Interface. Farben können dabei als Feedback im Hintergrund dienen. Sie geben einen Hinweis auf den Zusammenhang (zum Beispiel sachlich, unterstützend oder Warnung), ohne dass dafür aufwendige Animationen nötig sind. In der Forschung wird diskutiert, dass Licht und Farbe die Wahrnehmung von Stimmungen beeinflussen können. [16]

Daraus lassen sich folgende einfache Ansätze für die Gestaltung ableiten:

- **Unterstützung des Zusammenhangs:** Hintergrund oder Akzentfarben können zeigen, welchen Charakter eine Antwort hat. Kühle Farben passen zu sachlichen Informationen, wärmere Farben zu unterstützenden Antworten. Wichtig ist, dass die Farbänderungen dezent bleiben, damit sie nicht ablenken.
- **Steuerung der Aufmerksamkeit:** 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 Nutzer weniger auf mögliche Unstimmigkeiten im Ausdruck des Avatars achten.

Diese Punkte sind Gestaltungsvorschläge. Sie sind nicht als Beweis zu verstehen, dass Farbe den Uncanny Valley Effekt sicher verhindert. Sie sind eher eine logische Ableitung aus zwei Erkenntnissen: Erstens, dass das Problem oft durch widersprüchliche Eindrücke bei Avataren entsteht. Zweitens, dass auch die Umgebung und die visuelle Gestaltung die Bewertung von Avataren beeinflussen können. [14, 15]

6 Präferenzen von Avataren in unterschiedlichen Altersgruppen

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

6.1 Literatur zu altersabhängigen Designpräferenzen

Nurhas et al. (2022) untersuchte mittels Q-Methodologie subjektive Designpräferenzen für Chatbots in intergenerationalen Innovationssettings. In der Studie gab es 11 Teilnehmer im Alter von 18-23 Jahren, welche Vertreter der jüngeren Erwachsenen waren, bestehend aus Studenten und 13 Teilnehmer mit über 55 Jahren, welche regelmäßig das Internet und Online-Medien nutzten und die älteren Erwachsenen vertreten. Es wurde die Q-Methodologie eingesetzt, in der es vorgegebene Aussagen zu Chatbot-Designs relativ zueinander auf einer Zustimmungsskala sortiert werden mussten, um typische Meinungsprofile zu identifizieren, wodurch sich vier typische Nutzer-Typen mit unterschiedlichen Prioritäten herausgestellt haben. [12]

- (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. [12]
- (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. [12]
- (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. [12]
- (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. [12]

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. [12]
- **Immediacy / Reaktionsverfügbarkeit:** Schnelle Verfügbarkeit und unmittelbare Rückmeldung wurden von Benutzern erwünscht. Speziell für einen Avatar wäre dies die Interaktionslatenz, sowie die Präsenzgestaltung. [12]
- **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. [12]
- **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. [12]

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. [12]

7 Fazit

Moderne Chatbots mit Avataren sind das Ergebnis des Zusammenspiels aus generativer KI-Architektur und soziotechnischer Gestaltung, dabei hängt der Erfolg eines Chatbots nicht allein von seiner technischen Leistungsfähigkeit ab. 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 [17] 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 [9] untersucht die Bedeutung von anthropomorphen versus minimalistischen Chatbot-Interfaces in Experten-System-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. [18] 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.

In Bezug auf die Sprachausgabe verdeutlichen die Ergebnisse, dass künstliche Stimmen eine wichtige soziale Funktion übernehmen und Eigenschaften wie Freundlichkeit oder Kompetenz transportieren. Interessanterweise scheint die Akzeptanz bei Stimmen mit steigender Natürlichkeit zuzunehmen, ohne die typischen negativen Reaktionen auszulösen, die oft bei fast menschlichen visuellen Darstellungen auftreten. Allerdings führt eine sehr menschliche Aussprache auch zu höheren Erwartungen an die Intelligenz des Chatbots. Es ist daher entscheidend, dass der Klang der Stimme zum tatsächlichen Funktionsumfang des Systems passt, damit keine falschen Versprechungen gemacht werden und die Nutzerzufriedenheit stabil bleibt.

