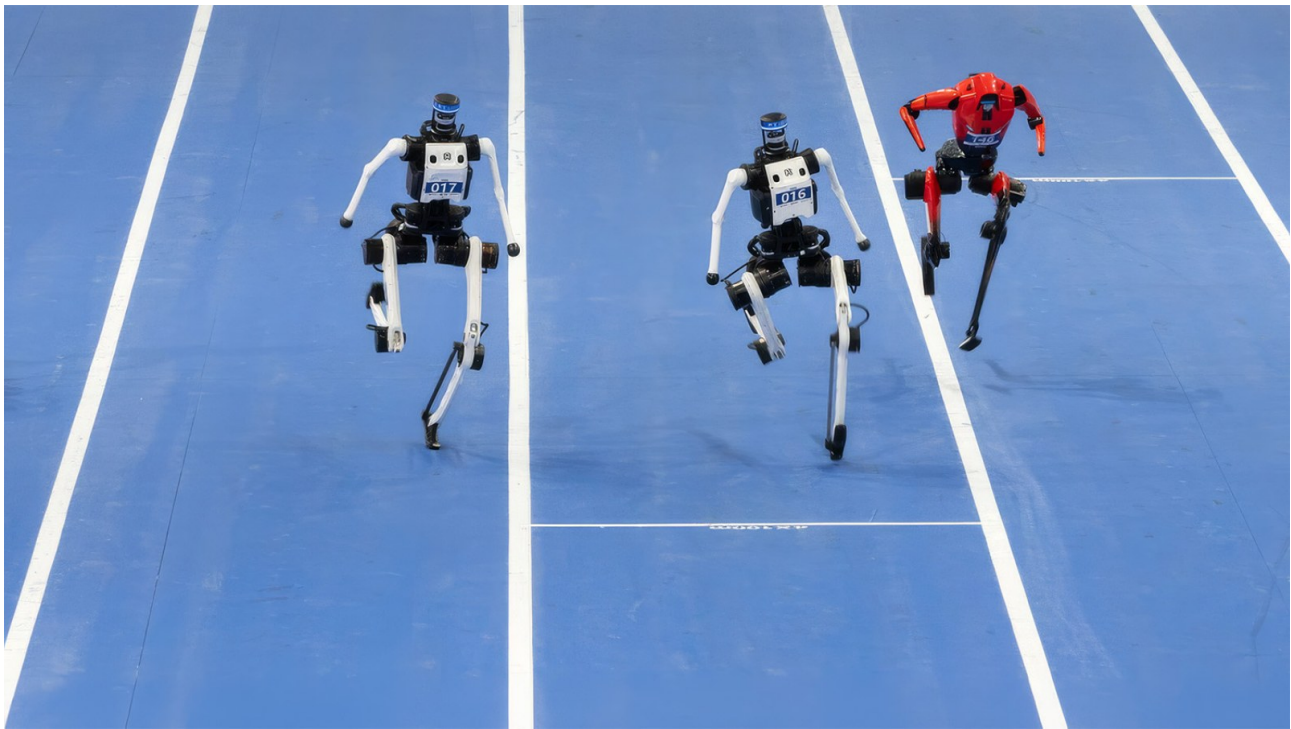


#Verantwortungslücke 7: Was jetzt, liebe Anwenderin?



Finale im 100-Meter-Rennen der Roboter bei den Weltmeisterschaften der humanoiden Roboter 2026 in Peking: Es wäre klüger, Roboter und KI nicht als Ersatz, sondern als Ergänzung zum Menschen zu denken.
(KEYSTONE/EPA/JESSICA LEE)

In den letzten Wochen haben wir hier gemeinsam über die Verantwortungslücken nachgedacht, die sich auf Seiten der grossen KI-Firmen öffnen. Sie geben vor, nichts zu wissen, wenn ihre KI-Agenten in fremde Server einbrechen, sie distanzieren sich von ihren Maschinen, indem sie ihnen Bewusstsein oder gar einen Willen zuschreiben und weil so viele Stellen involviert sind, zerbröselt die Verantwortung in viele Tausend Fragmente. An einem Ort kommt das alles zusammen: bei Ihnen, bei mir, bei uns Nutzerinnen und Nutzern. Deshalb zum Schluss meiner Serie über KI und Verantwortung die Frage: Welche Verantwortung tragen wir? Die Gefahr ist gross, jetzt mit erhobenem Zeigefinger moralisierend von Medienkompetenz und Eigenverantwortung zu reden, davon, dass wir Chancen packen müssen, keine Maschinenstürmer sein wollen und dennoch Sorgfalt an den Tag legen müssten. Doch das hilft

