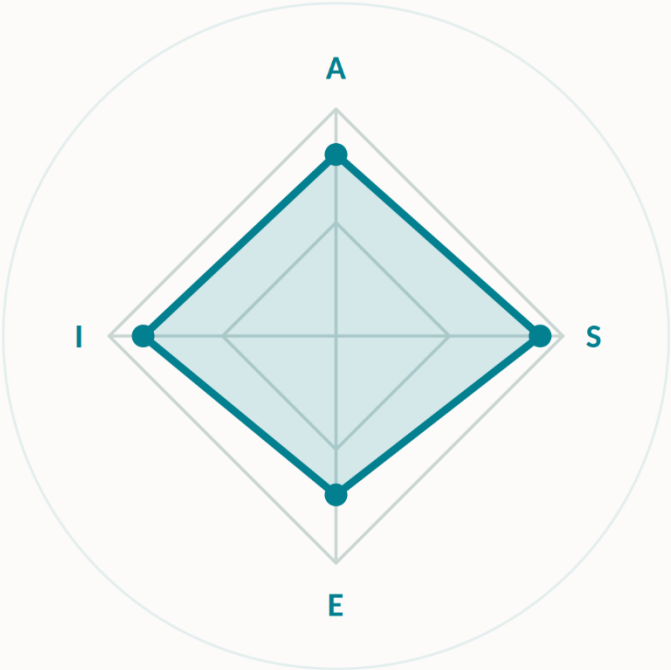
WISSENSCHAFTLICHES ARBEITSPAPIER

INDEPENDENT WORKING PAPER

ASEI-10

Ein heuristisches Klassifikationsmodell zur Bewertung des Risikopotenzials von KI-Systemen

AUTONOMIE · SEVERITÄT · EXPOSITION · IRREVERSIBILITÄT



RISIKOKLASSEN

RK1	RK2	RK3	RK4	RK5	RK6

sehr niedrig — sehr hoch

Dipl.-Kfm. Thomas H. Anhut

Berlin · 2026 · Version 1.0

Angaben zum Arbeitspapier

ASEI-10: Ein heuristisches Klassifikationsmodell zur Bewertung des Risikopotenzials von KI-Systemen

Zitiervorschlag

Anhut, T. H. (2026): ASEI-10 – Ein heuristisches Klassifikationsmodell zur Bewertung des Risikopotenzials von KI-Systemen. Unabhängiges Arbeitspapier, Version 1.0.
DOI: https://doi.org/10.5281/zenodo.20810443

Status des Arbeitspapiers

Bei diesem Dokument handelt es sich um ein unabhängiges wissenschaftliches Arbeitspapier (Working Paper). Es gibt einen vorläufigen Bearbeitungsstand wieder, wurde keinem externen Peer-Review unterzogen und wird zur fachlichen Diskussion und Weiterentwicklung veröffentlicht. Inhalte können in späteren Fassungen ergänzt, präzisiert oder revidiert werden.

Autor und Kontakt

Dipl.-Kfm. Thomas H. Anhut

ORCID-iD: https://orcid.org/0009-0001-5159-1248

Kontakt: thomasanhut@outlook.de

Hinweis zur Nutzung von KI-Werkzeugen

Bei der Erstellung dieses Arbeitspapiers wurden KI-gestützte Werkzeuge (ChatGPT von OpenAI sowie Claude von Anthropic) unterstützend eingesetzt, insbesondere zur Strukturierung, sprachlichen Überarbeitung, Recherche und kritischen Prüfung von Argumenten. Die inhaltliche Verantwortung für Konzeption, Aussagen, Bewertungen und Schlussfolgerungen liegt vollständig beim Autor.

Urheberrecht und Nutzung

Haftungsausschluss

Das ASEI-10-Modell ersetzt keine rechtliche, technische oder normative Prüfung und stellt keine Rechtskonformität fest. Die Anwendung erfolgt in eigener Verantwortung. Eine Haftung für Entscheidungen, die auf Grundlage dieses Arbeitspapiers getroffen werden, ist ausgeschlossen.

Vorwort

Das ASEI-10-Modell ist unabhängig entstanden – ohne Auftrag, ohne institutionelle Anbindung und ohne kommerzielles Interesse an einem bestimmten Bewertungsergebnis. Diese Unabhängigkeit ist kein Mangel, sondern eine Stärke des Ansatzes: Das Modell ist keinem Produkt, keinem Anbieter, keiner Behörde und keinem Verband verpflichtet. Es folgt allein der fachlichen Frage, wie Organisationen das Risikopotenzial konkreter KI-Systeme transparent, nachvollziehbar und anwendungsbezogen einordnen können.

Der Ausgangspunkt dieses Papiers ist die Überzeugung, dass die praktische Bewertung von KI-Systemen eine klarere, handhabbare und zugleich methodisch reflektierte Struktur benötigt. Bestehende rechtliche, normative und wissenschaftliche Rahmenwerke leisten wichtige Beiträge, beantworten aber nicht immer die operative Frage, wie ein konkretes KI-System im konkreten Nutzungskontext systematisch eingeordnet werden kann. Genau an dieser Stelle setzt ASEI-10 an.

Das Modell erhebt nicht den Anspruch, institutionell validierte Forschung zu ersetzen. Es versteht sich jedoch ausdrücklich als eigenständiger, begründeter und überprüfbarer Modellvorschlag. Alle Begriffe, Skalen, Formeln, Bewertungslogiken und Quellen werden offengelegt, hergeleitet und kritisch eingeordnet. Damit soll das Modell nicht durch institutionelle Autorität überzeugen, sondern durch seine innere Schlüssigkeit, seine praktische Anwendbarkeit und seine fachliche Anschlussfähigkeit.

