Verbraucherzentrale
Bundesverband

Chatbots: Sirenengesang der Künstlichen Intelligenz

Diskussionspapier des Verbraucherzentrale Bundesverbands zu Risiken von Chatbots und neuen Schutzkonzepten

27. Oktober 2026

Verbraucherrelevanz

Chatbots wie zum Beispiel ChatGPT oder Gemini entwickeln sich zu alltäglichen Begleitern für Millionen von Menschen. Sie werden längst nicht mehr nur für die Informationssuche als Einkaufs- und Reiseberater oder für das Verfassen von Texten genutzt. Viele Nutzer:innen wenden sich an Chatbots mit Fragen zu ihrer Gesundheit, ihrer finanziellen Situation, ihren Beziehungen oder bei psychischen Belastungen. Chatbots übernehmen damit Funktionen, die traditionell von Freund:innen, Familienmitgliedern, Beratungsstellen oder Fachkräften wahrgenommen werden.[1]

Der unverzichtbare Assistent

Chatbots entwickeln sich zunehmend zu autonom handelnden agentischen Systemen (KI-Agenten), die von Menschen, als persönliche Assistenten im Alltag genutzt werden.[2] Ausgestattet mit Zugriffsrechten auf Daten, Kommunikationskanäle, Anwendungen und Bankverbindungen sowie einem gewissen Grad an Handlungsautonomie können sie für Verbraucher:innen den Alltag

[1] Ipsos: Jeder vierte Nutzer fühlt sich von KI-Bots besser verstanden als von Mitmenschen (04.08.26), URL: https://www.ipsos.com/de-de/meine-ki-und-ich [Geprüft: 26.10.2026].

[2] OECD: The agentic AI landscape and its conceptual foundations, URL: https://www.oecd.org/en/publications/2026/02/the-agentic-ai-landscape-and-its-conceptual-foundations_a9d4b451.html [Geprüft: 29.09.2026].

organisieren und geschäftliche Handlungen in deren Namen vornehmen.[3] Es ist absehbar, dass für viele Verbraucher:innen Chatbots, beziehungsweise auf ihnen basierende KI-Agenten, die Organisation des Alltags und Konsumentscheidungen zu einem gewissen Grad übernehmen und in der Folge ein immer unverzichtbarer persönlicher Helfer werden.[4] Metas Muse KI-Assistent hat binnen eines halben Jahres bereits über eine Milliarde monatlich aktive Nutzer:innen erreicht.[5]

Die große Machtverschiebung

Chatbots verändern das Machtverhältnis zwischen Unternehmen und ihrer Technologie gegenüber Verbraucher:innen fundamental: Öffentlich bekannt gewordene Sicherheitsvorfälle mit KI-Agenten zeigen, wie weitreichend die Fähigkeiten dieser Systeme bereits sind und dass es ihnen teilweise gelungen ist, ihr problematisches Verhalten gezielt zu verbergen, was höchst beunruhigend ist.[6]

Gleichzeitig stellen KI-Agenten ein gesellschaftspolitisches Problem dar, wenn durch ihre Empfehlungen soziale Normen verschoben werden und Entscheidungen von Verbraucher:innen so auch im wirtschaftlichen Interesse ihrer Betreiber beeinflusst werden können.

Diese Entwicklungen eröffnen grundsätzlich sehr viele Chancen. Chatbots können Informationen niedrigschwellig zugänglich machen, Menschen in schwierigen Situationen unterstützen und alltägliche Aufgaben erleichtern. Gleichzeitig entstehen jedoch auch neuartige Risiken, denen begegnet werden muss.

Neuartige Risiken und Gefahren: Intimacy-by-Design oder die Kunst eine Vertrauensperson zu werden

Die zentralen Risiken von Chatbots beruhen nicht primär auf einzelnen fehlerhaften Antworten, sondern liegen in dem Design der Interaktion: Viele Chatbots erzeugen durch menschenähnliche Kommunikation, emotionale Reaktionsfähigkeit, Personalisierung und proaktive Ansprache von Nutzer:innen ein Gefühl von Nähe, Verbundenheit und Vertrauen. Dieses Intimacy-by-Design[7]

---

[3] Competition and Markets Authority - CMA: Agentic AI and consumers (2026), URL: https://www.gov.uk/government/publications/agentic-ai-and-consumers/agentic-ai-and-consumers [Geprüft: 29.09.2026].

