KI-Richtlinie im Unternehmen

(EU AI Act & ISO 42001 konform)

Die fortschreitende Entwicklung und Anwendung von Künstlicher Intelligenz (KI) bringt sowohl Chancen als auch Herausforderungen für Unternehmen mit sich. Um eine verantwortungsvolle, sichere und regelkonforme Nutzung von KI sicherzustellen, definiert diese Richtlinie grundlegende Prinzipien und Anforderungen für den Einsatz, die Entwicklung und den Betrieb von KI-Systemen im Unternehmen. Sie basiert auf internationalen Standards wie der ISO/IEC 42001 (Managementsystem für KI) und orientiert sich an den Vorgaben des EU AI Act sowie weiteren relevanten Normen und ethischen Leitlinien.
Die Richtlinie stellt sicher, dass rechtliche und regulatorische Anforderungen erfüllt werden und bewährte Verfahren aus den Bereichen Risikomanagement, Datenschutz, Sicherheit und Transparenz in den gesamten KI-Lebenszyklus integriert sind. Insbesondere werden die Risikoklassifizierungen des EU AI Act berücksichtigt, sodass für Hochrisiko-KI-Systeme angemessene Schutzmaßnahmen und Qualitätsstandards gewährleistet sind. Durch einen strukturierten und systematischen Ansatz wird eine vertrauenswürdige und zukunftssichere KI-Strategie etabliert, die Innovation mit ethischer und rechtlicher Verantwortung vereint.

Zweck und Geltungsbereich

Diese KI-Richtlinie legt fest, wie unser Unternehmen Künstliche Intelligenz verantwortungsvoll einsetzt, entwickelt und betreibt, um den Anforderungen der EU AI Verordnung (AI Act) sowie der ISO/IEC 42001 (Managementsystem für KI) gerecht zu werden. Sie gilt für alle Geschäftsbereiche, Projekte und Prozesse, in denen KI-Systeme entworfen, implementiert oder genutzt werden. Durch diese Richtlinie stellen wir sicher, dass rechtliche Vorgaben eingehalten werden und Best Practices aus Risikomanagement, Datenschutz und Ethik berücksichtigt sind.
Insbesondere werden die Risikoklassifizierungen des EU AI Act angewendet (von geringem bis hohem Risiko), und für Hochrisiko-KI-Systeme erfüllen wir sämtliche Sicherheitsanforderungen, Dokumentationspflichten und Qualitätsstandards. Die Richtlinie orientiert sich neben EU-Recht und ISO 42001 auch an ISO 31000 (Risikomanagement), ISO/IEC 27701 (Datenschutz- Management), den OECD-Prinzipien für KI sowie dem NIST • AI Risk Management Framework, um eine ganzheitliche, zukunftssichere KI-Strategie zu gewährleisten.

Grundsätze und Leitlinien

Wir verpflichten uns zu grundlegenden Prinzipien für den vertrauenswürdigen Einsatz von KI-Rechtmäßigkeit & Compliance: Alle KI-Anwendungen müssen im Einklang mit geltenden Gesetzen und Regulierungen stehen (insb. EU AI Act, DSGVO) sowie mit anerkannten Standards (ISO 42001 u.a.). Verbotene KI-Praktiken (z.B. manipulative oder diskriminierende KI gemäß EU AI Act) werden strikt ausgeschlossen.

- Ethik & Menschzentrierung: KI-Systeme sollen menschliche Werte und Grundrechte achten. Wir folgen den OECD-KI-Prinzipien und fördern fairen, nicht-diskriminierenden KI-Einsatz, Transparenz und Erklärbarkeit, Robustheit und Sicherheit sowie Rechenschaftspflicht. KI darf keine ungerechtfertigten Nachteile für Einzelne oder Gruppen verursachen.
- Risikobasierter Ansatz: Entwicklung und Einsatz von KI richten sich nach dem Risikoniveau. Bei hohem Risikogelten strenge Kontrollen und Freigabeprozesse. Wir betreiben proaktives Risikomanagement über den gesamten Lebenszyklus jeder KI-Lösung. Identifizierte Risiken (z.B. Bias, Fehlentscheidungen, Sicherheitslücken) werden frühzeitig adressiert.
- Transparenz & Verantwortung: Entscheidungen über KI-Einsatz werden dokumentiert und offengelegt. Nutzer und Betroffene sollen angemessen informiert werden, wenn sie mit einem KI-System interagieren. Klare Verantwortlichkeiten werden definiert - letztlich bleibt ein menschlicher Verantwortlicher für die Ergebnisse von KI-Systemen rechenschaftspflichtig.
- Datenschutz & Sicherheit: KI-Systeme müssen Datenschutz by Design umsetzen und technisch sowie organisatorisch abgesichert sein. Personenbezogene Daten werden nur im

nötigen Umfang und gemäß DSGVO verarbeitet. Cybersicherheit Maßnahmen schützen KI-Systeme vor Manipulation und unbefugtem Zugriff.