ASEI-10 versteht sich als Diskussions- und Arbeitsmodell in einem sich schnell entwickelnden Feld. Es soll angewendet, geprüft, kritisiert, weiterentwickelt und bei Bedarf empirisch validiert werden. Fachliche Rückmeldungen, Hinweise auf Schwächen, Verbesserungsvorschläge und Mitwirkung an der in Kapitel 13 skizzierten Validierung sind ausdrücklich willkommen.

Rückmeldungen werden gerne unter thomasanhut@outlook.de entgegengenommen.

Abstract

Das ASEI-10-Modell ist ein heuristisches Klassifikationsmodell zur Bewertung des Risikopotenzials konkreter KI-Systeme in konkreten Anwendungskontexten. Ausgangspunkt ist die Annahme, dass KI-Risiken nicht allein aus der Leistungsfähigkeit eines Modells entstehen, sondern aus dessen Einbettung in reale Nutzungs-, Entscheidungs- und Wirkungskontexte. Bewertet wird daher nicht ein KI-Modell isoliert, sondern ein KI-System mit seinen Funktionen, Berechtigungen, Datenzugriffen, Nutzergruppen, Einsatzgrenzen und möglichen Folgen.

ASEI-10 stützt die Risikoeinstufung auf vier Bewertungsdimensionen: Autonomie, Severität, Exposition und Irreversibilität. Autonomie beschreibt die operative Selbstständigkeit eines Systems, Severität die mögliche Schwere negativer Folgen im Einsatzbereich, Exposition die Breite des möglichen Wirkungsraums und Irreversibilität die Rückholbarkeit möglicher negativer Folgen. Die Autonomie wird qualitativ auf einer neunstufigen Skala bewertet und für die Scorebildung auf einen normierten Autonomiewert A_n im Wertebereich von 1 bis 5 übertragen. Anschließend werden A_n, Severität, Exposition und Irreversibilität zu einem multiplikativen Rohwert zusammengeführt und logarithmisch auf eine Skala von 1 bis 10 normiert.

Der resultierende ASEI-10-Score ist als heuristischer Klassifikationsindex zu verstehen. Er dient der Zuordnung zu Risikoklassen und unterstützt die Priorisierung, Dokumentation und Kommunikation des strukturellen Risikopotenzials. Das Modell bewertet nicht die Eintrittswahrscheinlichkeit konkreter Schadensereignisse und ist daher kein vollständiges Risikomessinstrument im klassischen Sinne. Ergänzende Prüfschwellen und die Prüfung korrelativer Beziehungen zwischen den Dimensionen sollen verhindern, dass qualitativ bedeutsame Einzelmerkmale verdeckt oder dieselben Risikotreiber unreflektiert mehrfach gezählt werden.

Für die praktische Anwendung sieht ASEI-10 einen strukturierten Bewertungsprozess und einen standardisierten Bewertungsbogen vor. Dadurch werden Bewertungsgegenstand, Systemkontext, Wirkungskontext, Bewertungsmodus, Dimensionseinstufungen, Scorebildung, Schwellenprüfung, Mindestanforderungen und Neubewertungsanlässe nachvollziehbar dokumentiert. ASEI-10 ersetzt keine rechtlichen, technischen oder normativen Prüfungen und ist in der vorliegenden Fassung noch kein empirisch validiertes Messinstrument. Es kann bestehende Governance-, Risiko- und Prüfansätze jedoch durch eine operative, transparente und dokumentierbare Klassifikationslogik ergänzen.

Abstract

The ASEI-10 model is a heuristic classification model for assessing the risk potential of specific AI systems in specific application contexts. It is based on the assumption that AI risks do not arise solely from the performance of a model, but from its embedding in real-world contexts of use, decision-making and impact. Accordingly, the model does not assess an AI model in isolation, but an AI system with its functions, permissions, data access, user groups, operational boundaries and possible consequences.

ASEI-10 bases its classification on four assessment dimensions: autonomy, severity, exposure and irreversibility. Autonomy describes the operational independence of a system. Severity captures the seriousness of potential negative outcomes in the application domain. Exposure describes the breadth of the possible impact space, while irreversibility refers to the extent to which negative consequences can be corrected, retrieved or compensated. Autonomy is assessed qualitatively on a nine-level scale and then transformed into a normalized autonomy value A_n ranging from 1 to 5 for score calculation. A_n, severity, exposure and irreversibility are then combined into a multiplicative raw score, which is logarithmically normalized to a scale from 1 to 10.

The resulting ASEI-10 score should be understood as a heuristic classification index. It supports assignment to risk classes and helps prioritize, document and communicate structural risk potential. The model does not assess the probability of specific adverse events and should therefore not be understood as a complete risk measurement instrument in the classical sense. Additional threshold criteria and the assessment of correlative relationships between the dimensions are intended to prevent qualitatively significant individual risk characteristics from being obscured or the same risk drivers from being counted repeatedly without reflection.

For practical application, ASEI-10 provides a structured assessment process and a standardized assessment form. This makes it possible to document the object of assessment, system context, impact context, assessment mode, dimensional ratings, score calculation, threshold assessment, minimum requirements and triggers for reassessment in a transparent and auditable manner. ASEI-10 does not replace legal, technical or normative assessments and, in its present form, is not yet an empirically validated measurement instrument. It can, however, complement existing governance, risk management and assessment approaches through an operational, transparent and documented classification logic.

Management Summary

Das ASEI-10-Modell ist ein heuristisches Klassifikationsmodell zur Bewertung des Risikopotenzials konkreter KI-Systeme in konkreten Anwendungskontexten. Es wurde entwickelt, um Organisationen dabei zu unterstützen, KI-Anwendungen nicht pauschal, sondern strukturiert, nachvollziehbar und vergleichbar einzuordnen. Im Mittelpunkt steht nicht die abstrakte Leistungsfähigkeit eines KI-Modells, sondern das konkrete KI-System mit seinen Funktionen, Berechtigungen, Datenzugriffen, Nutzergruppen, Einsatzgrenzen und möglichen Folgen.