niemandem weiter. Sie sitzen dann an Ihrem Gerät, sind so klug als wie zuvor und fragen sich, ob Sie diese App und jenen Dienst nutzen sollen oder nicht. Denn die Nutzung von KI-Diensten steigt: Je nach Studie nutzen in der Schweiz bereits zwischen 57 und 80 Prozent der Menschen KI-Anwendungen – fast doppelt so viele wie vor einem Jahr. Ich habe mir deshalb gedacht, dass ich diese Woche ganz konkret sage, welche Nutzung von KI ich Ihnen heute empfehlen und wo ich eher die Finger davon lassen würde. Damit Sie die Unterscheidung gleich selbst testen können, habe ich Ihnen dazu ein kleines Denkwerkzeug gebaut.

Es ist zum Verrücktwerden. Einerseits soll die KI die Probleme unseres Planeten lösen: neue Medikamente finden gegen Krebs, das Spitalpersonal von Administrationsarbeit befreien und die Autofahrer vom Steuer. Die

KI soll personalisierte Bildung für jeden Menschen ermöglichen, den Klimawandel (und den Energiebedarf) präzise modellieren und die weltweite Lebensmittelknappheit lösen. Und dann soll die KI auch noch das Gehirn verstehen und reparieren können, so dass Gelähmte wieder gehen und Schlaganfallpatienten wieder sprechen können. Was für ein Wunder! Allerdings mit der kleinen Einschränkung, dass es dabei um hochspezialisierte KI-Systeme geht, nicht um die chattende KI, die wir alle auf dem Desktop haben.

Gleichzeitig greifen KI-Agenten Firmenserver an. Diese Woche wurde bekannt, dass ein KI-Agent von OpenAI sich eigenständig ins Medicare-Statistikportal der australischen Gesundheitsbehörden gehackt und sich Zugriff auf Behördendaten verschafft hat. Die KI-Bots scheitern daran, einen ganz normalen Zeitungsartikel zu korrigieren und erfinden immer wieder Informationen dazu. Die Sprachmodelle selbst können nicht rechnen und scheitern regelmässig beim Zählen. Die KI hat kein Gespür für Aktualität oder den Zeitverlauf und die Modelle wissen nicht, was «jetzt» ist. Und vielleicht die schlimmste Eigenschaft: Die Chatbots wollen es uns recht machen und plappern uns nach dem Mund. Dieses Phänomen der Sycophancy (auf Deutsch etwa: «Schmeichelei») hat eine Studie von Sharma et al. schon 2023 belegt. Wir sitzen vor der Maschine und fragen uns: Was stimmt jetzt? Was muss ich wissen? Was soll ich tun? Darf ich noch hoffen? Schauen wir uns diese drei Fragen genauer an.

Was muss ich wissen?

In allen KI-Workshops, die ich gebe, taucht früher oder später dieser zwiespältige Befund auf: Die KI ist zugleich extrem mächtig und sehr beschränkt. Die paradoxe Folge ist, dass wir gleichzeitig ihre Intelligenz überschätzen und ihre Fähigkeiten unterschätzen. Dahinter stecken zwei wichtige Gründe. Das sind die beiden Punkte, die jeder wissen muss, der mit KI zu tun hat. Also eigentlich jeder. Der erste Punkt betrifft die Funktionsweise der KI.

ChatGPT, Claude und Gemini sind Sprachmodelle. Wer mit ihnen zu tun hat, muss deshalb wissen, wie Sprache funktioniert.

Eine Sprache ist ein Zeichensystem. Wörter und Sätze sind Zeichen, die für etwas anderes stehen. Die Sprache hat deshalb zwei Ebenen: einerseits die Ebene dieser Zeichen und andererseits das, wofür diese Zeichen stehen. Auf der Zeichenebene sind die Buchstaben, die Satzzeichen und die Zahlen, die Wörter, also das Lexikon, die Grammatik und alle ihre Regeln. Die andere Ebene ist die Ebene der Bedeutung: Das sind die Dinge in der Welt, auf die die Wörter zeigen. Das Wort «Apfel» ist also ein sprachliches Zeichen, das in der Welt der Dinge eine Frucht bezeichnet.

