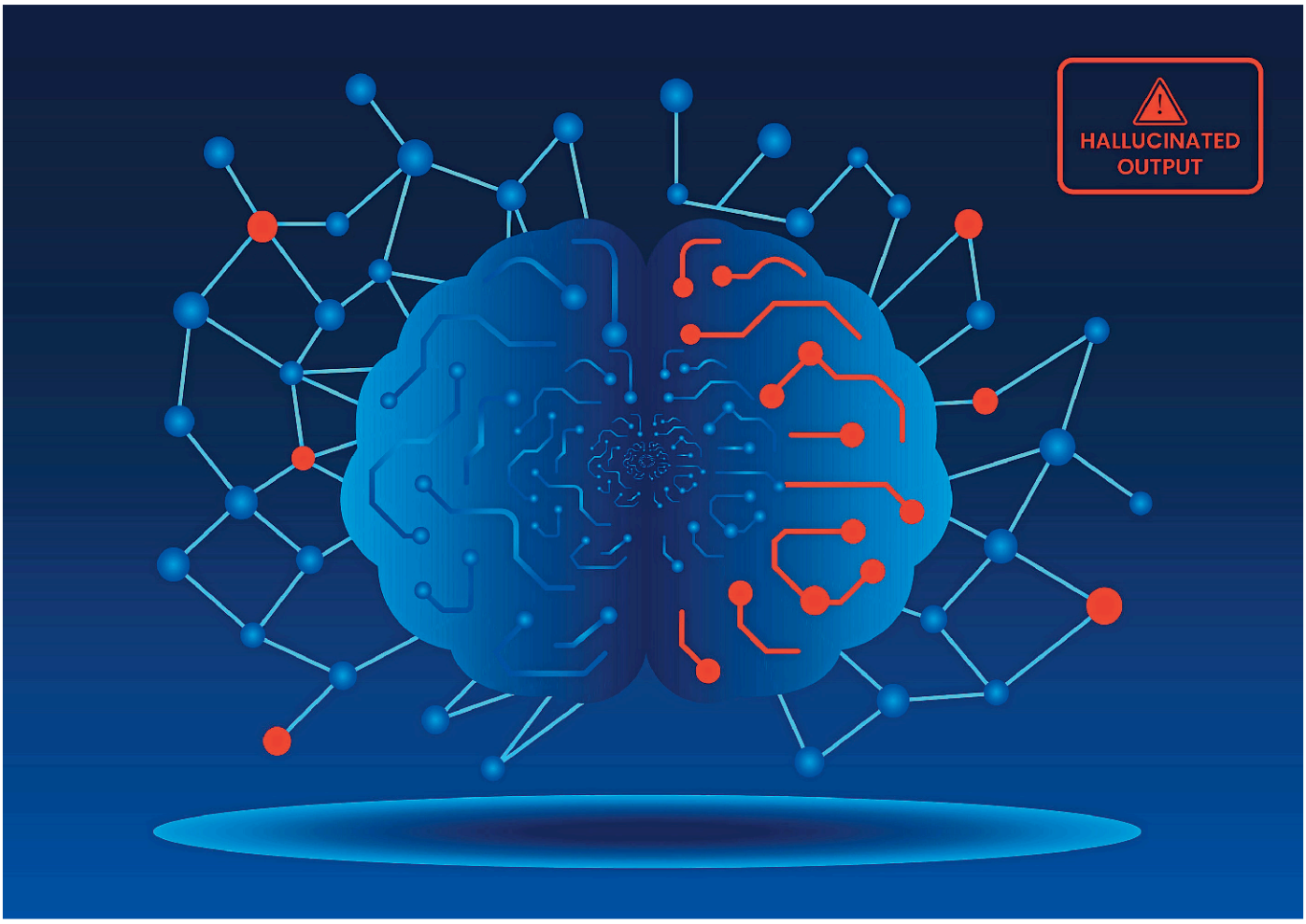


KI-Halluzinationen sind keine technischen Ausnahmen, sondern ein struktureller Bestandteil lernender Modelle.