Ausgangspunkt des Modells ist die Beobachtung, dass KI-Systeme in Organisationen häufig entweder überschätzt oder verharmlost werden. Einerseits werden sie als bloße Assistenzwerkzeuge betrachtet, obwohl sie bereits Entscheidungen vorbereiten, Prozesse beeinflussen oder operative Handlungen auslösen können. Andererseits werden sie teilweise abstrakt als grundsätzlich gefährlich beschrieben, ohne den konkreten Nutzungskontext, die tatsächlichen Systemrechte und die Rückholbarkeit möglicher Folgen zu berücksichtigen. ASEI-10 soll zwischen diesen beiden Verkürzungen vermitteln und eine sachlichere, dokumentierbare Bewertung ermöglichen.

ASEI-10 bewertet nicht das klassische Risiko im Sinne von Eintrittswahrscheinlichkeit mal Schadensausmaß. Das Modell bestimmt keine empirische Wahrscheinlichkeit dafür, dass ein bestimmtes Schadensereignis tatsächlich eintritt. Stattdessen klassifiziert es das strukturelle Risikopotenzial eines KI-Systems. Gefragt wird also: Welche Handlungsmöglichkeiten besitzt das System? In welchem Einsatzbereich wird es verwendet? Wie weit können sich seine Wirkungen ausbreiten? Und wie schwer wären negative Folgen nachträglich zu erkennen, zu stoppen, zu korrigieren oder zu kompensieren?

Die Grundannahme des Modells lautet, dass das Risikopotenzial eines KI-Systems aus dem Zusammenspiel von vier Dimensionen entsteht: Autonomie, Severität, Exposition und Irreversibilität. Ein KI-System ist nicht allein deshalb risikorelevant, weil es technisch leistungsfähig ist. Ebenso wenig ist ein System automatisch unproblematisch, nur weil es keine vollständig autonomen Entscheidungen trifft. Entscheidend ist, was das System im konkreten Anwendungskontext tun kann, welche Schutzgüter betroffen sind, wie weit seine Wirkungen reichen und wie gut mögliche Fehlwirkungen rückholbar bleiben.

Autonomie beschreibt die operative Selbstständigkeit eines KI-Systems. Sie reicht von passiven oder reaktiven Assistenzsystemen bis zu persistenten, verteilten, selbstoptimierenden oder nicht zuverlässig kontrollierbaren agentischen Systemen. Für die fachliche Beschreibung wird eine neunstufige Autonomieskala verwendet. Für die Scorebildung wird diese qualitative Einstufung auf einen normierten Autonomiewert A_n im Wertebereich von 1 bis 5 übertragen, damit Autonomie rechnerisch nicht allein aufgrund ihrer feineren Skala stärker gewichtet wird als die übrigen Dimensionen.

Severität beschreibt die mögliche Schwere negativer Folgen im konkreten Einsatzbereich. Private, kreative oder vorbereitende Nutzungen sind anders zu bewerten als Anwendungen in Bildung, Beschäftigung, Gesundheit, Recht, öffentlicher Sicherheit, kritischer Infrastruktur oder Cybersecurity. Bewertet wird nicht, ob ein Schaden wahrscheinlich eintritt, sondern welche Bedeutung die berührten Schutzgüter und möglichen Konsequenzen im jeweiligen Anwendungskontext besitzen.

Exposition beschreibt die Breite des möglichen Wirkungsraums. Ein Fehler kann auf eine einzelne Person begrenzt bleiben, eine Gruppe betreffen, organisationale Prozesse beeinflussen, öffentlich sichtbar werden oder systemische Wirkungen über mehrere Organisationen, Plattformen oder Infrastrukturen hinweg entfalten. Die Exposition macht sichtbar, ob ein System nur lokal wirkt oder ob seine Ergebnisse, Entscheidungen oder Handlungen eine größere Reichweite erreichen können.

Irreversibilität beschreibt, wie schwer negative Folgen nachträglich erkannt, gestoppt, korrigiert, zurückgeholt oder kompensiert werden können. Leicht korrigierbare Entwurfsfehler sind anders zu bewerten als Datenschutzverletzungen, Benachteiligungen, Reputationsschäden, körperliche Schäden, sicherheitskritische Fehlreaktionen oder Folgen, die sich über nachgelagerte Prozesse weiterverbreiten.

Aus den vier Dimensionen wird zunächst ein multiplikativer Rohwert gebildet. Dieser Rohwert wird anschließend logarithmisch auf eine Skala von 1 bis 10 normiert. Der daraus entstehende ASEI-10-Score ist ein heuristischer Klassifikationsindex. Er dient der Orientierung, Priorisierung und Dokumentation, ist aber keine exakte Messgröße im naturwissenschaftlichen, versicherungsmathematischen oder statistischen Sinne. Einzelne Dezimalstellen dürfen deshalb nicht überbewertet werden. Maßgeblich bleibt die Verbindung aus Score, Risikoklasse, Einzeldimensionen, Prüfschwellen, korrelativen Beziehungen und fachlicher Begründung.

Die Risikoklassen des Modells ordnen den ASEI-10-Score qualitativen Risikopotenzialen zu. Sie reichen von sehr niedrigem bis zu sehr hohem Risikopotenzial. Diese Risikoklassen sollen Organisationen helfen, KI-Systeme zu priorisieren, Mindestanforderungen abzuleiten und interne Steuerungsentscheidungen vorzubereiten. Sie ersetzen jedoch nicht die fachliche Prüfung der Einzeldimensionen. Gerade bei grenznahen Einstufungen ist nicht der gerundete Score allein ausschlaggebend, sondern der ungerundete Wert in Verbindung mit der Begründung der Dimensionseinstufungen und möglichen Prüfschwellen.

