Versicherungsforen Leipzig

GENAI HUB Leipzig

Vorschlag für eine KI-Richtlinie

Disclaimer

Dieser Vorschlag erhebt keinen Anspruch auf Vollständigkeit oder abschließende Richtigkeit. Er dient ausschließlich als Orientierungshilfe und stellt keine Rechtsberatung dar.

Es wird ausdrücklich empfohlen, die Inhalte mit den jeweils geltenden internen Compliance-Vorgaben sowie rechtlichen Anforderungen abzustimmen und gegebenenfalls durch qualifizierte Fachstellen prüfen zu lassen.

Der vorliegende Entwurf basiert auf den Ergebnissen der Werkstatt „AI Act: KI-Einsatz rechtskonform gestalten“ vom 20./21. April 2026.

Versicherungsforen Leipzig

GENAI HUB Leipzig

Inhaltsverzeichnis

I. Ziel und Zweck

Diese Richtlinie regelt und den Einsatz von Künstlicher Intelligenz (KI) im Unternehmen zum Zweck der Steigerung von Produktivität und Qualität.

Diese Richtlinie beschreibt den verantwortungsvollen Umgang mit (Generativer-) Künstlicher Intelligenz bei ... (Unternehmen) sowie deren Entwicklung, Beschaffung, Nutzung im Unternehmen. Ziel ist es, eine effektive, transparente und ethische Nutzung von KI sicherzustellen, die sowohl den Geschäftsinteressen und Unternehmenswerten als auch den Anforderungen unserer Kunden und regulatorischen Vorgaben gerecht wird.

II. Geltungsbereich

Dieser Geltungsbereich stellt sicher, dass die Richtlinie lückenlos für alle Aktivitäten im Namen des Unternehmens Anwendung findet.

a) sachlicher Anwendungsbereich

Die Richtlinie regelt die Nutzung aller Formen von Systemen der Künstlichen Intelligenz (KI). Dies beinhaltet insbesondere:

- Generative KI-Tools (z. B. ChatGPT, Gemini, Claude, Midjourney, DALL-E).
- KI-gestützte Assistenzsysteme, die in bestehende Software integriert sind (z. B. Microsoft Copilot, Adobe Firefly, GitHub Copilot).
- Automatisierungstools und Algorithmen zur Datenanalyse oder Entscheidungsfindung.

b) personeller Anwendungsbereich

Diese Richtlinie ist verbindlich für alle Personen, die für oder im Namen von [Unternehmensname] tätig sind. Dies umfasst sämtliche festangestellten Mitarbeitenden (Voll- und Teilzeit), Führungskräfte, Auszubildende, Praktikanten sowie Werkstudierende. Ebenso gilt sie für externe Dienstleister, Honorarkräfte und Freelancer, sofern sie im Rahmen ihrer Tätigkeit Zugriff auf Unternehmensdaten haben oder Ergebnisse für das Unternehmen produzieren.

c) räumlicher Anwendungsbereich

Orts- und Geräteunabhängigkeit: Die Richtlinie gilt unabhängig davon, ob die Nutzung am Arbeitsplatz, im Homeoffice oder mobil erfolgt. Sie erstreckt sich sowohl auf unternehmenseigene IT-Infrastruktur und Hardware als auch auf private Endgeräte (Bring Your Own Device - BYOD), sofern diese für geschäftliche Zwecke oder zur Verarbeitung von Unternehmensdaten genutzt werden.

d) zeitlicher Anwendungsbereich

Die Richtlinie ist unbefristet gültig ab 01.06.2026. Sie wird regelmäßig an die technologische Entwicklung sowie aktuelle Rechtslage (z. B. AI Act) angepasst.