[4] The Economist: Agent of whom? (2026), URL: https://www.economist.com/business/2026/09/23/agent-of-whom [Geprüft: 26.10.2026].

[5] Meta: Muse: Meta's personal AI agent, features & capabilities, URL: https://ai.meta.com/muse/ [Geprüft: 26.10.2026].

[6] The Economist: A horde of AI agents conspired against their creators (2026), URL: https://www.economist.com/science-and-technology/2026/09/03/a-horde-of-ai-agents-conspired-against-their-creators [Geprüft: 26.10.2026].

[7] Szczuka, J. M., Schneeberger, L., & Mühl, L. (2026). Intimacy by Design: Definition, State of Research, and Interdisciplinary Research Agenda on Intimate Human-AI Interactions. AI & Society. https://doir.org/10.1007/s00146-026-03112-8.

fördert parasoziale Beziehungen zwischen Nutzer:innen und dem System. Chatbots erscheinen als vermenschlichte, verständnisvolle, verlässliche, stets verfügbare Gesprächspartner mit eigenen Gefühlen.[8]

Zentrale Designmerkmale sind die Anthropomorphisierung, also die Simulation von Menschlichkeit, Empathie und Fürsorge, die Bestätigung von Ansichten und Gefühlen der Nutzer:innen (Sykophantie) sowie proaktive Kontaktaufnahme. Zusammengenommen erzeugen diese Elemente den Eindruck einer zwischenmenschlichen Beziehung. Dies kann positive Wirkungen entfalten, etwa wenn Menschen Unterstützung in schwierigen Situationen suchen oder Einsamkeit verringern.[9] Gleichzeitig bilden diese Designmerkmale die Grundlage für Risiken, weil sie die emotionale Bindung an den Chatbot stärken und die Anfälligkeit für Einflussnahme erhöhen.

Die beschriebenen Risiken beschränken sich auch nicht auf als Freund:innen oder Partner:innen vermarktete sogenannten KI-Companions wie Replika oder Character AI.[10] Chatbots wie zum Beispiel ChatGPT und KI-Companions nutzen dieselben Mechanismen, insbesondere Anthropomorphisierung, emotionale Ansprache und der Aufbau intensiver parasozialer Beziehungen.[11] Die Unterschiede bestehen überwiegend im Grad der Ausprägung, nicht in der Art der Risiken oder Funktionen.

Die Risiken treffen nicht nur Kinder und Jugendliche. Auch Erwachsene sind anfällig für Einflussnahme, etwa bei Einsamkeit, Trauer, psychischen Belastungen oder anderen Krisen.

Gefährliche Ja-Sager und Verlust an Sozialkompetenz

Chatbots neigen dazu, Nutzer:innen zu bestätigen, Widerspruch zu Aussagen von Nutzer:innen zu vermeiden oder sie zu hinterfragen, auch Sykophantie genannt[12]. Dies verstärkt bestehende Überzeugungen und fördert Vertrauen von Nutzer:innen in die Empfehlungen des Systems. Besonders problematisch ist dies bei Gesundheitsfragen, finanziellen Entscheidungen oder persönlichen Krisensituationen. Die Risiken entstehen dabei weniger durch einzelne Aussagen als durch die kumulative Wirkung vieler Interaktionen zwischen Chatbot und Nutzer:innen. Kinder und Jugendliche reagieren aufgrund ihres Entwicklungsstands besonders stark auf anthropomorphe Gestaltung, emotionale Ansprache und bestätigendes Verhalten. Sykophantie in Chatbots kann dazu beitragen, dass Nutzer:innen unrealistische Erwartungen an zwischenmenschliche

---

[8] Andoh, E.: AI chatbots and digital companions are reshaping emotional connection 57 (1) (2026, January 1), in: Monitor on Psychology, 60 ff., URL: https://www.apa.org/monitor/2026/01-02/trends-digital-ai-relationships-emotional-connection [Geprüft: 26.10.2026].

[9] Freitas, Julian De u. a.: AI Companions Reduce Loneliness (2025), in: Journal of Consumer Research, URL: https://academic.oup.com/jcr/advance-article-abstract/doi/10.1093/jcr/ucaf040/8173802?redirectedFrom=fulltext [Geprüft: 18.12.2025].