Eine besondere Funktion übernehmen die Prüfschwellen des Modells. Sie verhindern, dass qualitativ bedeutsame Einzelmerkmale durch den Gesamtscore verdeckt werden. Ein KI-System kann zusätzliche Mindestanforderungen auslösen, obwohl der Gesamtscore noch nicht in der höchsten Risikoklasse liegt. Solche Mindestanforderungen können etwa zusätzliche Freigaben, fachliche Prüfungen, Protokollierung, Monitoring, Rollen- und Rechtebegrenzungen, Rückholmechanismen, Datenschutz- oder Sicherheitsprüfungen sowie regelmäßige Neubewertungen betreffen.

Zusätzlich berücksichtigt ASEI-10, dass die vier Bewertungsdimensionen analytisch getrennt sind, in realen Anwendungskontexten aber korrelativ zusammenwirken können. Hohe Severität kann mit hoher Exposition verbunden sein, breite Exposition kann die Rückholbarkeit erschweren, und hohe Autonomie kann Korrekturfenster verkürzen. Diese Zusammenhänge sind methodisch bedeutsam, dürfen aber nicht zu einer unreflektierten Mehrfachzählung führen. Jede Dimension muss eigenständig begründet werden. Die Prüfung korrelativer Beziehungen dient daher der Plausibilitätskontrolle und der sauberen Abgrenzung der Risikotreiber.

Für die praktische Anwendung sieht ASEI-10 einen strukturierten Bewertungsprozess vor. Zunächst werden Bewertungsgegenstand und Systemkontext festgelegt. Anschließend werden die vier Dimensionen einzeln eingestuft, die Scorelogik angewendet, Prüfschwellen geprüft, Mindestanforderungen abgeleitet und mögliche Anlässe für eine spätere Neubewertung dokumentiert. Der zugehörige Bewertungsbogen macht die Annahmen, Einstufungen, Begründungen und Anwendungskonsequenzen nachvollziehbar. Er ist damit nicht nur ein Rechenformular, sondern ein Dokumentations- und Governance-Instrument.

Das Modell unterscheidet zwischen Brutto-Bewertung, Netto-Bewertung und der Bewertung eines systemimmanent begrenzten Systemdesigns. Schutz-, Kontroll- und Begrenzungsmaßnahmen dürfen nur dann scorewirksam berücksichtigt werden, wenn sie eine oder mehrere Bewertungsdimensionen tatsächlich, nachweisbar und dauerhaft reduzieren. Bloße Absichtserklärungen, unverbindliche Nutzungshinweise oder nicht überprüfbare Kontrollen rechtfertigen keine niedrigere Einstufung. Maßnahmen, die das Risikopotenzial nicht unmittelbar senken, können dennoch als Mindestanforderungen oder Governance-Maßnahmen relevant sein.

Die Fallbeispiele zeigen die Spannweite des Modells. Ein privater Chatbot für kreative Texte weist ein sehr niedriges Risikopotenzial auf. Ein KI-Assistent für Unterrichts- und Prüfungsvorbereitung kann bereits ein erhöhtes Risikopotenzial erreichen, weil fachliche Fehler Lern- und Prüfungskontexte beeinflussen. Ein Kundenservice-System kann trotz begrenzter Autonomie durch breite Kundenexposition und personenbezogene Daten grenznah zu hohem Risikopotenzial liegen. Unternehmensagenten, Personalvorauswahlsysteme und Systeme in kritischer Infrastruktur zeigen, dass hohe Risiken auf unterschiedlichen Wegen entstehen können: durch operative Handlungsausführung, Entscheidungsnähe, sensible Schutzgüter, systemische Reichweite oder schwer rückholbare Folgen.

ASEI-10 ist anschlussfähig an bestehende rechtliche, normative und organisatorische Rahmenwerke. Dazu gehören insbesondere der EU AI Act, das NIST AI Risk Management Framework, ISO/IEC 23894, ISO/IEC 42001, internationale AI-Safety-Berichte sowie praxisorientierte Prüf- und Qualitätsbewertungsansätze wie der KI-Prüfkatalog des Fraunhofer IAIS, MISSION KI und VDE-nahe Qualitätsstandards. ASEI-10 ersetzt diese Ansätze nicht. Es ist kein Rechtsgutachten, kein Compliance-Nachweis, keine technische Sicherheitsprüfung und kein vollständiges Managementsystem. Sein Beitrag liegt darin, bestehende Governance- und Risikomanagementansätze durch eine operative, skalierbare und dokumentierbare Klassifikationslogik zu ergänzen.

Die Grenzen des Modells sind ausdrücklich zu beachten. ASEI-10 misst keine Eintrittswahrscheinlichkeiten, ersetzt keine empirische Systemevaluation, bewertet keine Modellqualität im technischen Sinne und kann keine rechtliche Zulässigkeit garantieren. Es ist in der vorliegenden Fassung ein heuristisches, operationalisiertes und weiter validierungsbedürftiges Klassifikationsmodell. Seine wissenschaftliche Weiterentwicklung sollte insbesondere Inter-Rater-Reliabilität, Validität, Sensitivität der Scorelogik, Trennschärfe der Dimensionen und praktische Anwendbarkeit in realen KI-Inventaren untersuchen.

Der zentrale Nutzen von ASEI-10 liegt in der Verbindung von qualitativer Analyse, nachvollziehbarer Indexbildung, Risikoklassen, Prüfschwellen und dokumentierter Anwendung. Das Modell schafft eine Brücke zwischen abstrakter KI-Risikodiskussion und praktischer Organisationssteuerung. Es hilft, KI-Systeme nach ihrem tatsächlichen Risikopotenzial zu unterscheiden, Prioritäten zu setzen, Mindestanforderungen abzuleiten und Verantwortungsstrukturen zu stärken. Damit kann ASEI-10 einen Beitrag zu einer sachlicheren, transparenteren und verantwortbareren KI-Governance leisten.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Variablenverzeichnis

Das ASEI-10-Modell verwendet wenige zentrale Variablen und Rechenschritte. Die folgenden Angaben dienen der Orientierung und ersetzen nicht die Herleitung und methodische Begründung in den Kapiteln 2 und 9.