III. Begriffsdefinitionen

siehe Anhang

IV. Rollen und Verantwortlichkeiten

Position / Rolle	Zuordnung / Verantwortlichkeiten
Geschäftsführung	▪ Gesamtverantwortung ▪ Finanzielle und strategische Freigabe
Datenschutzbeauftragter (DSB)	▪ Datenschutzfolgeabschätzung (DSFA) ▪ Überprüfung AVV ▪ Schließung und Überprüfung von TOMs und SLAs im Rahmen der Freigabe eines KI-Systems
Informationssicherheitsbeauftragter (ISB)	▪ Informationssicherheitsbewertung eines Tools bzw. Use-Cases unter Berücksichtigung der Informationssicherheitsrichtlinien ▪ Abschätzung der Business Continuity Implikationen
Ggf. Risikomanagement	▪ Risikobewertung und Folgenabschätzung ▪ Einpreisung des Risikos
IT-Abteilung	▪ Zugänge ermöglichen ▪ Nutzerverwaltung ▪ Lizenzmanagement
Fachbereiche	▪ Festlegung der Use-Cases (am besten gleich Templates mitliefern) ▪ Ggf. Beschaffung und Vertrieb• Ggf. Einbindung anderer Fachbereiche ▪ Ggf. Monitoring der Einhaltung der Leitlinie (Compliance)
Jede/r einzelne Mitarbeitende	▪ Ist für die Einhaltung und Umsetzung der Richtlinie verantwortlich

V. Grundprinzipien

Unser Unternehmen bekennt sich bei der Einführung und Nutzung von Künstlicher Intelligenz (KI) zu folgenden ethischen und operativen Grundprinzipien:

- **Der Mensch im Mittelpunkt (Human-in-the-Loop):** KI ist ein Werkzeug zur Unterstützung unserer Arbeit, nicht zum vollständigen Ersatz menschlicher Entscheidungen. Die finale Beurteilung, Freigabe und Verantwortung für die Arbeitsergebnisse liegen stets bei den Mitarbeitenden des Unternehmens entsprechend ihrer Position im Unternehmen.
- **Transparenz und Offenheit:** Der Einsatz von KI muss nachvollziehbar sein. Wir kommunizieren ehrlich und transparent – sowohl intern als auch gegenüber unseren Kunden und Partnern –, wenn KI-Systeme maßgeblich zur Erstellung von Inhalten, Produkten oder Entscheidungen beigetragen haben.
- **Fairness und Diskriminierungsfreiheit:** Wir sind uns bewusst, dass KI-Modelle Vorurteile (Bias) aus ihren Trainingsdaten übernehmen können. KI-generierte Ergebnisse sind stets kritisch daraufhin zu prüfen, dass sie objektiv sind und niemanden benachteiligen oder diskriminieren.
- **Datensouveränität (Privacy first):** Der Schutz von Unternehmens-, Kunden- und Mitarbeiterdaten hat oberste Priorität. Wir speisen keine sensiblen Informationen in Systeme ein, bei denen der Datenabfluss nicht vertraglich ausgeschlossen ist.
- **Verantwortungsvolle Innovation:** Wir fördern Neugier und das Experimentieren mit neuen Technologien. Dies geschieht jedoch stets in einem sicheren, definierten Rahmen, um rechtliche, finanzielle oder Reputationsrisiken für das Unternehmen zu vermeiden.

VI. Risikoabschätzung und Monitoring

Inhalte einer KI-Risikoabschätzung (Assessment-Kriterien)

Wird ein neues KI-Tool für den Unternehmenseinsatz beantragt, müssen im Rahmen der Risikoabschätzung folgende Aspekte zwingend geprüft und dokumentiert werden:

- **Datenschutz und Datenverwendung:** Es muss geklärt werden, welche Datenkategorien in das System eingegeben werden (z. B. öffentliche Daten, interne Geschäftsgeheimnisse, personenbezogene Daten nach DSGVO). Zudem ist entscheidend, ob der KI-Anbieter die eingegebenen Daten (Prompts) vertraglich vom Training seiner eigenen Modelle ausschließt.
- **Informationssicherheit und Hosting:** Die IT-Sicherheit prüft die technischen Standards des Anbieters. Dazu gehören der Serverstandort (innerhalb der EU vs. Drittland), Verschlüsselungsstandards bei der Datenübertragung sowie bestehende Zertifizierungen (z. B. ISO 27001).
- **Rechtliche Konformität (AI Act):** Das System muss rechtlich eingestuft werden. Es ist zu prüfen, in welche Risikoklasse das geplante Einsatzszenario nach dem EU AI

Act fällt (z. B. minimales Risiko für reine Textgenerierung vs. Hochrisiko bei KI-gestütztem Recruiting) und welche gesetzlichen Auflagen damit einhergehen.