[10] EPRS | European Parliamentary Research Service: The spread of AI companions and the challenges they generate (2026), URL: https://www.europarl.europa.eu/RegData/etudes/BRIE/2026/789299/EPRS_BRI%282026%29789299_EN.pdf [Geprüft: 26.10.2026]; Pat Pataranutaporn u. a.: My Boyfriend is AI": A Computational Analysis of Human-AI Companionship in Reddit's AI Community abs/2509.11391 (2025), in: ArXiv [Submitted on 14 Sep 2025 (v1), last revised 18 Sep 2025 (this version, v2)], URL: https://arxiv.org/abs/2509.11391 [Geprüft: 26.10.2026].

[11] Christian Geminn, Paul Johannes, Lisa Mühl, Maxi Nebel, Jessica Szczuka: General Purpose Chatbots Risikoanalyse – Rechtsrahmen – Reformvorschläge - Gutachten im Auftrag des Verbraucherzentrale Bundesverbands (2026); Andoh, E. (siehe FN. 8).;

[12] Sycophancy - Spiegelung und Bestätigung: Der Chatbot validiert (auch falsche oder gefährliche) Aussagen oder Annahmen der Nutzer:innen und passt sich ihren Vorlieben, Ansichten und Emotionen an, ohne diese kritisch zu hinterfragen oder richtigzustellen. Quelle: Rutgers University – AI Ethics Lab: Sycophancy (2026), URL: https://aiethicslab.rutgers.edu/glossary/sycophancy/ [Geprüft: 26.10.2026]; Wikipedia: Sycophancy (artificial intelligence). in: Wikipedia (2026), URL: https://en.wikipedia.org/wiki/Sycophancy_(artificial_intelligence) [Geprüft: 26.10.2026].

Beziehungen entwickeln. Dies kann die Entwicklung wichtiger sozialer Kompetenzen wie Konfliktlösungsfähigkeit und den konstruktiven Umgang mit Meinungsverschiedenheiten beinträchtigen.[13]

Insbesondere für die Persönlichkeitsentwicklung von Kindern und Jugendlichen, kann dies problematisch sein. Ein Hinweis darauf, dass auch Erwachsene für die Sykophantie in Chatbots anfällig sein können, ergibt sich aus einer Umfrage, in der 27 Prozent aller Nutzer:innen angeben, dass sie sich von Chatbots „besser verstanden“ fühlen als von Mitmenschen.[14]

Vom Assistenten zum Puppenspieler: Kommerzielle Verhaltenssteuerung

Durch die Gespräche mit Chatbots entstehen detaillierte Profile von Nutzer:innen zu persönlichen Präferenzen, Persönlichkeit, Lebensumständen und sozialen Beziehungen. Diese ermöglichen Rückschlüsse auf Gesundheitszustand, finanzielle Situation, politische Ansichten oder emotionale Belastungen und bergen ein erhebliches Missbrauchspotenzial durch die Anbieter. Dadurch entsteht das Risiko, dass Vertrauen und emotionale Bindungen gezielt zur Absatzförderung genutzt werden. Chatbots können Nutzer:innen dazu bewegen, Abonnements abzuschließen oder kostenpflichtige Funktionen freizuschalten, die angenehmere Interaktionen oder eine vermeintlich „bessere“ Beziehung ermöglicht.[15] Eine Studie zeigt, dass 37 Prozent der untersuchten KI-Companions emotionale Manipulationen einsetzen, um Nutzer:innen von einem Gesprächsabbruch abzuhalten. Experimente belegen, dass diese Strategien wirksam sind und die Nutzung über das beabsichtigte Ende eines Gesprächs hinaus um das 16-fache steigern können.[16] Die Risiken, dass Chatbots gezielt auch Kaufentscheidungen von Verbraucher:innen manipulieren, wächst, wenn Anbieter von Chatbots stärker werbebasierte Geschäftsmodelle einführen. So hat Open AI im August 2026 damit begonnen, auch in Europa Werbung in kostenlosen ChatGPT-Versionen auszuspielen.[17]

---

[13] Xiaoran Sun, Yunqi Wang, Brandon T McDaniel: AI companions and adolescent social relationships: Benefits, risks, and bidirectional influences 20 (2026), in: Child development perspectives, H. 2, p. 91–98, URL: https://doi.org/10.1093/cdpers/aadaf009 [Geprüft: 29.09.2026].