Hinsichtlich der Platzierung im Interface zeigt sich, dass die räumliche Anordnung des Avatars die Aufmerksamkeit der Nutzer stark lenkt. Eine Positionierung am Rand des Bildschirms wird meist als angenehmer empfunden, da der Avatar so als Begleiter fungiert, ohne vom Text oder den Bedienelementen abzulenken. Zudem spielt die Blickrichtung eine Rolle, da ein zu intensiver, ständiger Blickkontakt als aufdringlich wahrgenommen werden kann. Indem man den Nutzern die Kontrolle über die Größe und Sichtbarkeit des Avatars überlässt, können diese die Interaktion nach ihren eigenen Bedürfnissen gestalten. Zusammen mit einer passenden farblichen Gestaltung der Umgebung wird so ein harmonisches Gesamtbild erzeugt, das die soziale Bindung stärkt, ohne die Bedienbarkeit zu stören.

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 wirklichen 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] 2025. *What is a chatbot?* <https://www.ibm.com/think/topics/chatbots>
- [2] Saleema Amershi, Daniel S. Weld, Mihaela Vorvoreanu, Adam Fournay, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ron Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3290605.3300233
- [3] 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
- [4] Oliver Bendel. 2025. *Definition: Was ist "Chatbot"?* <https://wirtschaftslexikon.gabler.de/definition/chatbot-54248>
- [5] Yaqin Cao, Yu Liu, Wei Lyu, and Ming Li. 2025. Human or AI Customer Service? The Role of Anthropomorphic Avatars in Shaping First Impressions of Chatbots. *Human Factors and Ergonomics in Manufacturing & Service Industries* 35, 5 (2025), e70024. doi:10.1002/hfm.70024
- [6] 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
- [7] Ulrich Gnewuch, Jasper Feine, Stefan Morana, and Alexander Maedche. 2020. *Soziotechnische Gestaltung von Chatbots*. Springer Fachmedien Wiesbaden, 169–189. doi:10.1007/978-3-658-27941-7_7
- [8] Anna Maria Hadjiev and Kenji Araki. 2021. Evaluation of various avatar designs for conversational chatbot systems. In *Japanese Society for Artificial Intelligence, Type 2 Workshop: Language Engineering*, Vol. 66. 5–10.
- [9] Gulnara Z Karimova. 2025. Not in our image: rethinking anthropomorphism in expert chatbot design. *AI & SOCIETY* (2025), 1–18.
- [10] Ewa Luger and Abigail Sellen. 2016. "Like having a really bad PA: The Gulf Between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5286–5297. doi:10.1145/2858036.2858288

- [11] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & automation magazine* 19, 2 (2012), 98–100.
- [12] Irawan Nurhas, Pouyan Jahanbin, Jan Pawlowski, Stephen Win-green, and Stefan Geisler. 2022. *Patterns of Sociotechnical Design Preferences of Chatbots for Intergenerational Collaborative Innovation : A Q Methodology Study*. <https://arxiv.org/pdf/2212.03485>
- [13] Mashrur Rashik, Mahmood Jasim, Kostiantyn Kucher, Ali Sarv-ghad, and Narges Mahyar. 2024. Beyond Text and Speech in Conversational Agents: Mapping the Design Space of Avatars. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, 1875–1894. doi:10.1145/3643834.3661563
- [14] 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
- [15] Valentin Schwind, Katrin Wolf, Niels Henze, and Oliver Korn. 2018. Avoiding the Uncanny Valley in Virtual Character Design. *ACM Transactions on Applied Perception (TAP)* 15, 4 (2018).
- [16] 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
- [17] Chek Tien Tan, Indriyati Atmosukarto, Budianto Tandianus, Song-jia Shen, and Steven Wong. 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
- [18] Markus Thaler, Stephan Schlögl, and Aleksander Groth. 2020. Agent vs. Avatar: Comparing Embodied Conversational Agents Concerning Characteristics of the Uncanny Valley. In *2020 IEEE International Conference on Human-Machine Systems (ICH-MS)*. 1–6. doi:10.1109/ICHMS49158.2020.9209539
- [19] 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