**KI-EINSATZ IM UNTERNEHMEN**

# WENN KÜNSTLICHE INTELLIGENZ **HALLUZINIERT**

Die KI-Systeme zeigen immer mehr Eigenschaften und Fähigkeiten, die vor nicht allzu langer Zeit noch als ureigen menschlich galten. Sie reagieren (scheinbar) empathisch, sie halluzinieren. Deshalb sollte man ihren Aussagen und Empfehlungen nicht blind vertrauen.

Die Künstliche Intelligenz imitiert immer überzeugender die menschliche Kommunikation. Systeme wie Chatbots oder generative Modelle können inzwischen nicht nur faktenbasiert argumentieren, sondern auch Empathie simulieren, Stimmungen spiegeln und sensibel auf emotionale Signale reagieren.

Deshalb werden KI-Systeme, die früher als reine Rechenmaschinen galten, heute zunehmend als ein Gegenüber mit Persönlichkeit empfunden. Die menschenähnliche Wirkung der Chatbots & Co. verleitet ihre Anwender allerdings oft zum Fehlschluss: Wenn sich die KI wie ein Mensch verhält, dann versteht sie auch

Bild: Graphicsign – stock.adobe.com

wie ein Mensch Ereignisse, Arbeitsaufträge und Sachverhalte.

Diese Schlussfolgerung ist falsch und gefährlich. Denn neben der Fähigkeit, beeindruckend realistische Dialoge zu führen, Gefühle zu spiegeln und Beratungsgespräche zu führen, haben KI-Systeme noch eine Eigenschaft mit Menschen gemeinsam: Sie können halluzinieren. Oder umgangssprachlich formuliert: Sie spinnen zuweilen. Das heißt, sie erzeugen zuweilen Informationen beziehungsweise einen Output, der zwar auf den ersten und nicht selten sogar zweiten Blick plausibel klingt, jedoch falsch ist. Anders als bei Menschen geschieht dies aber nicht aufgrund einer überbordenden Fantasie oder Selbsttäuschung, sondern aufgrund der statistischen Mechanismen, die KI-Systemen zugrunde liegen. Der Effekt ist aber derselbe: Ein souverän vorgetragenes oder präsentiertes (Arbeits-)Ergebnis, das inhaltlich jedoch nicht stimmt.

## WARUM KI-SYSTEME HALLUZINIEREN

Die Halluzinationen der KI-Systeme entstehen nicht durch Fehler im herkömmlichen Sinne, sondern durch das Funktionsprinzip großer Sprachmodelle. Sie erzeugen Antworten, indem sie Wort für Wort die wahrscheinlichste Fortsetzung eines Textes berechnen. Sie wissen und verstehen sozusagen nichts: sie rechnen! Je unvollständiger oder unklarer die Eingaben/Prompts sind, je weniger Daten zu einem Sachverhalt existieren oder je höher der Druck ist, eine kohärente Antwort zu liefern, desto wahrscheinlicher werden Elemente erfunden. Dies auch deshalb, weil KI-Systeme nicht darauf programmiert sind, im Bedarfsfall auf Prompts zu antworten „Das weiß ich nicht“ oder „Dafür fehlen mir die nötigen (aktuellen) Infos“. Oder: „Ich bin kein Hellseher. Ich kann die Zukunft ebenso wie Sie nicht vorhersehen; ich kann nur Wahrscheinlichkeiten berechnen.“

Häufige Ursachen der KI-Halluzinationen sind:

- **Datenlücken:** Wenn ein Modell zu einem Thema zu wenig Trainingsdaten hatte, generiert beziehungsweise erfindet es fehlende Informationen.
- **Übergeneralisierung:** KI-Systeme verallgemeinern Muster, manchmal zu sehr. Aus häufigen Strukturen leiten sie vermeintlich logische, jedoch nicht selten unzutreffende Aussagen ab.
- **Mehrdeutige oder schlechte Prompts:** Wenn die Eingaben der Anwender mehrdeutig sind, entscheidet sich das Modell für die wahrscheinlichste, jedoch nicht unbedingt richtige Interpretation.
- **Zwang zur Kohärenz:** Sprachmodelle sind darauf optimiert, flüssige und konsistente Texte zu erzeugen – im Bedarfsfall auch auf Kosten der Wahrheit.