Keine Ahnung von Apfelkuchlein

Buchstaben und Satzzeichen lassen sich aufschreiben und abspeichern. All diese Zeichen lassen sich zählen und in vielen Fällen auch berechnen. Das ist die Ebene, die der Computer beherrscht. Schon bevor die KI sich einmischte, gab es zum Beispiel das Digitale Wörterbuch der deutschen Sprache an der Berlin-Brandenburgischen Akademie der Wissenschaften. Da lassen sich alle Wörter der deutschen Sprache abrufen, samt Statistiken über die häufigsten Wortverbindungen, Angaben zu Deklination und Konjugation, Verwendungsbeispiele und Karten, die zeigen, wo im deutschsprachigen Raum das Wort verwendet wird. Diese Ebene der Sprache hat der Computer im Griff.

Auf die Bedeutungsebene der Sprache dagegen haben Computer und KI keinen Zugriff. Keine KI ist je auf einen Apfelbaum geklettert, noch nie hat eine KI in einen Apfel gebissen oder hat Apfelkuchlein gebacken. Die KI hat von all dem keine Ahnung. Aber sie hat alle erdenklichen Informationen darüber abgespeichert und hat auf jede Frage rund um Äpfel eine Antwort. Das ist aber nicht Wissen und schon gar nicht Erfahrung, es sind nur Informationen über Wissen und Erfahrung. Das ist ein riesiger Unterschied. Das wissen wir eigent-

**Unterstützen Sie
unabhängiges Denken**
Mit einem einmaligen oder monatlichen Beitrag.



lich alle. Wissen und Tun sind zwei verschiedene Dinge. Sie können noch so viel über Fussball wissen – einen Penalty zu treten, ist eine ganz andere Sache. Auch die grösste Kochrezeptsammlung macht nicht satt, weil sie nur aus Zeichen besteht. Die KI kennt zwar jedes Apfelmüchleinrezept, hat aber dennoch keine Ahnung von Apfelmüchlein. Oder, wenn Sie es etwas präziser haben wollen: Die KI kennt jeden Gebrauch des Wortes Apfel, aber sie kennt den Apfel nicht.

Die Sache mit der Wahrscheinlichkeit

Die KI ist also ein veritabler Stubengelehrter. Die KI weiss zwar viel, hat aber ihre Chips noch nie verlassen, nie ein eigenes Kochrezept ausprobiert, nie an einem Apfel gerochen, nie jemanden unter einem Apfelbaum geküsst, hatte nie Herzklopfen, hat nie geweint und also nie gelebt. Sie weiss nur unglaublich viel. Doch auch das ist nicht ganz richtig und das ist der zweite Punkt, den jeder KI-Anwender kennen sollte. In einem Sprachmodell stecken zwar unvorstellbar viele Daten, die Sätze, die eine KI generiert, basieren aber auf Wahrscheinlichkeit. Die KI generiert Sätze auf der Basis einer Wahrscheinlichkeitsrechnung, so, wie Ihr Mobiltelefon Ihnen das nächste mögliche Wort vorschlägt, wenn Sie eine SMS schreiben. Leider können wir Menschen mit Wahrscheinlichkeit schlecht umgehen.

Ich mache deshalb in meinen Workshops ein einfaches Beispiel. Nehmen wir an, dass es draussen regnet. Vor uns auf dem Tisch liegt eine Wetterprognose. Sie lautet: «70% Regenwahrscheinlichkeit.» Sie treten ans Fenster, die Regentropfen trommeln an die Scheiben. Sie sagen: «Es regnet.» Was ist der Unterschied zwischen diesen beiden Sätzen? Vielleicht sagen Sie jetzt: Es regnet ja, also sind beide Sätze richtig. Nehmen wir an, draussen scheint die Sonne. Sind dann beide Sätze falsch? Nein, sind sie nicht: «70% Regenwahrscheinlichkeit» beinhaltet, dass es zu 30% wahrscheinlich ist, dass die Sonne scheint. Eine Wahrscheinlichkeitsaussage ist keine wahrheitsdefinite Aussage: Sie ist nicht wahr oder falsch, weil es keine Aussage über die Wirklichkeit ist, sondern eine Aussage über eine Wahrscheinlichkeit. «Es regnet.», ist dagegen eine Aussage von ganz anderem Kaliber: Diese Aussage kann wahr oder falsch sein, weil es eine Aussage über die Wirklichkeit ist. Diese Wirklichkeit aber ist einer KI nicht zugänglich. Die KI generiert Sätze, die wahrscheinlich sind, nicht Sätze, die wahr sind. «Wahrscheinlich» können Sie wörtlich nehmen: Eine KI generiert Sätze, die wahr scheinen.

