

Positionspapier KI-Companions



Swiss Insights

POSITIONSPAPIER

KI-Companions: Vertrauen schafft Verantwortung

Ein Positionspapier zu Design- und Governance-Verantwortung über KI-Companions in Schweizer Unternehmen

Zielgruppe:

Grössere Schweizer Unternehmen und KMU

Mai 2026

Anouk Bergner, Samuel Schütz, & Christoph Bräunlich

Einführung

KI-Companions sind konversationelle KI-Systeme, die über einmalige Transaktionen hinausgehen und auf kontinuierliche, personalisierte Interaktion ausgelegt sind. Diese Companions klingen nicht mehr wie Chatbots der ersten Generation. Konversationelle KI ist heute standardmässig fließend und zugewandt, eine Entwicklung, die durch jüngste Fortschritte in der generativen KI erheblich beschleunigt wurde¹ und die Wahrnehmung, dass KI kalt und emotionslos ist, zusehends in den Hintergrund drängt.² Die relevante Frage für Unternehmen lautet daher nicht mehr, wie menschlich ihre KI klingen soll (sie tut es bereits) sondern welche Verantwortungen aus dieser wahrgenommenen Menschlichkeit folgen. Zentral ist dabei, dass KI-Companions im Unterschied zur menschlichen Kundenberatung, die durch individuelle Interaktion und begrenzte Skalierbarkeit geprägt ist, identische Interaktionsmuster gleichzeitig auf tausende Nutzende ausrollen können. Damit werden Verhaltensweisen, die nicht mit den Werten der Unternehmen für den Kundenkontakt vereinbar sind, nicht mehr zu Einzelfällen, sondern skalieren beispielsweise unmittelbar zu systematischen Reputationsrisiken.

¹ Bergner, Hildebrand & Häubl (2023), Journal of Consumer Research: zeigt, dass dialogische Gestaltungsmerkmale konversationeller KI die wahrgenommene Menschlichkeit des Systems erhöhen und engere Beziehungen zwischen Nutzenden und dem Unternehmen fördern.

² De Freitas, Agarwal, Schmitt & Haslam (2023), Nature Human Behaviour, 7, 1845–1854: identifiziert fünf psychologische Barrieren gegenüber KI-Systemen, darunter Emotionslosigkeit, die Wahrnehmung, dass KI Wärme und Empathie fehlt.

KI-Companions erzeugen Vertrauen

Nutzerinnen und Nutzer interagieren mit Companions anders als mit Suchmaschinen oder klassischen Interfaces. So teilen sie zum Beispiel eher persönlichere Informationen mit KI Companions als mit anderen Menschen, weil sie sich weniger stark beurteilt fühlen.³ Je menschlicher ein Companion wirkt, desto eher schreiben Nutzende ihm Absichten und Gefühle zu und desto stärker wird die Erwartung, dass er sie versteht und sich an sie erinnert.⁴ Diese Zuschreibung führt zu Vertrauen, Vertrauen wiederum reduziert die kritische Distanz gegenüber den Inhalten. Weiter entsteht daraus mit der Zeit etwas, das funktional wie eine auf Vertrauen-aufgebaute quasi-zwischenmenschliche Beziehung wirkt. Genau das macht Companions wirksam und folgenreich.

Zwei Verantwortungsbereiche

Informationelle Verantwortung. Vertrauen verringert den Impuls zur Überprüfung. Ein Suchergebnis wird hinterfragt, was ein vertrauter Companion sagt, deutlich seltener. Fehlinformationen, die über einen Companion vermittelt werden, treffen daher auf weniger Widerstand. Eigene Forschung zeigt, dass Nutzende Fehlempfehlungen eines konversationellen KI-Systems seltener korrigieren als solche eines klassischen Interface.⁵ Analog zu den Richtlinien und Werten beim menschlichen Kundenkontakt oder klassischen Kundeninterfaces tragen Unternehmen daher auch beim Einsatz von Companions Verantwortung für die faktische Richtigkeit von Inhalten, die Transparenz über Unsicherheiten sowie dafür, dass Antworten der jeweiligen Situation der Nutzenden gerecht werden (z.B. indem bei sensiblen Themen auf professionelle Beratung hingewiesen wird).

Relationale Verantwortung. Über längere Nutzungszeit können sich bei einigen Nutzenden relationale Dimensionen formen, etwa Gefühle von Vertrautheit und emotionaler Nähe zum Companion⁶. Dies kann Risiken wie emotionale Abhängigkeit oder eine erhöhte Beeinflussbarkeit bei Entscheidungen mit sich bringen, wobei das Ausmass und die Anfälligkeit je nach Nutzenden variieren können. Was diese Entwicklung für das Wohlbefinden bedeutet, ist noch wenig erforscht. Die Erfahrungen mit sozialen Medien mahnen jedoch zur Vorsicht, denn dort wurden potenzielle Langzeiteffekte, insbesondere bei jüngeren Nutzenden, erst sichtbar, als die

³ Lucas et al. (2014), Computers in Human Behavior, 37, 94-100: Nutzende offenbarten einem virtuellen Agenten signifikant mehr sensible Informationen, wenn sie annahmen, dass kein Mensch zuhört.

⁴ Blut, Wang, Wunderlich & Brock (2021), Journal of the Academy of Marketing Science, 49(4), 632–658: eine Meta-Analyse über 108 Studien zeigt, dass Nutzende KI-Systemen systematisch menschliche Eigenschaften zuschreiben.