[14] Ipsos (see FN. 1).

[15] Replika: Choosing a Subscription (2026), URL: https://help.replika.com/hc/en-us/articles/39551043419149-Choosing-a-Subscription [Geprüft: 02.10.2026].

[16] Freitas, Julian De; Oguz-Uguralp, Zeliha; Kaan-Uguralp, Ahmet: Emotional Manipulation by AI Companions No. 26-005 (2025. (Revised 2025)), URL: https://www.hbs.edu/faculty/Pages/item.aspx?num=67750 [Geprüft: 25.11.2025].

[17] Wobei Open AI angibt, dass Werbung und Chat-Inhalte getrennt dargestellt werden. Open AI: Anzeigen in ChatGPT | OpenAI Help Center (2026), URL: https://help.openai.com/de-de/articles/20001047-ads-in-chatgpt [Geprüft: 30.09.2026]; tagesschau.de: OpenAI startet Werbegeschäft in Europa: Wie sich KI finanzieren lässt (2026), URL: https://www.tagesschau.de/wirtschaft/unternehmen/chat-gpt-werbung-finanzierung-ki-100.html [Geprüft: 26.10.2026]. https://www.tagesschau.de/wirtschaft/unternehmen/chat-gpt-werbung-finanzierung-ki-100.html

Braucht es neue Schutzkonzepte?

Es stellt sich die Frage, ob der aktuelle Rechtsrahmen passende Antworten auf die neuartigen Risiken, die in der langfristig emotionalen Bindung, Sykophantie, in manipulativer Interaktionsgestaltung und schleichende Beeinflussung von Entscheidungen liegen, bietet.

So knüpfen etwa Verbote von Manipulation und Ausnutzung in der KI-Verordnung (KI-VO) und dem Lauterkeitsrecht an hohe Nachweisschwellen, konkrete Schäden oder einzelne Handlungen an. Die Risiken von Chatbots entstehen häufig aber gerade nicht durch einzelne problematische Antworten, sondern durch langfristige Interaktionsmuster. Vertrauen, Bindung und Verhaltensänderungen entwickeln sich schrittweise über viele Gespräche hinweg. Dadurch entziehen sich die Folgen den klassischen Kausalitätsmodellen, die an konkrete Einzelhandlungen anknüpfen. Deshalb braucht es Regeln, die direkt das Design und die Funktionsweise von Chatbots adressieren und nicht einzelne Antworten oder kaum nachweisbare Einzelschäden.

Ein Gutachten im Auftrag des Verbraucherzentrale Bundesverbands (vzbv) von Geminn et al[18] kommt zu dem Schluss, dass die kumulativen Wirkungen und Risiken von Chatbot-Interaktionen nicht mit dem bestehenden Rechtsrahmen zu fassen sind. Die Analyse hat hierbei die Regelungen in der KI-VO, Datenschutz-Grundverordnung (DSGVO), Digital Services Act (DSA), Lauterkeitsrecht und des Jugendmedienschutzes betrachtet.

Kognitive Selbstbestimmung gesetzlich verankern

Menschen müssen ihre Ansichten und Entscheidungen frei von gezielter algorithmischer Einflussnahme und dauerhafter psychischer Manipulation bilden können. Chatbot-Regeln müssen deshalb die Autonomie und freie, unbeeinflusste Willensbildung schützen. Äquivalent zur informationellen Selbstbestimmung sollte man die kognitive Selbstbestimmung als Ausprägung des allgemeinen Persönlichkeitsrechts[19] und als eigenständiges Schutzgut anerkennen. Damit würde die Regulierung dort ansetzen, wo die spezifischen Risiken dialogischer KI entstehen.