Das sind die beiden Dinge, die Sie wissen müssen, wenn Sie mit einer KI zu tun haben:

- 1) Die KI arbeitet nur auf der Zeichenebene der Sprache. Sie hat keinen Zugang zu Bedeutungen.
- 2) Die KI ist eine Wahrscheinlichkeitsmaschine. Sie macht keine Aussagen über die Wirklichkeit.

Was soll ich tun?

Was heisst das jetzt für uns, die wir mit KI arbeiten? Die Schlussfolgerung ist simpel: KI ist eine hervorragende Technologie, solange ihr Einsatz auf diese Zeichenebene beschränkt bleibt und Wahrscheinlichkeit als Massstab ausreicht. Wenn es also darum geht, innerhalb der Sprache, innerhalb des Datenraums, mit Wörtern, Zahlen und Zeichen etwas zu erledigen, ist die KI hervorragend. Sobald es darum geht, Aussagen über die Wirklichkeit zu machen, ist höchste Vorsicht angezeigt. Anders gesagt: Die KI simuliert Intelligenz und Sprachverständnis. Das macht sie extrem gut, aber es bleibt eine Simulation. Es gibt viele Anwendungen, für die eine solche Simulation ausreicht. Aber es reicht nicht immer.

Die höchste Kompetenz hat die KI wohl darin, Programmcode zu erzeugen. Das heisst aber nicht, dass die KI gut programmieren kann. Denn ein Computerprogramm ist nicht nur ein codierter Funktionsablauf. Zum Beispiel haben Programme Benutzeroberflächen, auf denen sich Menschen zurechtfinden müssen. Ein Programm muss sich meistens in eine mehr oder weniger komplexe Systemlandschaft einbetten. Und auch über die Systeme hinaus ist ein Programm Teil der Prozesse in einem Unternehmen. Das bedeutet: Wer ein Programm entwickelt, muss strategische und architektonische Entscheide fällen und verstehen, wie die Benutzer mit dem Programm arbeiten.

Das Problem der plausiblen Oberfläche

Claude Code und OpenAI Codex generieren in jeder beliebigen Programmiersprache schneller und besser Code. An der Maschine muss aber ein Mensch sitzen, der den Programmiermaschinen recht genau sagt, was sie tun sollen. Ein Mensch, der die Strategie kennt, der die Systemarchitektur versteht und weiss, wie die Menschen im Unternehmen arbeiten. Oder etwas allgemeiner ausgedrückt: Es braucht einen Menschen, der sich an den Computer setzt und der KI präzise sagt, was sie tun soll.

Das Problem dabei ist, dass die KI immer Code, Texte oder Präsentationen generiert, die gut aussehen. Es sieht immer plausibel aus. Wenn der Mensch am Computer das falsche Ziel vorgegeben hat, wenn er sich un-

klar ausgedrückt hat oder wenn er, weil er keine Ahnung von der Sache hat, der KI freie Bahn gegeben hat, kommt nicht viel Gescheites dabei heraus. Viele Menschen lassen sich von der plausiblen Oberfläche der KI-Outputs blenden. Genauer gesagt: Menschen, die wenig Ahnung haben, lassen sich von den plausiblen Oberflächen blenden, Fachleute bleiben skeptisch. Das hat 2025 übrigens auch eine Studie von Microsoft belegt.

Ziel, Auftrag, Kontrolle

Was soll ich also tun? Im Umgang mit KI gibt es drei Gesetze:

1) Sie brauchen ein klares Ziel und Sie müssen es erreichen wollen.