- **Urheberrecht und geistiges Eigentum:** Es ist zu bewerten, wie das Tool mit geistigem Eigentum Dritter umgeht und ob bei der gewerblichen Nutzung der generierten Ergebnisse (z. B. Bilder, Code) ein erhöhtes Risiko für Urheberrechtsverletzungen besteht.
- **Technische Zuverlässigkeit und Bias:** Die Prüfung umfasst das Risiko von „Halluzinationen“ (falsche Faktenausgabe) für den spezifischen Anwendungsfall. Ebenso muss bewertet werden, ob die KI bekannte Verzerrungen (Bias) aufweist, die zu diskriminierenden Ergebnissen führen könnten.
- **Geschäftliche Auswirkungen (Business Impact):** Es muss ein Worst-Case-Szenario bedacht werden: Welcher finanzielle, rechtliche oder rufschädigende Schaden könnte für das Unternehmen entstehen, wenn die KI unerkannt falsche Ergebnisse liefert oder komplett ausfällt?
- **Laufendes Monitoring (Review-Prozess):** Die Freigabe eines KI-Tools gilt nicht dauerhaft. Da sich KI-Modelle und die Nutzungsbedingungen der Anbieter schnell ändern können, führen die IT-Abteilung und die KI-Verantwortlichen regelmäßige Überprüfungen (Audits) der eingesetzten Systeme durch.
- **Meldepflicht bei Vorfällen (Incident Reporting):** Die kontinuierliche Überwachung der KI-Ausgaben im Arbeitsalltag obliegt den Nutzenden. Mitarbeitende sind verpflichtet, Auffälligkeiten oder Sicherheitsbedenken unverzüglich zu melden. Dies gilt insbesondere für:
    - Den Verdacht auf einen unbeabsichtigten Abfluss sensibler Daten.
    - Wiederkehrende “Halluzinationen” (grobe Falschinformationen) oder Leistungseinbrüche der KI.
    - Das Erkennen von systematisch diskriminierenden, voreingenommenen (Bias) oder unethischen Ergebnissen.

VII. Kommunikation und Awareness

Alle Mitarbeitenden werden mit Hilfe von Schulungen für den Umgang mit KI sensibilisiert. Die Schulungen finden mind. einmal pro Jahr statt und werden ggf. unter Einbeziehung des Change-Managements entwickelt.

weitere Inhalte:

- *Hinweis auf KI-Sprechstunde im Unternehmen und Prompting-Schulungen*

- *Eskalationskette im Schadensfall*
- *Exit-Management*

Über Änderungen an dieser Richtlinie werden die Mitarbeitenden sofort informiert.

VIII. Genehmigungs- und Prüfprozesse

Um sicherzustellen, dass der Einsatz von KI-Systemen dauerhaft im Einklang mit unseren Sicherheitsstandards und rechtlichen Vorgaben (insbesondere der DSGVO und dem EU AI Act) steht, etablieren wir einen strukturierten Prozess zur Bewertung und Überwachung dieser Technologien:

- **Vorab-Risikobewertung (Assessment):** Bevor ein neues KI-System für den geschäftlichen Einsatz freigegeben wird, muss es einen internen Prüfprozess durchlaufen. Dieser umfasst:
    - *Datenschutz und Sicherheit:* Prüfung der Informationssicherheit (z. B. Serverstandort, Datenverschlüsselung) und, falls erforderlich, die Durchführung einer Datenschutz-Folgenabschätzung (DSFA) durch den Datenschutzbeauftragten.
    - *Rechtliche Einstufung (EU AI Act):* Abgleich des geplanten Anwendungsfalls mit den gesetzlichen Risikoklassen, um sicherzustellen, dass keine unzulässigen Hochrisiko-Systeme ohne die entsprechenden Compliance-Maßnahmen eingesetzt werden.
- **Anpassung der Prozesse:** Die Erkenntnisse aus dem laufenden Monitoring sowie gemeldete Vorfälle fließen direkt in die Aktualisierung dieser Richtlinie und in die regelmäßigen Schulungsmaßnahmen für unsere Mitarbeitenden ein.