A	Qualitative Autonomiestufe des KI-Systems auf der Skala A1 bis A9.
A_n	Normierter Autonomiewert im Wertebereich von 1 bis 5. Für die Scorebildung wird die qualitative Autonomiestufe A in den normierten Autonomiewert A_n überführt.
E	Exposition. Bewertungsdimension für die Breite des möglichen Wirkungsraums. Wertebereich: E1 bis E5.
I	Irreversibilität. Bewertungsdimension für die Rückholbarkeit, Korrigierbarkeit oder Kompensierbarkeit möglicher negativer Folgen. Wertebereich: I1 bis I5.
R	Multiplikativer Rohwert des ASEI-10-Modells. Er ergibt sich aus dem Produkt des normierten Autonomiewerts und der drei übrigen Bewertungsdimensionen.
RK	Risikoklasse. Qualitative Einordnung des ASEI-10-Scores in eine der sechs Risikoklassen RK1 bis RK6.
S	Severität. Bewertungsdimension für die mögliche Schwere negativer Folgen im konkreten Anwendungskontext. Wertebereich: S1 bis S5.

Formelverzeichnis

Normierung der Autonomie: $$\boldsymbol{A_n = 1 + 4 \cdot \frac{A-1}{8}}$$

Äquivalent vereinfacht: $$\boldsymbol{A_n = \frac{A+1}{2}}$$

Rohwertbildung: $$\boldsymbol{R = A_n \cdot S \cdot E \cdot I}$$

Logarithmische Normierung auf die ASEI-10-Skala: $$\boldsymbol{\text{ASEI-10} = 1 + 9 \cdot \frac{\ln(R)}{\ln(625)}}$$

Hinweis zur Interpretation: Der ASEI-10-Score ist ein heuristischer Klassifikationsindex. Er dient der Zuordnung zu Risikoklassen und der strukturierten Priorisierung, ist aber keine exakte quantitative Risikomessung. Maßgeblich bleiben neben dem Score die Einzeldimensionen, Prüfschwellen, korrelativen Beziehungen und fachlichen Begründungen.

Abkürzungsverzeichnis

AI	Artificial Intelligence; englische Bezeichnung für Künstliche Intelligenz.
AI Act	Kurzbezeichnung für die Verordnung (EU) 2024/1689 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz.
AIQI	AI Quality Infrastructure; Konsortium bzw. Initiative im Umfeld von KI-Qualität, KI-Prüfung und KI-Vertrauenswürdigkeit.
AIMS	Artificial Intelligence Management System; Managementsystem für künstliche Intelligenz, insbesondere im Zusammenhang mit ISO/IEC 42001.
AIR	AI Risk Categorization; im Kontext des zitierten AIR-2024-Ansatzes verwendete Abkürzung für einen KI-Risikokategorisierungsansatz.
arXiv	Offenes Online-Repositorium für wissenschaftliche Vorabveröffentlichungen, insbesondere in Informatik, Mathematik, Physik und angrenzenden Disziplinen.
ASEI	Abkürzung der vier Bewertungsdimensionen Autonomie, Severität, Exposition und Irreversibilität.
DIN	Deutsches Institut für Normung.
DLR	Deutsches Zentrum für Luft- und Raumfahrt.
DOI	Digital Object Identifier; dauerhafter digitaler Identifikator für wissenschaftliche Veröffentlichungen.
DSIT	Department for Science, Innovation and Technology; Ministerium des Vereinigten Königreichs, unter anderem Herausgeber bzw. institutioneller Rahmen der International AI Safety Reports.
ELI	European Legislation Identifier; standardisierter europäischer Identifikator für Rechtsakte.
EU	Europäische Union.
FMEA	Failure Mode and Effects Analysis; Fehlermöglichkeits- und Einflussanalyse.
FMECA	Failure Mode, Effects and Criticality Analysis; erweiterte Fehlermöglichkeits- und Einflussanalyse mit zusätzlicher Kritikalitätsbetrachtung.
GPAI	General-Purpose AI; KI-Modell oder KI-System mit breitem Einsatzspektrum für unterschiedliche Aufgaben und Anwendungskontexte.
IAIS	Institut für Intelligente Analyse- und Informationssysteme; Fraunhofer IAIS.
IEC	International Electrotechnical Commission; internationale Normungsorganisation für Elektrotechnik, Elektronik und verwandte Technologien.
ISO	International Organization for Standardization; internationale Normungsorganisation.
ISO/IEC	Gemeinsame Normen von ISO und IEC, insbesondere im Bereich Informationstechnologie und künstliche Intelligenz.
KI	Künstliche Intelligenz.
MIT	Massachusetts Institute of Technology.
ML	Machine Learning; maschinelles Lernen.
NIST	National Institute of Standards and Technology; US-amerikanische Standardisierungs- und Forschungsbehörde.
OJ	Official Journal; Amtsblatt der Europäischen Union. In Quellenangaben häufig als „OJ L“ für die Reihe „Legislation“ verwendet.
PDF	Portable Document Format; Dateiformat zur plattformübergreifenden Darstellung von Dokumenten.
RMF	Risk Management Framework; Risikomanagement-Rahmenwerk, insbesondere im Zusammenhang mit dem NIST AI Risk Management Framework.
SPEC	Spezifikation; im Normungs- und Standardisierungskontext Bezeichnung für eine dokumentierte technische oder organisatorische Spezifikation.
TÜV	Technischer Überwachungsverein; in diesem Papier insbesondere im Zusammenhang mit dem TÜV AI.Lab verwendet.
VDE	Verband der Elektrotechnik Elektronik Informationstechnik e. V.

