KI Briefing #5

KI-RISIKEN IM ÜBERBLICK

Was real ist, was Panik – und was Sie differenziert einordnen sollten

Zwischen Innovation und Unsicherheit

Kaum eine Technologie der letzten Jahre hat so viel Aufbruchsstimmung und gleichzeitig so viele Warnrufe ausgelöst wie Künstliche Intelligenz. In Führungsetagen wird gefragt: Ist das sicher? Können wir das verantworten? Wo lauern die echten Gefahren?

Oft vermischen sich dabei konkrete Risiken mit spekulativen Szenarien – oder die Risiken anderer Technologien werden der KI zugeschrieben. Dieses Briefing soll Ihnen helfen, die tatsächlichen Risiken moderner KI-Systeme realistisch einzuordnen.

Vier Risikodimensionen

Nicht jedes KI-Risiko hat die gleiche Ursache – oder verlangt die gleiche Reaktion. Oft werden in der Praxis technische Probleme, rechtliche Unsicherheiten und gesellschaftliche Bedenken in einen Topf geworfen. Das erschwert die Einordnung und führt zu Über- oder Unterreaktionen. Für eine realistische Einschätzung ist es hilfreich, vier grundlegende Risikotypen zu unterscheiden:

Technische Risiken
Dazu gehören fehlerhafte Ergebnisse, sogenannte Halluzinationen, mangelhafte Trainingsdaten oder instabile Systeme nach Updates. Sie entstehen aus der Funktionsweise der Modelle selbst – oder durch falsche Anwendung.

Rechtliche Risiken
Diese betreffen vor allem Datenschutz, Urheberrecht, Geschäftsgeheimnisse und regulatorische Vorgaben wie den AI Act. Sie entstehen nicht durch die KI an sich – sondern durch ihre Nutzung im falschen Rahmen.

Soziale Risiken
KI kann Diskriminierung verstärken, Vertrauen untergraben oder interne Verunsicherung auslösen – auch wenn sie technisch korrekt funktioniert. Diese Risiken betreffen das Verhältnis zu Mitarbeitenden, Kund:innen oder der Öffentlichkeit.

Organisatorische Risiken
Hier geht es um strategische Abhängigkeiten, unklare Zuständigkeiten oder Schattennutzung. Sie entstehen oft nicht durch die Technologie – sondern durch fehlende Steuerung in der Organisation.

Diese Kategorien überschneiden sich oft – und genau deshalb ist es so wichtig, sie zu verstehen. Wer weiß, welcher Risikotyp vorliegt, kann gezielter handeln. Und wer Risiken strukturiert analysiert, stärkt nicht nur die Sicherheit – sondern auch die Handlungsfähigkeit seiner Organisation.

Typische Befürchtungen zu bestehenden Risiken

Viele Risiken, die mit Künstlicher Intelligenz verbunden werden, basieren nicht auf tatsächlichen Erfahrungen, sondern auf diffusen Ängsten oder spekulativen Szenarien. Genau deshalb ist eine sachliche Einordnung so wichtig: Nicht alle Risiken sind gleich real, nicht alle gleich relevant – und manche beruhen auf Missverständnissen.

Die folgenden Abschnitte greifen typische Befürchtungen auf, wie sie Führungskräfte häufig hören oder selbst äußern.

1. Die KI macht Fehler.

↳ ***Real, aber anders als gedacht – v. a. bei generativer KI***

Viele Führungskräfte befürchten, dass KI-Systeme schlicht „nicht richtig funktionieren". Tatsächlich ist der Begriff „Fehler" bei KI schwer zu fassen: Sprachmodelle wie ChatGPT machen keine absichtlichen Fehler – aber sie **erfinden Informationen, wenn die Datenlage unscharf ist** (sogenannte Halluzinationen). Klassifikationsmodelle wiederum treffen Wahrscheinlichkeits-aussagen – sie können also nie 100 % korrekt sein.