- Qualität & Kontinuierliche Verbesserung: Wir streben eine hohe Qualität in Daten, Modellen und Ergebnissen an. Durch ein formales KI-Qualitätsmanagementsystem (angelehnt an ISO 42001) gewährleisten wir laufende Überwachung, Evaluierung und Verbesserung unserer KI-Systeme. Eine Kultur des Lernens aus Fehlern und regelmäßige Reviews stellen sicher, dass wir uns an neue Erkenntnisse und Vorgaben anpassen.

Organisation, Verantwortung und Governance

Die Unternehmensleitung übernimmt die Verantwortung für KI-Governance und stellt die nötigen Ressourcen bereit. Ein internes KI-Lenkungsgremium (oder KI-Verantwortlicher) wird benannt, um die Umsetzung dieser Richtlinie zu steuern. Dieses Gremium definiert Rollen und Pflichten klar - von der Entwicklung über den Betrieb bis zur Aufsicht über KI-Systeme.

- Führung und Kultur: Das Management verankert ethische KI-Grundsätze in der Unternehmenskultur und fördert Bewusstsein für KI-Risiken. Führungskräfte stellen sicher, dass alle Abteilungen (IT, Fachbereiche, Recht, Datenschutz etc.) eingebunden werden.
- Rollen und Zuständigkeiten: Konkrete Verantwortliche werden bestimmt für Bereiche wie Risikomanagement, Datenschutz, IT-Sicherheit und Compliance im Kontext von KI. Beispielsweise überwacht der Datenschutzbeauftragte KI-Anwendungen mit Personenbezug, und ein KI-Projektleiter verantwortet die Risikobewertung und Dokumentation eines neuen KI-Systems.
- Schulung und Kompetenz: Wir stellen sicher, dass Mitarbeitende ausreichend im Umgang mit KI geschult sind und die Richtlinie kennen. Ab Februar 2025 fordert der EU AI Act sogar explizit, KI-Kompetenzen im Unternehmen aufzubauen. Daher führen wir regelmäßige Trainings zu KI-Risiken, ethischen Leitlinien und den richtigen Umgang mit KI- Tools durch.
- Einbindung von Stakeholdern: Relevante Stakeholder - intern (Management, Betriebsrat, Fachteams) und bei Bedarf extern (Aufsichtsbehörden, Kundenvertretungen) - werden in die Entwicklung und Prüfung von KI-Systemen einbezogen. So stellen wir Mehr-Perspektiven sicher, wie vom NIST-Rahmen gefordert („Einbeziehung aller betroffenen Parteien in die Bewertung von KI-Risiken").
- Externe Partner und Lieferanten: Wenn wir KI-Komponenten oder -Daten von Dritten beziehen, verpflichten wir diese vertraglich auf unsere KI-Standards. Das Lieferantenmanagement stellt sicher, dass Drittanbieter unsere Prinzipien teilen und vergleichbare Kontrollen anwenden. Zulieferer von KI-Systemen müssen z.B. ihre Trainingsdaten offenlegen und nachweisen, dass keine unvertretbaren Biases vorliegen.
- Dokumentation & Berichtswesen: Es wird ein zentrales Verzeichnis aller KI-Systeme geführt, inkl. Zweck, Verantwortlichen und Risikokategorie. Wichtige Entscheidungen (Freigaben, Ausnahmegenehmigungen, Vorfälle) werden dokumentiert. Das Lenkungsgremium berichtet der Geschäftsführung regelmäßig zum Status der KI-Compliance und etwaigen Risiken.

Risikomanagement entlang des KI-Lebenszyklus

Ein systematisches Risikomanagement ist Kernbestandteil dieser Richtlinie. Gemäß EU AI Act müssen Anbieter Hochrisiko-KI ein Risikomanagementsystem über den gesamten Lebenszyklus etablieren. Wir setzen dies für alle kritischen KI-Systeme um und orientieren uns dabei an ISO 31000 und dem NIST AI Risk Management Framework.

- Identifizierung von Risiken: Bereits in der Planungsphase eines KI-Projekts werden potenzielle Risiken und Auswirkungen identifiziert (z.B. Fehlerrisiken, Verzerrungen, Auswirkungen auf Benutzer oder Betroffene). Wir betrachten technische Risiken und gesellschaftliche Auswirkungen (ISO 42001 fordert z.B. ein KI-Impact-Assessment für Nutzer und Umfeld). Auch durch den EU AI Act geforderte Auswirkungen auf Grundrechte werden in die Bewertung einbezogen.