1 Einleitung

Die Diskussion über Künstliche Intelligenz wird häufig von pauschalen Gegensätzen geprägt. Einerseits wird KI als produktivitätssteigerndes Werkzeug für Wirtschaft, Forschung, Bildung, Verwaltung und kreative Arbeit beschrieben. Andererseits stehen Warnungen vor Manipulation, Missbrauch, Fehlfunktionen, Autonomieverlust, Kontrollproblemen und systemischen Risiken im Vordergrund. Beide Sichtweisen bleiben unzureichend, solange KI-Systeme nicht nach ihren konkreten Einsatzbedingungen, Zugriffsmöglichkeiten und Handlungsfolgen unterschieden werden.

Entscheidend ist nicht allein, wie leistungsfähig ein KI-Modell sprachlich, analytisch oder technisch erscheint. Risikorelevant wird ein KI-System vor allem durch seine konkrete Einbettung: Reagiert es nur auf Eingaben? Nutzt es Werkzeuge? Darf es Daten verändern, Nachrichten versenden, Entscheidungen vorbereiten oder selbstständig ausführen? Wirkt es nur in einem privaten Einzelfall oder in organisationsweiten, öffentlichen oder systemkritischen Zusammenhängen? Und lassen sich fehlerhafte oder schädliche Folgen nachträglich erkennen, stoppen, korrigieren oder rückgängig machen?

Für die Bewertung von KI-Systemen bestehen inzwischen zahlreiche rechtliche, normative und methodische Anschlussstellen. Der EU AI Act[1] arbeitet risikobasiert und berücksichtigt unter anderem Autonomie, Hochrisiko-Anwendungen, Risikomanagement, menschliche Aufsicht, Robustheit, Grundrechtsfolgenabschätzung und besondere Pflichten für General-Purpose-AI-Modelle mit systemischem Risiko. Die International AI Safety Reports 2024 bis 2026[2] betonen ergänzend die besondere Bedeutung von General-Purpose AI, agentischen Systemen, Evaluationslücken, Missbrauchsrisiken, Fehlfunktionen, systemischen Risiken und möglichen Kontrollverlustszenarien. Weitere Anschlussstellen bestehen zum NIST AI Risk Management Framework[3], zum NIST Generative AI Profile[4] sowie zu internationalen Normen des KI-Risikomanagements und der KI-Governance.[5]

Trotz dieser vielfältigen Ansätze bleibt eine praktische und methodische Lücke bestehen. Bestehende rechtliche, normative und organisatorische Rahmenwerke beschreiben wichtige Anforderungen an Regulierung, Governance, Risikomanagement, Qualitätssicherung und technische Prüfung von KI-Systemen. Sie beantworten jedoch nur begrenzt die operative Frage, wie das Risikopotenzial eines konkreten KI-Systems in einem konkreten Anwendungskontext kompakt, nachvollziehbar, vergleichbar und dokumentierbar klassifiziert werden kann.

Daraus ergibt sich die leitende Forschungs- und Entwicklungsfrage dieses Papiers:

> *Wie kann das Risikopotenzial konkreter KI-Systeme so erfasst werden, dass Autonomie, Severität, Exposition und Irreversibilität systematisch bewertet und für organisatorische Governance-Entscheidungen nutzbar gemacht werden?*

[1] Verordnung (EU) 2024/1689 (EU AI Act), Art. 3, 6, 9, 14, 15, 27, 51, 53, 55

[2] Bengio et al. 2024; 2025; 2026

[3] NIST AI RMF 1.0

[4] NIST 2024, AI 600-1

[5] ISO/IEC 23894:2023; ISO/IEC 42001:2023

Das ASEI-10-Modell wurde als heuristischer Modellvorschlag zur Bearbeitung dieser Frage entwickelt. Es liefert keine abschließende theoretische Antwort, sondern stellt eine strukturierte, nachvollziehbare und praktisch anwendbare Klassifikationslogik bereit, mit der das Risikopotenzial von KI-Systemen im jeweiligen Anwendungskontext systematisch eingeordnet werden kann. Bewertet wird daher nicht ein KI-Modell in abstrakter oder isolierter Form, sondern das konkrete KI-System mit seinen Funktionen, Berechtigungen, Datenzugriffen, Nutzergruppen, Einsatzgrenzen und möglichen Folgen.

Die Risikoeinstufung stützt sich auf vier Bewertungsdimensionen:

1. Autonomie – wie selbstständig ein KI-System Ziele verfolgt, Handlungsschritte plant, Werkzeuge auswählt und über Zeit aktiv bleibt;
2. Severität – wie sensibel oder folgenreich der Einsatzbereich ist;
3. Exposition – wie breit Personen, Daten, Organisationen, Systeme oder öffentliche Räume den Wirkungen des Systems ausgesetzt sind;
4. Irreversibilität – wie schwer negative Folgen nachträglich rückholbar, korrigierbar oder kompensierbar sind.

ASEI-10 versteht sich nicht als Ersatz für bestehende rechtliche, normative oder technische Prüfungen. Es soll vielmehr eine zusätzliche, operationalisierbare Klassifikationslogik bereitstellen, mit der das Risikopotenzial von KI-Systemen systematisch beschrieben, verglichen, dokumentiert und geprüft werden kann.