Was stimmt: KI kann fehlerhafte, verzerrte oder falsche Ergebnisse liefern – oft ohne dass es auf den ersten Blick erkennbar ist.

Was nicht stimmt: KI ist grundsätzlich unzuverlässig – sie kann sehr konsistent arbeiten, wenn der Kontext und die Datenlage passen.

Beispiel: Ein KI-generierter Text enthält plausible, aber frei erfundene Quellen.

Risikominimierung: Ergebnisse niemals ungeprüft übernehmen. Qualitätssicherung, manuelle Freigabeprozesse und gesicherte Anwendungsfälle einführen.

2. Die KI weiß alles.

↳ ***Nicht richtig – v. a. bei generativer KI***

Ein verbreitetes Missverständnis ist, dass generative KI-Modelle über umfassendes Weltwissen verfügen oder „alles gespeichert haben". In Wahrheit haben sie **kein aktives Wissen**, sondern **statistische Sprachmuster aus Trainingsdaten** gelernt – und kennen keine aktuellen Entwicklungen, keine internen Informationen und keine Fakten im eigentlichen Sinne.

Was stimmt: Die Ausgaben wirken oft allwissend – das kann zu Überschätzung führen.

Was nicht stimmt: KI-Modelle „wissen" etwas – sie simulieren Wissen durch Sprach-vervollständigung.

Beispiel: Eine KI antwortet selbstbewusst auf eine Rechtsfrage – mit veraltetem oder falschem Inhalt.

Risikominimierung: Mitarbeitende schulen, die Grenzen des Modells zu verstehen. Besonders bei kritischen Themen immer mit Fachexpertise gegenprüfen lassen.

3. KI kann manipulieren.

↳ ***Teils berechtigt – aber nicht im Sinne von Absicht***

KI lügt nicht – weil sie keine Absicht hat. Aber sie kann **plausibel klingende Unwahrheiten** erzeugen, z.B. durch fehlende Quellen, verzerrte Trainingsdaten oder unkritische Prompts. Gerade generative KI-Systeme sind hier anfällig: Sie erzeugen Ergebnisse, die nicht objektiv prüfbar sind, aber überzeugend klingen – das kann als „manipulativ" wahrgenommen werden, obwohl keine Täuschungsabsicht vorliegt.

Was stimmt: KI kann zur Manipulation verwendet werden – etwa für Deepfakes oder gezielte Meinungsbeeinflussung.

Was nicht stimmt: Die KI selbst verfolgt ein Ziel oder täuscht bewusst.

Beispiel: Ein Bild, das scheinbar dokumentarisch ist, wurde in Wahrheit vollständig KI-generiert.

Risikominimierung: Kennzeichnungspflicht einführen, Quellenverantwortung klären, kritische Inhalte nicht automatisiert erzeugen lassen.

4. KI verletzt unsere Datenschutzregeln.

↳ ***Reale Gefahr – aber kontrollierbar***

Diese Sorge ist sehr verbreitet – und nicht unberechtigt. Gerade bei öffentlich zugänglichen KI-Diensten besteht die Gefahr, dass **personenbezogene Daten, Geschäftsgeheimnisse oder interne Inhalte ungewollt weitergegeben** werden. Das betrifft jedoch nicht die KI als solche, sondern die Art und Weise, wie Tools eingesetzt und konfiguriert werden.

Was stimmt: Wer ungeschützt personenbezogene Daten in öffentlich trainierte KI-Tools eingibt, riskiert Datenschutzverstöße.

Was nicht stimmt: KI an sich ist verboten oder pauschal nicht DSGVO-konform – es kommt auf den Anwendungsrahmen an.

Beispiel: Ein Protokoll mit Mitarbeitendennamen wird über ein öffentliches GPT-Modell überarbeitet.

Risikominimierung: Klare Nutzungspolicies, Datenschutzfreigaben, sichere Unternehmens-lösungen und Zugriffsbeschränkungen einführen.