- Risikobewertung: Für jedes identifizierte Risiko erfolgt eine Analyse und Einschätzung der Eintrittswahrscheinlichkeit und Schwere der potenziellen Schäden. Dieser Prozess folgt einem etablierten Schema (Kontext festlegen, Risiko identifizieren, analysieren, bewerten). Risiken werden kategorisiert (akzeptabel, tolerierbar mit Maßnahmen, inakzeptabel).
- Risikominimierung: Wo immer möglich ergreifen wir Maßnahmen zur Behandlung erkannter Risiken. Beispielsweise werden Trainingsdaten bereinigt, wenn Bias festgestellt wird oder zusätzliche Sicherheitsmechanismen eingebaut, falls Ausfallrisiken bestehen. Für hochriskante KI-Systeme sind laut AI Act ausdrücklich geeignete Steuerungs- und Minderungsmaßnahmen vorzusehen. Unvertretbare Restrisiken führen dazu, dass ein KI-System nicht eingesetzt oder nur mit engen Auflagen betrieben wird.
- Dokumentation der Risiken: Alle Ergebnisse der Risikoanalyse sowie implementierte Gegenmaßnahmen werden nachvollziehbar dokumentiert. Dies ist Teil der technischen Dokumentation, die der EU AI Act für Hochrisiko-Systeme fordert. Sie dient intern als Entscheidungs-grundlage und kann auf Anforderung auch Aufsichtsbehörden vorgelegt werden.
- Kontinuierliches Monitoring: Risiken werden über den gesamten Lebenszyklus hinweg überwacht. Wir überprüfen regelmäßig, ob angenommene Risiken sich verändern (z.B. durch neue Nutzungsszenarien oder Updates des KI-Modells) und passen die Bewertung entsprechend an. Ein Frühwarnsystem (z.B. automatisierte Logs-Analysen, Benutzerfeedback) soll neue Risiken oder Vorfälle schnell erkennen lassen.
- Notfallplan: Für den Umgang mit realisierten Risiken (Vorfällen) gibt es definierte Prozesse: Meldewege, Incident-Response-Teams und Maßnahmenkataloge, um Schäden einzudämmen. Beispielsweise existiert ein Eskalationsprozess, falls ein KI-System Fehlentscheidungen mit Kundenimpact produziert - das System kann abgeschaltet und das Problem analysiert werden, bevor es wieder in Betrieb geht.
- Durch dieses proaktive Risikomanagement stellen wir sicher, dass KI-Systeme sicher, nachvollziehbar und unter Kontrolle bleiben. Es ist integraler Bestandteil unseres KI-Qualitätsmanagementsystems, wie es ISO 42001 vorgibt, und verzahnt mit dem allgemeinen Unternehmens-Risikomanagement.

Datenmanagement und Datenschutz

Ein verantwortungsvoller Umgang mit Daten ist Voraussetzung für vertrauenswürdige KI. Diese Richtlinie fordert strikte Daten-Governance und Datenschutz bei allen KI-Anwendungen:

- Datenqualität und Repräsentativität: Wir stellen sicher, dass Trainings-, Validierungs- und Testdaten für KI-Systeme relevant, hinreichend repräsentativ und möglichst fehlerfrei sind. Daten werden vor Verwendung bereinigt (Cleansing) und auf Bias überprüft. Etwaige Datenverzerrungen (Bias gegen bestimmte Gruppen) werden analysiert und durch Anpassung der Datenauswahl oder -gewichtung gemindert, um faire Modelle zu erhalten.
- Datengovernance-Prozess: Es existieren definierte Prozesse für die Auswahl, Beschaffung und Pflege von KI-Trainingsdaten. Die Verantwortlichkeiten für Datensätze sind benannt. Jede Änderung an kritischen Datensätzen (z.B. neues Dataset, andere Quellen) muss freigegeben und dokumentiert werden.
- Privacy by Design: Sobald KI-Systeme personenbezogene Daten verarbeiten, gelten besonders strenge Anforderungen. Wir halten uns an die DSGVO und implementieren ein Privacy Information Management System (PIMS) gemäß ISO/IEC 27701, um Risiken für persönliche Daten gezielt zu steuern. Bereits bei der Entwicklung wird das Prinzip Datensparsamkeit verfolgt: Es werden nur die Daten erhoben und genutzt, die zwingend nötig sind. Pseudonymisierung oder Anonymisierung kommen zum Einsatz, wo immer möglich.
- Einwilligung und Betroffenenrechte: Falls KI-Systeme persönliche Daten von Kunden verarbeiten, holen wir erforderliche Einwilligungen ein oder stellen sicher, dass eine andere Rechtsgrundlage greift. Die Betroffenenrechte (Auskunft, Löschung etc.) sind gewährleistet - etwa indem wir Daten in KI-Modellen löschbar halten oder alternative Lösungen anbieten, falls

jemand der KI-Nutzung widerspricht.