Die zentrale Modellannahme lautet:

> *Für die Zwecke des ASEI-10-Modells wird das Risikopotenzial eines KI-Systems nicht aus der Modellleistung allein abgeleitet. Es wird als Zusammenspiel von Autonomie, Severität, Exposition und Irreversibilität verstanden. Entscheidend ist, wie selbstständig ein System im konkreten Anwendungskontext handeln kann, welche Schwere mögliche negative Folgen besitzen, wie weit seine Wirkungen reichen können und wie schwer mögliche Fehlwirkungen rückholbar, korrigierbar oder kompensierbar wären.*

Das Modell unterscheidet daher zwischen rein assistiven, agentischen und strategisch-systemischen KI-Systemen. Es verbindet eine qualitative neunstufige Autonomieskala mit drei ergänzenden fünfstufigen Bewertungsdimensionen für Severität, Exposition und Irreversibilität. Für die Scorebildung wird die Autonomiestufe auf einen normierten Autonomiewert A_n im Wertebereich von 1 bis 5 übertragen. Aus den vier Faktoren wird ein multiplikativer Rohwert gebildet, der anschließend logarithmisch auf einen ASEI-10-Score von 1 bis 10 normiert wird. Zusätzlich werden qualitative Prüfschwellen, Begründungspflichten und Plausibilitätsprüfungen vorgesehen, insbesondere die Prüfung korrelativer Beziehungen zwischen den Dimensionen. Dadurch soll verhindert werden, dass bedeutsame Einzelmerkmale durch den zusammenfassenden Score verdeckt oder dieselben Risikotreiber unreflektiert mehrfach gezählt werden.

Ziel des ASEI-10-Modells ist eine sachlichere und prüfbarere Risikodiskussion. Es soll nicht alarmistisch behaupten, dass KI-Systeme grundsätzlich gefährlich seien. Ebenso wenig soll es Risiken verharmlosen, wenn KI-Systeme zunehmend autonom handeln, in kritischen Kontexten eingesetzt werden, große Exposition erzeugen oder schwer reversible Folgen auslösen können. ASEI-10 soll deshalb als wissenschaftlich anschlussfähiges Arbeitsmodell dienen, das vorhandene Regulierungs-, Governance- und Risikomanagementansätze ergänzt und die Bewertung konkreter KI-Systeme transparenter macht.

2 Methodische Einordnung des ASEI-10-Modells

2.1 Methodische Grundposition des Modells

2.1.1 Funktion der methodischen Einordnung

Das ASEI-10-Modell ist ein heuristisches Klassifikationsmodell zur Bewertung des Risikopotenzials konkreter KI-Systeme in konkreten Anwendungskontexten. Es soll Organisationen dabei unterstützen, KI-Systeme strukturiert zu beschreiben, risikorelevante Merkmale transparent zu erfassen, Systeme vergleichbar zu priorisieren und Mindestanforderungen aus Prüfschwellen abzuleiten. Damit besitzt das Modell eine operative und heuristische Funktion: Es ordnet Risikopotenziale, ersetzt aber keine vollständige quantitative Risikoanalyse.

Diese methodische Einordnung ist notwendig, weil der Begriff Risiko in unterschiedlichen Fachkontexten unterschiedlich verwendet wird. In klassischen Risikomanagementansätzen wird Risiko häufig über die Verbindung von Eintrittswahrscheinlichkeit und Auswirkung bzw. Schadensausmaß betrachtet. Das ASEI-10-Modell folgt dieser Logik nicht vollständig. Es bewertet nicht die empirische Wahrscheinlichkeit, mit der ein bestimmtes Fehlverhalten tatsächlich eintritt, sondern das potenzielle Risikoprofil, das sich aus Systemeinbettung, Handlungsmöglichkeiten, Einsatzbereich, Wirkungsreichweite und Rückholbarkeit möglicher Folgen ergibt.

ASEI-10 ist daher nicht als validiertes Messinstrument im engeren messtheoretischen Sinne zu verstehen. Es ist ein strukturierender Klassifikationsindex. Seine Stärke liegt nicht in der exakten Messung objektiver Risikowerte, sondern in der systematischen Offenlegung von Bewertungsannahmen. Das Modell macht sichtbar, warum ein KI-System als niedrig, erhöht, hoch oder sehr hoch risikobehaftet eingestuft wird und welche Dimensionen diese Einstufung tragen.

2.1.2 Abgrenzung zum klassischen Risikobegriff

Im klassischen Risikomanagement wird Risiko regelmäßig als Kombination aus Unsicherheit, Ereignis, Auswirkung und Wahrscheinlichkeit verstanden.[6] Für viele technische, finanzielle oder sicherheitsbezogene Anwendungen ist diese Perspektive sachgerecht, weil Risiken dort zumindest teilweise über Eintrittshäufigkeiten, Ausfallwahrscheinlichkeiten, Schadensstatistiken, Testdaten oder Betriebserfahrungen quantifiziert werden können.

Bei KI-Systemen ist eine solche Wahrscheinlichkeitsbewertung häufig schwierig. Die Eintrittswahrscheinlichkeit fehlerhafter, verzerrter, manipulativer oder unzureichend kontrollierter Systemwirkungen hängt von vielen Faktoren ab: Modellqualität, Trainingsdaten, Systemarchitektur, Prompting, Toolzugriffen, Nutzungsverhalten, Schutzmechanismen, Monitoring, Angriffsmethoden, organisatorischer Kontrolle und konkreter Betriebspraxis. Viele dieser Faktoren sind dynamisch, schwer beobachtbar oder erst im laufenden Einsatz belastbar bestimmbar.

ASEI-10 grenzt sich deshalb vom klassischen Risikobegriff ab. Das Modell fragt nicht primär: Wie wahrscheinlich ist ein konkretes Schadensereignis? Es fragt vielmehr: Welches Risikopotenzial besitzt ein konkretes KI-System, wenn Fehler, Fehlverwendung, Missbrauch oder Kontrollprobleme auftreten? Der Begriff Risikopotenzial ist dabei bewusst gewählt. Er bezeichnet die potenzielle Gefährdungs- und Schadensrelevanz eines Systems, nicht die statistisch bestimmte Eintrittswahrscheinlichkeit eines bestimmten Ereignisses.