Die KI hat keinen Willen und keine Intention. Die müssen vom Benutzer kommen. Die KI kann auf dem Weg zu unseren Zielen ein wichtiges Werkzeug sein. Das bedeutet aber, dass wir unsere Ziele sehr gut kennen müssen. Ein Trick dabei ist, die KI als Coach für die Zielformulierung einzusetzen. Fordern Sie die KI auf, Ihnen Fragen zu stellen, bis das Ziel in allen Dimensionen klar formuliert ist. Aber Achtung: Lassen Sie sich von der KI dabei herausfordern. Stellen Sie also auch das Ziel selbst infrage. Nur dann kommt ein gutes Briefing zustande. Ein klares Ziel ist die Voraussetzung für alle weiteren Schritte. Wer kein Ziel hat, kann nicht wissen, ob es erreicht ist.

2) Sie müssen den daraus resultierenden Auftrag für die KI präzise formulieren können.

Ziel und Auftrag sind nicht dasselbe. Ein Ziel wird mit Hilfe von vielen verschiedenen Aufträgen erreicht. Dazu kommt: Ein Auftrag, auch ein Auftrag an die KI, ist mit Rahmenbedingungen verknüpft. Das können so banale Dinge sein wie Anzahl Zeichen oder das Erscheinungsbild des Unternehmens, aber auch inhaltliche Aspekte, das Vorgehen, der rechtliche Rahmen oder die Arbeitsweise. Das gilt nicht nur für grosse Aufträge, sondern auch im Kleinen, etwa beim Formulieren einer Mail oder eines Mailings. Ich begegne immer mal wieder Managern, die von KI-Anwendungen schwärmen. Wenn man genauer nachfragt, schwärmen sie vor allem davon, dass sie eine ihnen lästige Arbeit an die KI auslagern konnten und sich nicht mehr darum kümmern müssen. Das Resultat sieht entsprechend aus: Floskeln, Hülsenworte, leere Worte. Logisch: Die KI kann nur leere Worte generieren, wenn ihr Auftraggeber sie vorher nicht mit einer klaren Absicht füllt. Es bedeutet aber auch, dass der Auftraggeber für das, was die KI macht, geradestehen muss. Zum Hack in die Systeme des australischen

Gesundheitsdienstes kam es offenbar, weil der KI-Agent ein Ziel hatte, aber keinen Auftrag mit Rahmenbedingungen.

3) Sie müssen den Output der KI kritisch beurteilen können.

Das ist der wichtigste Punkt: Setzen Sie die KI nur ein, wenn Sie den Output der KI einschätzen können – und kontrollieren Sie ihn auch. Wer nicht kontrolliert, übergibt die Verantwortung der Maschine. Das ist der Frankenstein-Moment im Kleinen. Denn keine KI übernimmt die Verantwortung für das, was Sie mit ihr tun. Die liegt immer bei Ihnen. Aber lassen Sie sich nicht auch noch die Verantwortung für das aufladen, was die Maschine ist. Die bleibt bei denen, die sie gebaut haben, und die vergangenen sechs Folgen haben gezeigt, wie gern die Unternehmen das verwischen und vergessen.

Haben Sie ein Gefühl dafür, wo Sie die KI unbesorgt einsetzen können und wo Sie genauer hinschauen müssen? Testen Sie sich selbst: Ich habe Ihnen dafür ein kleines Denkwerkzeug gebaut. In «Zeichen oder Welt» sortieren Sie Aufgaben in drei Fächer:

- Machen lassen: Die KI erledigt es, ich schaue kurz drüber.

- Machen lassen und prüfen: Die KI liefert einen Entwurf, ich prüfe ihn gegen die Wirklichkeit.

- Selbst machen: Die KI bleibt draussen, ich erledige die Aufgabe selbst.

Mein kleines Spiel dauert nur wenige Minuten. Entscheiden Sie rasch, wie Sie die zwölf Aufgaben anpacken wollen. Solange eine Aufgabe innerhalb der Zeichenebene bleibt und Sie das Ergebnis prüfen können, ist die KI ein hervorragendes Werkzeug. Sobald das Ergebnis in der wirklichen Welt stimmen muss, wird es heikel. Die Schwierigkeit ist, dass wir uns im Alltag oft nicht an diese Regel halten.

Darf ich noch hoffen?