- Schutz sensibler Daten: Besondere Kategorien personenbezogener Daten (z.B. Gesundheitsdaten) werden nur in eng abgegrenzten Fällen von KI-Systemen genutzt und unterliegen zusätzlichen Kontrollen (Verschlüsselung, isolierte Verarbeitung). KI-Systeme dürfen keinesfalls unerlaubt sensible Attribute wie ethnische Herkunft oder politische Meinung der Nutzer erhalten.
- Datensicherheit: Alle Daten, die in KI-Systemen verwendet werden, werden angemessen gegen Verlust, Manipulation und unbefugten Zugriff gesichert. Hier gelten unsere generellen IT-Sicherheitsrichtlinien (z.B. Zugriffskontrollen, regelmäßige Backups) weiter. Insbesondere Trainingsdaten, die aus externen Quellen stammen, werden vor Nutzung auf Integrität geprüft (um z.B. Datenvergiftungsangriffe zu verhindern).
- Protokollierung und Nachverfolgbarkeit: Änderungen an Datensätzen und Datenflüssen werden versioniert und protokolliert. So ist stets nachvollziehbar, welche Datenbasis ein Modell trainiert hat. Diese Nachverfolgbarkeit der Datenherkunft ist Teil der technischen Dokumentation und unterstützt sowohl Qualitätssicherung als auch Compliance-Nachweise.
- Durch diese Maßnahmen stellen wir sicher, dass KI-Systeme datenschutzkonform und mit qualitativ hochwertigen, biasarmen Daten arbeiten, was Grundlage für zuverlässige und ethische KI-Ergebnisse ist.

Entwicklung und Test von KI-Systemen

In der Entwicklungsphase legen wir durch professionelle Verfahren den Grundstein für sichere, robuste und nachvollziehbare KI-Systeme. Die Richtlinie fordert für Design, Implementierung und Test von KI insbesondere:

- Anforderungsdefinition: Vor Start der Entwicklung werden klare Ziele, Anwendungsgrenzen und Leistungsanforderungen definiert. Dazu gehören auch die vom EU AI Act verlangten Kriterien wie Genauigkeit, Robustheit und Cybersicherheit des KI-Systems. Wir legen fest, welche Genauigkeit z.B. ein Modell mindestens erreichen muss und wie es auf fehlerhafte Eingaben reagieren soll.
- Technische und ethische Spezifikationen: Für jedes KI-Projekt wird eine Spezifikation erstellt, die neben funktionalen Anforderungen auch ethische und regulatorische Vorgaben einbezieht. Etwa wird festgelegt, dass das Modell keine geschützten Merkmale als Entscheidungsgrundlage nutzt (Anti-Diskriminierung) und dass eine menschliche Kontrollmöglichkeit eingeplant wird.
- Secure Development Lifecycle: Die Entwicklung erfolgt in kontrollierten Umgebungen mit Qualitätssicherungs-Gates. Code und Modelle werden peer-reviewed. Sicherheitsrichtlinien (z.B. keine Verwendung unsicherer Bibliotheken) werden eingehalten. Dies ähnelt dem Vorgehen bei ISO 27001, nun aber erweitert um KI-Aspekte.
- Modelltraining und -auswahl: Beim Training von ML-Modellen stellen wir sicher, dass Überanpassung vermieden wird und das Modell generalisiert. Verschiedene Modellvarianten werden verglichen, und diejenige mit dem besten Trade-off aus Genauigkeit und Erklärbarkeit wird gewählt.

Test und Validierung: Vor Einsatz durchlaufen KI-Systeme eine umfangreiche Testphase. Wir testen mit repräsentativen Szenarien und Edge-Cases, um Verhalten und Fehlerraten zu prüfen. Maßgebliche Vertrauenswürdigkeits-Merkmale (z.B. Fairness, Transparenz) werden überprüft. Für Hochrisiko-Systeme lassen wir ggf. externe Gutachten oder Audits durchführen. Tests umfassen:

- Funktionale Tests: Erfüllt das System die fachlichen Anforderungen?
- Leistungstests: Erreicht das Modell die definierte Genauigkeit und Robustheit?
- Sicherheitstests: Ist das System vor Cyber-Angriffen und Manipulation geschützt? (z.B. Resistenz gegen adversarial examples)
- Usability-Tests: Kann ein Mensch die Ergebnisse verstehen und sinnvoll mit dem

System interagieren?