⁵ Hildebrand & Bergner (2021), Journal of the Academy of Marketing Science: Nutzende akzeptierten fehlerhafte Portfolioempfehlungen eines konversationellen Robo-Advisors häufiger als identische Empfehlungen eines klassischen Interface.

⁶ De Freitas et al. (2025), Journal of Consumer Research: KI-Companions können Einsamkeit reduzieren und soziale Bedürfnisse erfüllen.

Plattformen bereits weit verbreitet waren. Beim Einsatz von KI-Companions sollten daher Unternehmen bereits jetzt Verantwortung für die bewusste Gestaltung der Beziehungsebene übernehmen. Dazu gehört insbesondere die Autonomie der Nutzenden zu stärken und Interaktionen so zu gestalten, dass sie keine verzerrten Erwartungen an Gegenseitigkeit oder Verfügbarkeit erzeugen. Dazu sollten Unternehmen Feedbackschlaufen vorsehen, die problematische Nutzungsmuster frühzeitig erkennen und begrenzen, und klare Übergänge zu menschlicher Unterstützung ermöglichen.

Was das für Schweizer Unternehmen bedeutet

Schweizer Unternehmen, die KI-Companions einsetzen, agieren in einem Umfeld hoher Vertrauenserwartungen und einer sich verschärfenden Regulierungslandschaft (DSG, EU AI Act). Es geht hier jedoch nicht in erster Linie um Compliance, sondern um Fürsorgepflicht. Unternehmen, die durch KI Vertrauensbeziehungen schaffen, tragen Verantwortung dafür, wie diese Beziehungen gestaltet werden und was sie leisten.

Verantwortungsvolles Design wird zunehmend zum Differenzierungsmerkmal. Bestehende Instrumente wie das Label «Data Fairness» von SWISS INSIGHTS bieten dabei einen nützlichen Ausgangspunkt, der mit companion-spezifischen Anforderungen erweitert werden kann. Unternehmen, die diesen Anspruch ernst nehmen und ihn durch nachvollziehbare Prozesse belegen können, sind besser für die Zukunft positioniert.

Die folgenden Empfehlungen für das verantwortungsvolle Design von KI-Companions orientieren sich an den sechs Werten des Ethik-Kodex der Data Innovation Alliance.⁷

- **Schadensvermeidung:** Companion-Outputs regelmässig auf Korrektheit prüfen, insbesondere in sensiblen Bereichen wie Gesundheit, Finanzen, und Recht
- **Gerechtigkeit:** Sicherstellen, dass vulnerable Gruppen (ältere, bildungsferne oder psychisch belastete Nutzende) durch Companion-Interaktionen nicht systematisch schlechter beraten oder stärker manipulierbar werden als andere
- **Autonomie:** Klare Grenzen bei emotionaler Abhängigkeit setzen (z.B. keine Simulation exklusiver Beziehungen) und Zugänge zu menschlicher Unterstützung schaffen
- **Transparenz:** Nutzende transparent darüber informieren, dass sie mit KI interagieren, auch wenn sich die Interaktion persönlich anfühlt; klar kommunizieren, wo der Companion an seine Grenzen stösst

⁷ https://data-innovation.org/wp-content/uploads/2021/01/w-che-EthicalCodex_A4_Broschure_Grundlagen_DE.pdf

- **Kontrolle:** Nutzungsmuster beobachten und Frühwarnsysteme einrichten; Interne Rückkopplungsschleifen etablieren, einschliesslich gezielter Tests auf unerwünschte Outputs
- **Rechenschaft:** Klare fachliche, technische und organisatorische Verantwortlichkeiten für den gesamten Lebenszyklus (von Design, Training und Einführung bis zu Monitoring, Eskalation und Korrektur) festlegen
- **Rechenschaft:** Interne Governance Strukturen etablieren (Monitoring, Eskalationsmechanismen, klare Verantwortlichkeiten)

Über Swiss Insights

Swiss Insights ist der Schweizer Verband und die Interessenvertretung der datengetriebenen Wirtschaft – von der Markt-, Meinungs- und Sozialforschung über Data Science und Anwendungsentwicklung bis hin zu Insight-, Marketing- und Innovationsteams in Unternehmen. Swiss Insights vereinigt damit alle Akteure, die entlang der Datenwertschöpfungskette tätig sind: jene, die Daten erheben und analysieren, jene, die daraus prädiktive Modelle und datengetriebene Anwendungen entwickeln, und jene, die Erkenntnisse in unternehmerische Entscheidungen überführen.

Als Verband setzt sich Swiss Insights für Wissenschaftlichkeit, Qualität und ethische Standards ein und pflegt einen aktiven Dialog mit politisch und gesellschaftlich relevanten Akteuren. Swiss Insights vertritt die Interessen seiner Mitglieder gegenüber Öffentlichkeit und Politik, entwickelt Leitlinien zur Qualitätssicherung und fördert den Austausch mit nationalen und internationalen Fachorganisationen.

Der Einsatz von KI-Companions ist nicht nur eine technologische Innovation, sondern auch eine neue Form unternehmerischer Verantwortung. Wo Unternehmen durch KI Vertrauen schaffen und langfristige Beziehungen gestalten, tragen sie Verantwortung für die Qualität der Informationen, die Gestaltung der Interaktion und den Schutz der Autonomie ihrer Kundinnen und Kunden. Swiss Insights versteht es als Aufgabe, diesen Wandel mitzugestalten und Unternehmen praxisnahe Orientierung für den verantwortungsvollen Einsatz von KI-Companions zu bieten.