## BEISPIELE FÜR HALLUZINATIONEN

Unter anderem folgende Beispiele für halluzinierende KI-Systeme haben in jüngster Vergangenheit anlässlich einer Studie der Europäischen Rundfunkunion eine breite mediale Aufmerksamkeit erfahren: So behauptete etwa der Chatbot ChatGPT noch vor Kurzem, dass Papst Franziskus noch lebe, obwohl er bereits im April 2025 verstarb. Microsoft Copilot wusste nicht, dass Schweden seit März 2024 im Gefolge des Ukraine-Kriegs NATO-Mitglied ist. Und Google Gemini bezeichnete die Wiederwahl von Donald Trump als US-Präsident als „möglich“, obwohl dieser inzwischen seit über einem Jahr wieder im Amt ist.

Über solche Fehler aufgrund von Datenlücken kann man schmunzeln, weil fast jeder, der sich etwas für das Weltgeschehen interessiert, sie beim Lesen sofort erkennt und sie deshalb keine größeren Schäden verursachen können. Doch häufig ist dies nicht der Fall. Nicht selten müssen sich sogar Experten sehr tief in den KI-Output vertiefen, um zu erkennen, dass das, was das KI-System schreibt oder empfiehlt oder die Schlussfolgerungen, die es aus den Daten zieht, falsch sind. Auch hierfür einige reale Beispiele:

**Erfundene Quellenangaben:** Wissenschaftler und Forscher berichteten immer wieder, dass Chatbots Literaturverweise oder Studien erfanden, die plausibel klingen, aber nicht existieren; inklusive gefälschter DOI-Nummern, die beim Online-Publizieren wissenschaftlicher Arbeiten verwendet werden, um diese als seriös und vertrauenswürdig zu kennzeichnen.

> **Die KI beantwortet Fragen nicht, weil sie die Welt versteht, sondern weil sie Muster fortsetzt.** Prof. Dr. Georg Kraus

**Fiktive Gerichtsurteile:** Nicht nur in den USA reichten Rechtsanwälte wiederholt Schriftsätze ein, die von einer KI erstellt wurden – und zwar samt ausgedachter Präzedenzfälle und erfundener juristischer Begründungen.

**Erfundene biografische Daten:** Personen des öffentlichen Lebens wie Politiker, Wirtschaftsvertreter und Wissenschaftler oder auch Schauspieler und Spitzensportler finden sich immer wieder in KI-generierten Texten mit Aussagen von ihnen aus angeblichen Interviews zitiert, die nie stattfanden; des Weiteren werden sie mit Ereignissen in Verbindung gebracht, die keinerlei Bezug zu ihnen haben.

**Falsche Bilderklärungen:** Bildgenerierende Systeme ordnen immer wieder visuelle Objekte wie Fotos falschen Kontexten und Zusammenhängen zu, weil sie ein erkanntes Muster überinterpretieren. ▶

### BLINDES VERTRAUEN GEFÄHRDET UNTERNEHMEN

Die aufgeführten Beispiele verdeutlichen: KI-Halluzinationen sind keine technischen Ausnahmen, sondern ein struktureller Bestandteil lernender Modelle. Sie sind also ein Phänomen, mit dem man stets rechnen muss. Darin schlummern große Gefahren, wenn Unternehmen KI-Systeme zum Beispiel für die Analyse von Kundendaten, zum Recruiting, für das Erstellen von Marktberichten, für das Wissensmanagement oder strategische Prognosen nutzen. Und das nicht nur, weil den Usern häufig das nötige Fach- beziehungsweise Expertenwissen fehlt, um KI-Bugs zu erkennen.

> **Wer der KI blind vertraut, setzt sein Unternehmen unnötigen Risiken aus.**
>
> Prof. Dr. Georg Kraus

Mindestens ebenso groß ist die Gefahr, dass sie sich aufgrund der menschenähnlichen Interaktion der Systeme dazu verleiten lassen, die Analysen, Schlussfolgerungen et cetera der KI-Systeme nicht ausreichend zu hinterfragen: sei es aus Zeitmangel, Naivität oder Bequemlichkeit.

