POLICY BRIEF

Empfehlungen zu Companion-AI (CAI)

Vier Handlungsfelder mit konkreten regulatorischen, durchsetzungsbezogenen und förderpolitischen Maßnahmen.

LEGENDE

- CAI Apps: Eigenständige Companion-AI-Anwendungen wie Replika, Character.AI, Nomi.
- CAI Funktion in LLM: Companion-artige Interaktion innerhalb von Mehrzweck-Sprachmodellen wie ChatGPT, Gemini, Claude.

01

Individuelle Nutzer schützen

Datenschutz wirksam durchsetzen

Daten aus intimer Selbstoffenbarung sind strikt auf die Kernfunktionalität der Anwendung zweckzubinden. Eine Verwendung sensibler Daten für Werbezwecke oder sonstige kommerzielle Nutzungen im Rahmen einer „Intimacy Economy" ist auszuschließen.

Kontextbasierten Krisenmechanismus vorschreiben

Anbieter sollten verpflichtet werden, bei Anzeichen krisenhafter Zustände oder suizidaler Äußerungen sofort, aktiv und mit Bedacht zu intervenieren und niedrigschwellig auf professionelle Hilfsangebote hinzuweisen.

Companion-Funktionen bei LLM trennen

Risikoabschätzung, Behandlung und die Umsetzung von gesetzlichen Pflichten setzen einen definierten Verwendungszweck voraus. Companion-Funktionen, verstanden als persistente, Persona-basierte Konversation mit simulierter emotionaler Bindung, sollten in LLM in klar abgetrennten Modi mit eigener Risikobehandlung, getrennter Datenverwertung und eigener Altersprüfung angeboten werden.

Dialog- und kontextbasierten Jugendschutz bevorzugen

Als Mittel der Wahl bei der Umsetzung von Jugendschutzmaßnahmen empfiehlt sich eine dialogbasierte Schutzhandlung. Anbieter sollten bei im Dialog erkennbaren Anzeichen von Minderjährigkeit die Interaktion behutsam unterbrechen und auf altersgerechte Alternativen sowie Bezugspersonen verweisen. Eine solche Maßnahme wäre gegenüber einer vorgelagerten Identitäts- oder Altersverifikation milder, da sie keine zusätzliche Datenerhebung voraussetzt und erst situativ an erkennbare Schutzbedarfe anknüpft. Art. 28 Abs. 3 DSA stellt für Plattformen klar, dass keine Pflicht zur zusätzlichen Datenverarbeitung besteht, verlangt zugleich

aber ein hohes Schutzniveau. Dieses Schutzniveau sollte daher nicht vorrangig durch die Abfrage zusätzlicher Daten, sondern durch risikoangemessene Schutzmaßnahmen innerhalb der Interaktion erreicht werden.

Anhang III KI-VO um Manipulation erweitern

Anhang III sollte einen eigenständigen Bereich für KI-Systeme aufnehmen, deren Zweckbestimmung in der Manipulation menschlicher Entscheidungsfindung, Verhaltensweisen oder Emotionen liegt. Die Kommission sollte gemäß Art. 112 Abs. 4 lit. Für eine entsprechende Anpassung verbleibt hinreichend Zeit, da im Rahmen der Verhandlungen zum Digital Omnibus mit neuen Beschlüssen[1] eine Verlängerung der Umsetzungsfrist für die betreffenden Pflichten bis zum 2. Dezember 2027 vorgesehen wurde.

Bestehende KI-Verbote durchsetzen

Companion-AI-Systeme sind unverzüglich durch Bundesnetzagentur und AI Office auf verbotene Praktiken nach Art. 5 KI-VO zu prüfen. Bei Verstößen sind sofortige Maßnahmen zu ergreifen.

EU-Leitlinien zu verbotenen KI-Praktiken korrigieren

Das Companion-AI-Beispiel in Rn. 134 der Leitlinien zu Art. 5 KI-VO (Kommission 2025) sollte gestrichen werden, da es Companion-AI fälschlich als unbedenklich einordnet. Der Eintrag in Rn. 88, der diese Systeme als schädlich klassifiziert, sollte beibehalten werden.