IX. Überprüfung und Weiterentwicklung

Die Entwicklung im Bereich der Künstlichen Intelligenz schreitet rasant voran. Um als Unternehmen sowohl technologisch innovativ zu bleiben als auch rechtlich und sicherheitstechnisch stets auf der sicheren Seite zu sein, wird diese Richtlinie als “lebendes Dokument” (Living Document) behandelt:

- **Fester Review-Zyklus:** Diese Richtlinie wird regelmäßig (mindestens [einmal jährlich / alle sechs Monate]) durch ein dafür zuständiges Team (z. B. IT, Datenschutz, Legal und ggf. Betriebsrat) auf ihre Aktualität, Praktikabilität und Wirksamkeit hin evaluiert.

- **Anlassbezogene Updates:** Bei disruptiven technologischen Sprüngen (z. B. der Einführung neuer autonomer KI-Agenten) oder veränderten Marktpraktiken wird die Richtlinie auch außerhalb des regulären Zyklus umgehend angepasst.
- **Feedback aus der Praxis (Mitarbeiterbeteiligung):** Die Erfahrungen der Mitarbeitenden im täglichen Umgang mit KI sind für die Weiterentwicklung essenziell. Wir fördern eine offene Feedbackkultur. Über [Kanal einfügen, z. B. Intranet/Feedback-E-Mail] können jederzeit Best Practices geteilt oder Verbesserungsvorschläge zur Richtlinie eingereicht werden.
- **Rechtliche und regulatorische Anpassungen:** Das Dokument wird kontinuierlich an neue gesetzliche Rahmenbedingungen, Gerichtsurteile und behördliche Leitlinien (insbesondere Anpassungen der DSGVO und des EU AI Act) angeglichen.

Versionierung und Überprüfungs- /Änderungshistorie

Version	Überprüfung/Änderung	Beschluss	Verantwortlich
1.0	Ersterstellung	05.05.2026	Maxi Muster

Anhang

Begriffe und Definitionen:

(EU) AI Act	(European Union) Artificial Intelligence Act (KI-Verordnung der Europäischen Union)
AI-Agent	Ein KI-Agent (AI Agent) ist ein autonomes Softwareprogramm, das auf Basis Künstlicher Intelligenz Ziele selbstständig plant, Entscheidungen trifft und Aktionen ausführt.
AVV	Auftragsverarbeitungsverzeichnis
Bias	Systematische Verzerrungen und Vorurteile in künstlicher Intelligenz, die zu diskriminierenden Ergebnissen führen.
DSB	Datenschutzbeauftragter
DSGVO	Datenschutzgrundverordnung
GenAI	Generative Künstliche Intelligenz (GenAI oder Gen-KI) ist ein Teilbereich der Künstlichen Intelligenz (KI), der darauf spezialisiert ist, eigenständig neue Inhalte, Daten oder Ideen zu erstellen, anstatt nur bestehende Daten zu analysieren
Halluzination	Überzeugend formulierte, aber inhaltlich falsche oder frei erfundene Informationen, die von generativen KI-Modellen (wie ChatGPT) generiert werden. Sie wirken oft sehr plausibel, realistisch und fachmännisch.

ISB	Informationssicherheitsbeauftragter
KI	KI ist dabei ein maschinengestütztes System, das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt ist und das nach seiner Betriebsaufnahme anpassungsfähig sein kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet, wie Ausgaben wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können.
LLM	Large Language Model (deutsch: großes Sprachmodell), ist die softwaretechnische Realisierung eines mathematischen Sprachmodells, das sich durch seine Fähigkeit zur Textgenerierung auszeichnet.
Schatten-KI	KI-Schatten-IT (Shadow AI) bezeichnet die Nutzung nicht genehmigter oder unüberwachter KI-Tools durch Mitarbeiter im Arbeitsalltag.