- Bias- und Fairness-Prüfung: Wir führen spezielle BIA-Tests durch, um Ungleichbehandlungen aufzudecken. Zeigt sich z.B., dass ein Algorithmus systematisch bei bestimmten Gruppen schlechter funktioniert, werden die Ursachen analysiert und behoben (z. B. durch mehr Trainingsdaten für diese Gruppen oder Anpassung des Modells).
- Erklärbarkeit: Soweit möglich werden Modelle so entwickelt oder ergänzt, dass Erklärungen für ihre Ausgabengegeben werden können (z.B. durch Feature Importance, XAI-Methoden). Gerade bei Entscheidungs-KI in sensiblen Bereichen ist eine nachträgliche Erklärung des Ergebnisses wichtig, um Transparenz für Betroffene herzustellen.
- Technische Dokumentation: Zu jedem KI-System wird eine umfassende Dokumentation geführt, die Aufbau, Funktionsweise und Testnachweise enthält. Diese technische Dokumentation dient dem Nachweis der Compliance und muss bei Hochrisiko-KI alle im EU AI Act geforderten Informationen abdecken. Darin beschrieben sind u.a. Systemzweck, Algorithmus typ, Datenquellen, erzielte Genauigkeit, bekannte Restrisiken, implementierte Sicherheitsmechanismen, Ergebnisse der Tests sowie Anleitung zur sicheren Nutzung.
- Protokollierung im Design: Das KI-System wird so entworfen, dass es relevante Ereignisse automatisiert protokolliert (z.B. Entscheidungen oder Abweichungen). Logging ist bereits in die Softwarearchitektur integriert, um im Betrieb wichtige Informationen für Fehlersuche und Risikoüberwachung verfügbar zu haben.
- Menschliche Kontrollierbarkeit: Wir designen KI-Systeme immer mit "Human-in-the-Loop" oder "Human-on-the-loop" - d.h. es gibt Möglichkeiten für menschliche Eingriffe oder zumindest fortlaufende Überwachung während der Nutzung. Kein Hochrisiko-System agiert völlig autonom ohne definierte Eingriffsmöglichkeiten. Beispielsweise könnten kritische Parameter im Betrieb von verantwortlichen Personen justiert oder Ergebnisse vor endgültiger Umsetzung von einem Menschen freigegeben werden.
- Freigabeprozess: Ein KI-System wird erst dann zum Einsatz freigegeben, wenn alle obigen Punkte erfüllt und dokumentiert sind, die Risikoabwägung positiv ausfällt und - falls erforderlich - eine externe Konformitätsbewertung (Audit) erfolgt ist. Die Freigabe erteilt das KI-Lenkungsgremium bzw. die Geschäftsleitung.

Durch diesen strengen Entwicklungs- und Testprozess erfüllen wir Sicherheitsanforderungen schon bei der KI-Entwicklung und bauen vertrauenswürdige Systeme, die den Vorgaben des EU AI Act und der ISO 42001 genügen.

Einsatz und Betrieb von KI-Systemen

Auch nach der Entwicklung stellen wir durch klare Regeln sicher, dass KI im Betrieb kontrolliert und sicher bleibt. Wichtige Vorgaben für die Nutzungsphase sind:

Konformitätsbewertung vor Inbetriebnahme: Für Hochrisiko-KI-Systeme verlangen EU-Vorgaben eine Konformitätsprüfung, bevor sie auf den Markt oder in Betrieb gehen. Wir führen daher ggf. eine interne oder externe Abnahmeprüfung durch: Überprüfung, ob alle Anforderungen, Dokumentationen und Risikokontrollen umgesetzt sind. Falls notwendig, wird das System in der EU-Datenbank für KI-Systeme registriert (sofern vom AI Act für die jeweilige Anwendung vorgesehen).

- Nutzungsrichtlinien und Anleitungen: Vor dem Ausrollen eines KI-Systems erstellen wir klare Anweisungen für die Anwender. Diese Gebrauchsanleitungen enthalten Hinweise zur korrekten Nutzung, Grenzen der Systemleistung, Überwachungspflichten des Nutzers und Verfahren bei Fehlfunktionen. Externe Kunden erhalten transparente Benutzerhinweise; interne Nutzer werden geschult, das System gemäß den Vorgaben einzusetzen.
- Überwachung im Betrieb: Jedes produktive KI-System unterliegt einer laufenden Überwachung durch benannte Verantwortliche (z.B. System Owner im Fachbereich).
- Wir verfolgen definierte Metriken (Accuracy, Antwortzeiten, Fehlerquoten, etc.), um die Performance und Zuverlässigkeit im Feld zu überprüfen. Abweichungen oder Trends (z.B.

allmählicher Leistungsabfall durch Datenverschiebung) werden erkannt, bevor sie kritisch werden. Bei Hochrisiko-Anwendungen kann zusätzlich ein Monitoring-Dashboard mit Warnmeldungen eingerichtet werden.