Entsprechend wichtig ist es, dass sich die Führungskräfte in Unternehmen sowie die Personen, die ähnliche Entscheiderpositionen innehaben, immer wieder ins Gedächtnis rufen: Die KI beantwortet Fragen nicht, weil sie die Welt versteht, sondern weil sie Muster fortsetzt. Deshalb sollten sich die Verantwortlichen in den Betrieben bei ihrem Einsatz stets bewusst sein:

- Die KI ist ein Werkzeug und kein Entscheider. Ihr Output kann zwar ein wesentlicher Input für ihre Entscheidungen sein, er darf jedoch nie als die endgültige Wahrheit erachtet werden.
- Falsche Informationen können große wirtschaftliche Schäden verursachen: von fehlerhaften Kennzahlen über falsche Prognosen bis hin zu rechtlichen Risiken.
- Halluzinationen sind nicht völlig vermeidbar. Auch die nächsten Generationen der KI-Systeme werden Fehler machen – jedoch subtiler, also für uns Menschen noch schwieriger erkennbar.

### KI VERANTWORTUNGSVOLL EINSETZEN

Damit Unternehmen von der KI profitieren, ohne sich den genannten Risiken auszusetzen, bedarf es für deren Nutzung klarer Regeln. Ein verantwortungsvoller Umgang mit ihr sollte sich unter anderem an folgenden Leitlinien orientieren:

**Eine kritische Distanz bewahren:** Auch wenn der KI-Output souverän klingt, muss jede Information geprüft werden. Das gebietet die Sorgfaltspflicht.

**Die Ergebnisse verifizieren:** Oft ist die Präsentation eines KI-Outputs beeindruckend professionell. Doch Stil und Wahrheitsgehalt stehen häufig in keinem Zusammenhang. Führungskräfte müssen lernen, das eine vom anderen zu trennen.

**Nicht als alleinige Entscheidungsgrundlage nutzen:** Insbesondere bei Personalfragen, Finanzentscheidungen, strategischen Analysen und im Compliance-Bereich sollten KI-Systeme nur unterstützend, aber nie allein eingesetzt werden.

**Klare Prompting-Regeln etablieren:** Unklare Eingaben erhöhen das Risiko von Halluzinationen. Unternehmen sollten daher Standards für das Formulieren von Fragen an die KI entwickeln: präzise, eindeutig und kontextreich.

**Die Mitarbeitenden im Umgang mit der KI schulen:** Technische Kompetenz allein reicht hier nicht. Entscheidend ist das Verständnis der Funktionsweise und Grenzen der Modelle. Die Unternehmen sollten daher Schulungen implementieren, die das kritische Denken und die Fähigkeit, Fehler zu erkennen, stärken.

**Prozesse der Qualitätskontrolle definieren:** Je nachdem, wie sensibel der Aufgabenbereich ist, sollten mehrstufige Prüfmechanismen existieren: etwa Stichproben, Peer-Reviews oder Plausibilitätschecks anhand realer Daten.

**Transparenz gegenüber Kunden schaffen:** Wenn KI-Systeme am Erbringen einer Leistung beteiligt sind, sollten Unternehmen dies offen kommunizieren, damit für die User Fehler leichter erkenn- und nachvollziehbar sind und Vertrauen nicht verloren geht.

### DEN KI-EINSATZ REFLEKTIEREN

Solche Regeln für die Nutzung der KI-Systeme werden umso wichtiger, je menschenähnlicher die Systeme in ihrer Sprache, ihren Interaktionen und sogar Fehlern werden.

Fazit: Eine überzeugende, weil scheinbar in sich logische und schlüssige Darstellung darf nicht darüber hinwegtäuschen, dass KI-Systeme keine denkenden Wesen, sondern statistische Modelle sind. Dies bedeutet: Die Fähigkeit, zwischen einem menschlich beziehungsweise realistisch erscheinenden Output und der Realität zu unterscheiden, wird zu einer zentralen Kompetenz von Entscheidern. Wer die KI klug respektive reflektiert einsetzt, profitiert davon enorm. Wer ihr jedoch blind vertraut, setzt sein Unternehmen unnötigen Risiken aus. ■

**Prof. Dr. Georg Kraus**
guenter.herkommer@holzmann-medien.de