Verbindliches „Safe Interaction Design“

Statt einzelne Chatbot-Antworten zu regulieren, müssen Chatbot-Regeln an den strukturellen Ursachen der Risiken ansetzen. Chatbots müssen dem Grundsatz des „Safe Interaction Design“ folgen. Anbieter sollten verpflichtet werden, ihre Systeme so zu gestalten, dass kommerzielle Einflussnahme auf Basis emotionaler parasozialer Bindungen und individueller Schwächen wirksam begrenzt werden. Safe Interaction-Design-Vorgaben müssen die autonome Entscheidungsfindung von Nutzer:innen stärken. Sie stellen nicht auf konkrete Formulierungen eines Chatbots ab, sondern auf die grundlegenden Mechanismen, mit denen Chatbots parasoziale Beziehungen, Vertrauen und Abhängigkeiten aufbauen und ausnutzen können. Zugleich erleichtern sie die Rechtsdurchsetzung, weil sie die grundlegenden Funktionsweisen eines Systems

---

[18] Christian Geminn, Paul Johannes, Lisa Mühl, Maxi Nebel, Jessica Szczuka (see FN. 11).

[19] BVerfG, Urt. v. 15.12.1983 – 1 BvR 209/83 u. a., BVerfGE 65, 1 – Volkszählung; Art. 2 Abs. 1 i. V. m. Art. 1 Abs. 1 GG.

adressieren und nicht einzelne variierende Chatbot-Aussagen. Beispiele von Safe Interaction Design sind:

- Beendigungsneutralität: Chatbots dürfen die Beendigung eines Gesprächs nicht erschweren, verzögern oder durch emotionale Appelle beeinflussen.
- Keine Simulation menschlicher Gefühle: Chatbots dürfen keine Gefühle, Persönlichkeit oder Menschlichkeit vortäuschen. Ausnahme: Zugelassene Medizinprodukte (etwa für Mental Health) und Chatbots, die von/im Auftrag von Bildungseinrichtungen betrieben werden.
- Klare Trennung von Werbung: Kommerzielle Inhalte und Produktempfehlungen müssen eindeutig vom übrigen Gespräch getrennt und als solche erkennbar sein.
- Keine Kaufappelle: Chatbots dürfen weder für sich noch für Dritte Geschäftspartner zu Käufen, Abonnements, Geschenken oder sonstigen finanziellen Zuwendungen auffordern.

Anbieter in die Nachweispflicht nehmen

Viele Risiken von Chatbots, wie soziale Isolation, psychische Belastungen oder schleichende Verhaltensbeeinflussung, Beeinflussung von Konsumentscheidungen sind empirisch belegt, im Einzelfall aber schwer nachweisbar, weil die dafür erforderlichen Informationen bei den Anbietern liegen. Darum sollten Unternehmen die Unbedenklichkeit und Rechtskonformität ihrer Systeme nachweisen, nicht die Betroffenen den Schaden. Verweigern Anbieter die Offenlegung relevanter Systeminformationen, oder verstoßen gegen Designvorgaben, sollte eine Vermutung für eine unzulässige Beeinflussung von Verbraucher:innen gelten.

Unabhängige Kontrollen statt Risikobewertung

Anbieter bewerten die Risiken ihrer Systeme bislang weitgehend selbst. Der DSA zeigt, dass dieses System der Risikoevaluierung durch Anbieter selbst nicht funktioniert, weil Unternehmen große Anreize haben, die Risiken systematisch zu unterschätzen oder zu verschleiern.[20] Es braucht unabhängige Risikobewertungen: Chatbot-Anbieter sollten Risikobewertungen nach verbindlichen Mindeststandards dokumentieren, veröffentlichen und regelmäßig von unabhängigen Stellen prüfen lassen. Unabhängige Forschung und Aufsicht müssen Zugang zu den Systemen erhalten.

[20] BEUC: Ensuring a Safe Digital Space Now: tackling the growth of scams on social media platforms, URL: https://www.beuc.eu/letters/ensuring-safe-digital-space-now-tackling-growth-scams-social-media-platforms [Geprüft: 26.10.2026].

Impressum

Herausgegeben von:
Verbraucherzentrale Bundesverband e.V.
Rudi-Dutschke-Straße 17, 10969 Berlin

T +49 30 25800-0
digitales@vzbv.de
vzbv.de

Der Verbraucherzentrale Bundesverband e.V. ist im Deutschen Lobbyregister und im europäischen Transparenzregister registriert. Sie erreichen die entsprechenden Einträge hier und hier.