- Logging und Aufbewahrung: Relevante Systemereignisse werden auch im Betrieb konsequent geloggt. Für Hochrisiko-KI fordern die Regeln eine Archivierung von Logs, um Entscheidungen nachträglich nachvollziehen zu können. Wir bewahren Protokolle gemäß den vorgeschriebenen Fristen (mindestens 6 Monate oder länger, je nach Anwendung) sicher auf. Die Logs werden regelmäßig ausgewertet, um Verbesserungsbedarf zu erkennen.
- Menschliche Eingriffsmöglichkeiten: Im Betrieb ist stets ein verantwortlicher Mensch in der Lage, das KI-System zu überstimmen, auszusetzen oder zu deaktivieren, falls es zu unerwartetem Verhalten kommt. Prozesse definieren, in welchen Situationen z.B. ein Fail-Safe ausgelöst wird oder ein Alarm an einen Operator geht. Gerade in kritischen Anwendungen (Medizin, Fertigung, etc.) sind klare Eskalationspläne hinterlegt, damit im Zweifel der Mensch die Kontrolle übernimmt.
- Vorfallmanagement: Tritt ein Fehler, Zwischenfall oder ein potenziell schädliches Ergebnis eines KI-Systems auf, greift unser Incident-Response-Plan. Mitarbeiter melden Vorfälle umgehend an das KI-Governance-Team. Wir analysieren die Ursache (etwa neue Art von Eingabedaten, die das Modell fehlleitet) und ergreifen Korrekturmaßnahmen. Das kann ein sofortiger Patch, eine Anpassung des Modells oder notfalls das Abschalten des Systems sein. Schwere Vorfälle werden dokumentiert und der Geschäftsführung gemeldet; falls meldepflichtig, informieren wir Behörden und Betroffene in der vorgeschriebenen Weise.
- Wartung und Updates: KI-Systeme werden regelmäßig gewartet und aktualisiert. Dazu gehört, Modelle bei Bedarf neu zu trainieren (z.B. jährlich oder bei signifikanten Datenänderungen), um Genauigkeit zu erhalten. Änderungen am System (Updates, neue Daten, geänderte Parameter) durchlaufen erneut Tests und Freigaben gemäß Entwicklungsrichtlinie. Jede wesentliche Änderung an einem Hochrisiko-KI-System wird außerdem in der Dokumentation vermerkt (Änderungshistorie) und kann eine neue Konformitätsbewertung erfordern.
- Kontinuierliche Risiko-Neubewertung: Im Betrieb evaluieren wir periodisch, ob sich das Risiko-Profil eines KI-Systems verändert hat. Neue Nutzungszwecke oder Erkenntnisse aus dem Betrieb fließen in eine aktualisierte Risikoanalyse ein. Ggf. werden zusätzliche Schutzmaßnahmen implementiert, um neuen Risiken zu begegnen. Diese laufende Anpassung entspricht dem Check-Act Zyklus von ISO 42001, der fordert, Beobachtungen im Betrieb auszuwerten und Korrekturmaßnahmen abzuleiten.

Durch engmaschige Überwachung und definierte Betriebsprozesse stellen wir sicher, dass KI-Systeme verlässlich funktionieren und bei Problemen schnell eingegriffen wird. Somit bleibt die Nutzung der KI unter Kontrolle und im Einklang mit regulatorischen Vorgaben.

Transparenz und Erklärbarkeit

Transparenz ist ein Schlüsselfaktor für vertrauenswürdige KI. Unser Unternehmen sorgt dafür, dass sowohl die Nutzung von KI-Systemen als auch deren Entscheidungen im angemessenen Rahmen transparent sind:

- Kennzeichnung von KI-Systemen: Wann immer Personen mit einem KI-System interagieren, machen wir dies deutlich. Beispielsweise werden KI-Chatbots oder - Assistenten als solche kenntlich gemacht, sodass Nutzer wissen, dass sie es mit einer Maschine zu tun haben. Inhalte, die vollständig durch KI generiert wurden, kennzeichnen wir als KI-generiert, wo es angebracht oder rechtlich gefordert ist (z.B. Kennzeichnung von synthetischen Medien, Deepfakes).
- Offenlegung gegenüber Kunden: In unseren Produkten und Dienstleistungen informieren wir Kunden über den Einsatz von KI-Technologien, insbesondere wenn dies für die Servicequalität relevant ist oder persönliche Daten involviert sind. Die Funktion und Zweck des KI-Systems werden in verständlicher Sprache erläutert. Bei kritischen Anwendungen (etwa KI

in Entscheidungsprozessen über Kredite, HR etc.) bieten wir auf Nachfrage zusätzliche Informationen an, wie ein Ergebnis zustande kam (im Rahmen des technisch Möglichen).

- Erklärbarkeit von Entscheidungen: Wir streben an, entscheidungsunterstützende KI so zu gestalten, dass Mitarbeiter oder Kunden eine Begründung erhalten können. Wo vollautomatisierte Entscheidungen mit erheblicher Auswirkung getroffen werden, erfüllen wir die gesetzlichen Auskunftspflichten und erklären die Hauptfaktoren der Entscheidung. Z.B. könnte eine Kreditentscheidung durch KI mit den wichtigsten Einflussgrößen begründet werden. Vollständige Explainability ist nicht immer erreichbar, aber wir gewährleisten zumindest Transparenz über die Eingabedaten und das Ziel des Modells.
- Interne Nachvollziehbarkeit: Für alle KI-Modelle halten wir intern detaillierte Dokumentation bereit (siehe technische Dokumentation), auf die berechtigte Personen Zugriff haben. So ist für Audits oder Überprüfungen stets nachvollziehbar, wie das System funktioniert und konfiguriert ist.
- Kommunikation mit Stakeholdern: Wir pflegen eine offene Kommunikation zu unserem KI- Einsatz. Intern werden die Belegschaft und die Mitbestimmungsgremien regelmäßig über KI- Projekte informiert. Extern suchen wir den Dialog mit Regulierungsbehörden, Brancheninitiativen und ggf. der Öffentlichkeit, um Vertrauen aufzubauen. Sofern erforderlich oder sinnvoll, veröffentlichen wir Transparenzberichte über die Nutzung von KI im Unternehmen.
- Feedbackmöglichkeiten: Nutzer und Mitarbeiter haben Kanäle, um Bedenken oder Fragen zu KI-Systemen zu adressieren. Wir ermutigen dazu, etwaige unfaire Ergebnisse oder Fehler zu melden. Dieses Feedback nutzen wir, um die Systeme weiter zu verbessern.