5. KI ist voreingenommen.

↳ ***Richtig – und schwer erkennbar***

Alle datenbasierten KI-Systeme übernehmen **implizite Vorurteile aus den Trainingsdaten**. Das kann sich z. B. auf Geschlecht, Herkunft, Sprache oder gesellschaftliche Rollenbilder beziehen. Die Herausforderung: Diese Verzerrungen sind oft nicht direkt sichtbar – aber sie beeinflussen Bewertungen, Texte, Entscheidungen oder Rankings.

Was stimmt: Bias ist ein reales, strukturelles Risiko in vielen KI-Systemen.

Was nicht stimmt: Es lässt sich durch reine Technikkorrekturen vollständig vermeiden.

Beispiel: Eine KI bewertet Bewerbungen subtil anders, je nach Name oder Wortwahl.

Risikominimierung: Sensibilisierung der Nutzenden, Testläufe mit Fokus auf Fairness, manuelle Kontrollinstanzen, bewusste Sprache.

6. Wir machen uns abhängig.

↳ ***Langfristig real – häufig übersehen***

Ein unterschätztes Risiko ist die **Abhängigkeit von wenigen großen Anbietern**. Viele Organisationen bauen heute produktive Arbeitsprozesse auf Tools wie ChatGPT, Copilot oder Claude – ohne vertragliche Sicherheit, Preiskontrolle oder Zugriffsgarantie. Wenn sich Bedingungen ändern, kann das zu massiven Störungen führen.

Was stimmt: Strategische Abhängigkeit ist ein reales Risiko – besonders bei fehlender Kontrolle über Infrastruktur, Lizenzen oder Daten.

Was nicht stimmt: Es gibt keine Alternativen – Organisationen können Abhängigkeiten bewusst steuern.

Beispiel: Ein Anbieter ändert plötzlich die Preisstruktur oder beschränkt bestimmte Funktionen.

Risikominimierung: Technologische Redundanz sichern, Anbieter evaluieren, Exit-Strategien vorbereiten, Schlüsselprozesse dokumentieren.

Was aus Risiken wird, entscheidet die Organisation

Die meisten **Risiken** von KI entstehen **nicht im Modell, sondern im Kontext**, in dem es genutzt wird. Genau deshalb ist es Aufgabe der Organisation, diese Kontexte klar zu gestalten: Wer darf was einsetzen? Wie wird geprüft, was verlässlich ist? Wo endet Automatisierung – und wo beginnt Verantwortung?

Nicht jedes Risiko braucht aber sofort eine Policy. Aber jedes Risiko braucht eine bewusste Einordnung: Was ist tolerierbar? Was muss abgesichert werden? Und wo wäre es schlimmer, gar nichts zu tun?

Führung heißt in diesem Zusammenhang nicht, Technik zu kontrollieren – sondern **Verantwortung so zu strukturieren, dass mit Unsicherheit professionell umgegangen werden kann**. Nicht jedes Risiko lässt sich lösen. Aber jedes lässt sich führen.

Reflexionsfragen

1. Welche konkreten Risiken bestehen im aktuellen KI-Einsatz bei uns – und sind diese systematisch adressiert?
2. Wo herrscht Unsicherheit über die Datenverwendung, Nachvollziehbarkeit oder rechtliche Zulässigkeit von KI-Anwendungen?
3. Wie kann ich in meiner Rolle zur risikobewussten Nutzung von KI beitragen – ohne Innovationspotenzial zu blockieren?

Künstliche Intelligenz weiterdenken

Dieses Briefing ist Teil einer Serie für Führungskräfte, die Orientierung rund um KI suchen. Und es ist (natürlich) in enger Zusammenarbeit mit Künstlicher Intelligenz entstanden.

Nächstes Briefing: #6 KI ist kein IT-Thema - Was Sie jetzt als Geschäftsführung verantworten

Unverbindliche Erstberatung: https://calendly.com/freuding/beratung-ki

Bisherige Briefings & weitere E-Books zum Download: ki-briefing.kit.com

Kontakt: Dr. Beate Freuding, freuding@digital-leader.eu, 0152 05188026