Einhaltung des KI-Verhaltenskodex überprüfen

Bundesnetzagentur und Kommission sollten systematisch überprüfen, ob Unterzeichner des GPAI-Verhaltenskodex, etwa OpenAI mit ChatGPT, Google mit Gemini und Microsoft mit Copilot, die zur Umsetzung dieses Kodex erforderlichen Maßnahmen eingehalten haben.

KI-Haftungsrichtlinie neu vorlegen

Ein neuer Vorschlag sollte Beweiserleichterungen und eine Kausalitätsvermutung zugunsten von KI-Geschädigten vorsehen und die bereits erzielten Verhandlungsergebnisse aufgreifen. Wer eine potentiell gefährliche KI einführt und die Manifestierung bekannter Risiken in Kauf nimmt, sollte im Schadensfall beweisen müssen, dass eine Verschlechterung des Gesundheitszustands oder ein Suizid nicht auf die Konversation mit der Companion-App zurückzuführen ist.

[1] Siehe Pressemitteilung vom 07.05.2026.

CAI-Vorfall-Datenbank heranziehen

Schadensfälle im Zusammenhang mit Companion-AI, die insbesondere durch laufende Klageverfahren bekannt geworden sind, in einer CAI-Vorfall-Datenbank zusammengetragen. Diese wird laufend erweitert, um die Risikoevaluierung zu erleichtern und den Zugang zu laufenden Verfahren für Litigation-Arbeit von Verbraucherschutz, Aufsichtsbehörden und NGOs zu verbessern.

02

Informationsintegrität und Entscheidungsautonomie sichern

Werbliche Inhalte klar visuell trennen

Werbliche Inhalte sollten durch eine einheitliche visuelle Absetzung klar von Outputs getrennt werden, etwa durch einen farblich abgehobenen Kasten neben oder unter dem Text. Die gegenwärtig aufkeimende Kennzeichnungspraxis wird in Tests nachweislich nicht bemerkt. Eine Integration in den Output-Text ist auszuschließen. Werbung darf keinen Einfluss auf die Erstellung der Outputs haben.

ChatGPT als VLOSE einstufen

ChatGPT sollte unverzüglich als sehr große Online-Suchmaschine im Sinne des Art. 33 Abs. 1 DSA eingestuft werden.

Einordnung als VLOP ergänzend prüfen

Ergänzend sollte geprüft werden, ob ChatGPT als sehr große Online-Plattform nach Art. 33 DSA einzuordnen ist, um die Anwendung von Art. 25 Abs. 1 DSA zum Schutz vor manipulativen Gestaltungselementen sowie von Art. 28 DSA zum erhöhten Jugendschutz sicherzustellen.

Digital Fairness Act verabschieden

Der Digital Fairness Act (DFA) verspricht weiteren Schutz vor Risiken durch Companion-AI, indem er suchterzeugendes Design, Dark Patterns, manipulative Personalisierung und KI-gestützte Interaktionsformen wie Chatbots adressiert.

03

Absenkung des bestehenden Schutzniveaus verhindern

Absenkung des Schutzes sensibler Daten dringend stoppen

Die im Digital Omnibus vorgeschlagenen Änderungen zur Absenkung des Schutzes sensibler Daten im KI-Kontext sollten abgelehnt werden. Intime Gesprächsdaten Minderjähriger und Erwachsener werden durch kontinuierliche Animierung zur Selbstoffenbarung gewonnen und bereits kommerziell verwertet. Durch die

angekündigte Einführung von Werbung in LLM-Systemen steht eine weitere Ausweitung dieser Verwertung unmittelbar bevor.

04

Fairen Wettbewerb schützen

Schwarze Liste unlauterer Geschäftspraktiken anpassen

Der Anhang zu § 3 Abs. 3 UWG enthält Praktiken, die ohne Einzelfallprüfung stets als unlauter gelten. KI-gestützte Manipulationspraktiken sollten in die Schwarze Liste des UWG aufgenommen werden, um Wettbewerber zu schützen, die keine manipulativen Praktiken nutzen.

Kontakt: Ramak Molavi Vasse'i

Ramak.molavi@digitalreche.de