Mit diesen Transparenzmaßnahmen unterstützen wir die Nachvollziehbarkeit und Überprüfbarkeit unseres KI-Einsatzes. Sie entsprechen internationalen Empfehlungen für vertrauenswürdige KI (z.B. OECD fordert Transparenz und Erklärbarkeit) und helfen, Akzeptanz bei Nutzern und Öffentlichkeit zu schaffen.

Sicherheit und Robustheit der KI-Systeme

Die IT-Sicherheit und Robustheit unserer KI-Systeme hat höchste Priorität, um Zuverlässigkeit und Schutz vor Missbrauch zu gewährleisten:

- Cybersecurity Maßnahmen: Alle KI-Systeme werden in die bestehende IT-Sicherheitsarchitektur integriert. Das bedeutet: aktuelle Sicherheitsupdates, Firewalls, Zugangskontrollen und Überwachung auf Sicherheitsvorfälle gelten genauso für KI-Server und - Dienste wie für andere kritische Systeme. Bei cloudbasierten KI-Diensten achten wir auf die Einhaltung von Sicherheitsstandards seitens der Anbieter (ggf. Zertifizierungen).
- Schutz vor unbefugtem Zugriff: Insbesondere Trainingsdaten und Modelle werden vor unautorisiertem Zugriff geschützt. Zugriff erhalten nur befugte Personen nach dem Need-to-know-Prinzip. Sensible Modelle, die einen Wettbewerbsvorteil darstellen, werden durch geeignete Mechanismen (z.B. IP-Schutz, technische Schutzmaßnahmen gegen Modell-Exfiltration) gesichert.
- Robustheit gegen Angriffe: Wir testen KI-Modelle gezielt auf Robustheit. Dazu gehören adversarial testing(Überprüfung, ob kleine Eingabeänderungen zu drastischen falschen Outputs führen) und Belastungstests. Erkenntnisse daraus fließen in Modellverbesserungen ein, um die Resistenz gegen Manipulation oder fehlerhafte Eingaben zu steigern.
- Fehlertoleranz: KI-Systeme werden so gestaltet, dass sie bei Ausfällen oder unerwartetem Verhalten kontrolliert, degradieren, statt chaotisch zu versagen. Beispielsweise greift ein Fallback-Mechanismus oder das System verlangt menschliche Bestätigung, wenn Unsicherheiten zu groß werden.
- Überwachung von Sicherheitsrisiken: Das Security-Team überwacht kontinuierlich aktuelle Sicherheitsbedrohungen im KI-Bereich (z.B. neue Attacken auf ML-Modelle) und beurteilt, ob unsere Systeme verwundbar sind. Gegebenenfalls werden Patches oder neue Schutzmaßnahmen implementiert.

- Notfallpläne bei Sicherheitsvorfällen: Für den Fall eines sicherheitsrelevanten Vorfalls (z. B. Hackerangriff auf ein KI-System, Datendiebstahl) liegen Incident Response Pläne vor, die in unsere generellen Cybersicherheit-Notfallpläne integriert sind. Enthalten sind Verantwortlichkeiten (z.B. Information Security Officer, KI-Entwicklung, PR) und Schritte zur Eindämmung, forensischen Untersuchung, Benachrichtigung und Wiederherstellung des Normalbetriebs.
- Auditierung und Penetrationstests: Wir unterziehen kritische KI-Anwendungen regelmäßigen Sicherheitsauditsund ggf. Penetrationstests, um Schwachstellen aufzudecken. Dabei prüfen interne oder externe Experten sowohl die umgebende Infrastruktur als auch KI-spezifische Aspekte (etwa ob das Modell gegen Adversaria Exempels gehärtet ist). Ergebnisse werden dem KI-Governance-Team berichtet und zeitnah abgearbeitet.

Diese Maßnahmen gewährleisten, dass unsere KI-Systeme nicht nur im Labor, sondern auch in der realen Betriebsumgebung robust und sicher funktionieren. Indem wir Cybersicherheit und KI-Risiken gemeinsam betrachten, erfüllen wir die Forderung des EU Al Act nach angemessener Widerstandsfähigkeit und Cybersicherheit von Hochrisiko-KI und schützen unser Unternehmen sowie die Nutzer vor Schaden.