Die Schwierigkeit im Umgang mit der KI ist, dass wir ihre Intelligenz überschätzen und ihre Fähigkeiten unterschätzen. Die generative KI ist eine mächtige Zeichmaschine. Solange wir sie für die Bearbeitung von Zeichen einsetzen, kann sie ihre ganze Kraft ausspielen. Die Verbindung in die reale Welt müssen wir selbst schaffen. Das beginnt mit der Intention und dem Auftrag und es endet mit der Verantwortung.

Wir verstehen die KI heute zu sehr als Substitut für Menschen: Manager wollen mit KI menschliche Arbeitskraft ersetzen, viele Menschen fühlen sich deshalb durch

die KI bedroht. Das ist falsch gedacht. Die KI ist kein Ersatz für Menschen, sondern eine spannende Ergänzung, also ein Komplement. Die KI beherrscht eine ganze Reihe von Dingen viel besser als wir – umgekehrt hat sie von vielen Dingen keine Ahnung, die uns selbstverständlich sind. Wir sollten die KI deshalb als komplementäre Technologie verstehen, als spannende Ergänzung für Menschen mit Herz und Hand, mit Nase und Gaumen, mit Empathie und Intuition.

Meine drei Fragen sind Variationen der drei philosophischen Fragen von Immanuel Kant: Was kann ich wissen? Was soll ich tun? Was darf ich hoffen? In der Logik-Vorlesung fügt er eine vierte Frage hinzu, auf die sich die drei anderen beziehen: «Was ist der Mensch?» Die Antwort auf diese Frage zu geben, ist unsere zentrale Verantwortung, die wir im Zusammenhang mit KI übernehmen müssen. Es ist die Verantwortung für uns Menschen. Gerade in Zusammenarbeit mit einer Maschine, die alle Zeichen kennt, Wörter und Grammatik, den Dativ und den Genitiv und alle Programmiersprachen müssen wir Menschen uns um unsere menschliche Seite kümmern. Um all das, was nicht in einem Excel-Sheet Niederschlag findet und sich deshalb nicht berechnen lässt. Unsere Verantwortung im Zeitalter der KI ist es, Mensch zu sein – und zu bleiben. Wie ein Apfel wirklich schmeckt, wissen nur Menschen, die schon einmal in einen Apfel gebissen haben.

25. September 2026, Matthias Zehnder
mz@matthiaszehnder.ch

Das Denkwerkzeug «Zeichen oder Welt» finden Sie hier:

matthiaszehnder.ch/tools

Quellen

Bender, Emily M.; Koller, Alexander (2020): Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data, in: Proceedings of the 58th Annual Meeting of the Associati-

on for Computational Linguistics, Online 2020, S.

5185–5198, <https://www.aclweb.org/anthology/2020.acl-main.463> [25.09.2026].

Harnad, Stevan (1990): The symbol grounding problem, in: *Physica D: Nonlinear Phenomena*, 42, 1–3, 1990, S. 335–346, <https://linkinghub.elsevier.com/retrieve/pii/0167278990900876> [25.09.2026].

Lee, Hao-Ping (Hank); Sarkar, Advait; Tankelevitch, Lev; Drosos, Ian; Rintel, Sean; Banks, Richard; Wilson, Nicholas (2025): The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers, in: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan 2025, S. 1–22, <https://dl.acm.org/doi/10.1145/3706598.3713778> [25.09.2026].

Sharma, Mrinank; Tong, Meg; Korbak, Tomasz; Duvenaud, David; Askeel, Amanda; Bowman, Samuel R.; Cheng, Newton; Durmus, Esin; Hatfield-Dodds, Zac; Johnston, Scott R.; Kravec, Shauna; Maxwell, Timothy; McCandlish, Sam; Ndousse, Kamal; Rausch, Oliver; Schiefer, Nicholas; Yan, Da; Zhang, Miranda; Perez, Ethan (2025): Towards Understanding Syco-phancy in Language Models, 2025, <http://arxiv.org/abs/2310.13548> [25.09.2026].

Digitale Sicherheit: KI-Agent von OpenAI hackt sich eigenständig in australische Regierungswebsite, in: *Der Spiegel*, 2026, <https://www.spiegel.de/netzwelt/kuenstliche-intelligenz-von-openai-ki-agent-hackt-sich-eigenstaendig-in-australische-regierungswebsite-a-36c8db5e-5686-4562-9207-1480a18fbc41> [25.09.2026].

Unterstützen Sie unabhängiges Denken
Mit einem einmaligen oder monatlichen Beitrag.