Qualitätsmanagement, Überwachung und Verbesserung

Um die Einhaltung dieser Richtlinie sicherzustellen und unsere KI-Nutzung stetig zu optimieren, betreiben wir ein kontinuierliches KI-Qualitätsmanagement. Dies orientiert sich an ISO/IEC 42001 (KI-Managementsystem) und verzahnt sich mit bestehenden Managementsystemen (z.B. ISO 9001 Qualitätsmanagement, ISO 27001 1nformationssicherheit):

- Qualitätsmanagementsystem (QMS) für KI: Für Hochrisiko-KI-Systeme verlangt der EU Al Act explizit ein Qualitätsmanagementsystem mit definierten Richtlinien und Verfahren. Wir haben ein solches KI-Managementsystem (AIMS) etabliert, das Prozesse für Qualitätssicherung, Überwachung und Risikomanagement rund um KI beinhaltet. Das KI-QMS dokumentiert alle relevanten Verfahren, von Datengovernance über Entwicklungs-Workflows bis hin zu Incident-Management, und stellt deren Umsetzung sicher. Viele Anforderungen des Al Act an ein QMS decken sich mit ISO 42001, sodass unsere Orientierung an ISO 42001 eine hohe Compliance gewährleistet.
- Leistungsbewertung und interne Audits: Wir führen in regelmäßigen Abständen Überprüfungen der KI-Systeme und der gesamten KI-Governance durch. Ein internes Auditteam oder externe Prüfer evaluieren, ob Vorgaben eingehalten werden und wie wirksam die Kontrollen sind. ISO 42001 verlangt z.B. Verfahren zur Innenrevision, um die Wirksamkeit der KI-Maßnahmen zu messen. Auditfeststellungen werden dokumentiert und an das KI-Lenkungsgremium berichtet. Bei Abweichungen oder Schwachstellen werden Korrekturmaßnahmen initiiert.
- Management Review: Die oberste Leitung erhält mindestens jährlich einen Bericht über das KI-Managementsystem und die Performance unserer KI-Anwendungen. In diesem Managementreview werden Fortschritte, Risiken, Vorfälle und geplante Verbesserungen diskutiert. Die Geschäftsführung bewertet, ob Ressourcen oder Strategien angepasst werden müssen. Dieses Top-Management-Engagement stellt sicher, dass KI-Compliance und -Ethik auf höchster Ebene verankert bleiben.
- Kontinuierliche Verbesserung: Wir fördern eine Kultur der kontinuierlichen Verbesserung in allen KI-Prozessen. Basierend auf Auditergebnissen, Vorfällen, neuen Technologien oder geänderten Vorschriften passen wir die Richtlinie und die Maßnahmen regelmäßig an. Verbesserungen können z.B. aktualisierte Schulungen, optimierte Datenprozesse oder neue Monitoring-Tools sein. ISO 42001 folgt dem Plan-Do-Check-Act-Zyklus, was wir in unserer KI-Governance umsetzen: planen (Richtlinie und Maßnahmen festlegen}, umsetzen, prüfen (Monitoring/Audits) und verbessern.
- Dokumentation und Nachweisführung: Alle relevanten Dokumente (Richtlinie, Verfahren, Risikoberichte, technische Dokumentationen, Auditberichte) werden versioniert

aufbewahrt. Somit können wir jederzeit Nachweise unserer Compliance erbringen - intern gegenüber Audit und extern gegenüber Aufsichtsbehörden oder Kunden.

- Zertifizierung und externe Orientierung: Bei Bedarf streben wir externe Zertifizierungen an, z.B. eine ISO/IEC 42001 Zertifizierung unseres KI-Managementsystems, um unsere Bemühungen nach außen sichtbar zu machen. Dies schafft Vertrauen bei Geschäftspartnern und Kunden, da eine unabhängige Stelle die Verantwortungsvolle KI-Praxis unseres Unternehmens bestätigt.
- Zukünftige Regulierungen: Wir beobachten aktiv die Weiterentwicklung von Gesetzen, Normen und Leitlinien im KI-Bereich. Neue Anforderungen (z.B. zukünftige Anpassungen des AI Act, nationale KI-Regeln oder neue ISO-Normen) werden rechtzeitig in unsere Richtlinie integriert. Durch die frühe Implementierung von ISO 42001 sind wir gut vorbereitet aufkommende Vorschriften - Unternehmen, die bereits ISO 42001-konform handeln, haben einen Vorsprung bei der Erfüllung künftiger KI-Gesetze.

Zur effektiven Implementierung der in dieser Richtlinie beschriebenen Vorgaben ist eine strukturierte Vorgehensweise erforderlich. Eine umfassende Checkliste unterstützt bei der Entwicklung, Umsetzung und laufenden Überprüfung einer unternehmensweiten KI-Richtlinie. Sie enthält praxisnahe Schritte - von der Definition von Verantwortlichkeiten über die Risikobewertung bis hin zur Schulung von Mitarbeitenden und der kontinuierlichen Überwachung von KI-Systemen.
Die vollständige Checkliste zur Erstellung und Umsetzung einer KI-Richtlinie bietet eine systematische Anleitung, um alle relevanten Aspekte zu berücksichtigen und eine effiziente, regelkonforme und nachhaltige KI-Governance sicherzustellen.

Marktsteft, 03.03.2026

_____
(Geschäftsführer)