[6] ISO 31000:2018; IEC 31010:2019

Damit bewertet ASEI-10 keine vollständige Risikogröße im Sinn von Wahrscheinlichkeit mal Schadensausmaß. Es bewertet die strukturelle Risikointensität eines KI-Systems im Anwendungskontext. Die Eintrittswahrscheinlichkeit muss bei Bedarf zusätzlich durch technische Evaluation, Fehlerratenanalyse, Robustheitstests, Red Teaming, Monitoring, Betriebserfahrung oder andere empirische Verfahren bestimmt werden.

2.1.3 Risikopotenzial statt Eintrittswahrscheinlichkeit

Die bewusste Ausklammerung der Eintrittswahrscheinlichkeit ist keine unbeabsichtigte Lücke, sondern eine Modellentscheidung. ASEI-10 soll auch dann anwendbar sein, wenn noch keine belastbaren empirischen Daten über Fehlerraten, Angriffswahrscheinlichkeiten, Halluzinationshäufigkeiten oder reale Schadensereignisse vorliegen. Dies ist insbesondere bei neuen KI-Systemen, Pilotanwendungen, agentischen Konfigurationen oder dynamisch aktualisierten Modellen häufig der Fall.

Das Modell bewertet deshalb die potenzielle Tragweite eines Schadensraums. Ein System mit hoher Autonomie, hoher Severität, hoher Exposition und hoher Irreversibilität bleibt auch dann besonders prüfbedürftig, wenn die Eintrittswahrscheinlichkeit einzelner Schadensereignisse noch nicht zuverlässig quantifiziert werden kann. Umgekehrt bedeutet ein niedriger ASEI-10-Score nicht, dass ein System zuverlässig, fehlerfrei oder technisch sicher ist. Er bedeutet lediglich, dass das strukturelle Risikopotenzial im beschriebenen Anwendungskontext begrenzt ist.

Die Eintrittswahrscheinlichkeit kann als ergänzende Analyseebene verstanden werden. In einer erweiterten Risikobetrachtung könnte ASEI-10 mit empirischen Größen wie Fehlerraten, Robustheitskennzahlen, Evaluationsresultaten, Sicherheitsvorfällen, Nutzerfehlverhalten oder Missbrauchswahrscheinlichkeiten verbunden werden. Der Kernindex bleibt jedoch auf das kontextbezogene Risikopotenzial beschränkt.

Diese Trennung hat einen praktischen Vorteil. Organisationen können zunächst unabhängig von noch unsicheren Wahrscheinlichkeitsdaten klären, welche KI-Systeme aufgrund ihrer Struktur, Rechte, Reichweite und möglichen Folgen besonders aufmerksam zu behandeln sind. Erst danach kann entschieden werden, für welche Systeme eine vertiefte technische oder empirische Wahrscheinlichkeitsanalyse erforderlich ist.

2.1.4 ASEI-10 als heuristischer Klassifikationsindex

ASEI-10 ist als heuristischer Klassifikationsindex zu verstehen. Ein solcher Index dient der systematischen Ordnung, Priorisierung und Kommunikation komplexer Sachverhalte. Er ersetzt keine vollständige quantitative Modellierung, sondern schafft eine nachvollziehbare Struktur für fachliche Entscheidungen.

Der Indexcharakter des Modells hat drei Konsequenzen. Erstens sind die vergebenen Werte nicht als exakte Messwerte zu interpretieren. Ein ASEI-10-Score von 7,4 bedeutet nicht, dass ein System objektiv exakt 0,4 Risikoeinheiten riskanter ist als ein System mit 7,0. Zweitens dürfen knappe Scoreunterschiede nicht überbewertet werden. Entscheidender als einzelne Dezimalstellen sind Risikobereich, Einzeldimensionen, Prüfschwellen und Begründung der Einstufung. Drittens muss jede Bewertung dokumentiert werden, damit die getroffenen Annahmen überprüfbar bleiben.

Das Modell ist damit näher an einer strukturierten Bewertungs- und Priorisierungsmethode als an einem empirisch kalibrierten Messinstrument. Sein wissenschaftlicher Status ist der eines begründeten Artefakts, das weiter validiert, kalibriert und empirisch geprüft werden kann. Fallbeispiele illustrieren die Anwendung des Modells, stellen aber keine Validierung dar.

2.2 Skalen-, Normierungs- und Indexlogik

2.2.1 Zur Verwendung ordinaler Skalen

Die vier ASEI-Dimensionen beruhen auf geordneten Stufen. Diese Stufen sind ordinal: Eine höhere Stufe bedeutet eine stärkere Ausprägung der jeweiligen Risikodimension. Aus A5 folgt also eine höhere Autonomie als aus A3; aus E4 folgt eine höhere Exposition als aus E2. Daraus folgt jedoch nicht automatisch, dass die Abstände zwischen den Stufen gleich groß sind oder dass eine Stufe doppelt so stark ist wie eine andere.

Diese messtheoretische Einschränkung ist für die Interpretation des Modells wesentlich. Die Zahlenwerte der Stufen sind keine physikalischen Messgrößen, sondern Ordnungswerte. Ihre rechnerische Verknüpfung ist daher als heuristische Indexbildung zu verstehen, nicht als mathematisch exakte Risikomessung. ASEI-10 verwendet Zahlen, um Bewertungen vergleichbar und dokumentierbar zu machen, nicht um eine Scheingenauigkeit objektiver Risikometrik zu erzeugen.

Die Verwendung ordinaler Skalen ist in Risikobewertungen verbreitet, aber methodisch nicht unproblematisch. Besonders problematisch wird sie, wenn ordinalen Stufen ohne Begründung arithmetische Eigenschaften zugeschrieben werden.[7]

ASEI-10 begegnet diesem Problem durch drei Maßnahmen:

Erstens werden die Stufen ausführlich definiert und voneinander abgegrenzt.

Zweitens wird der Score stets gemeinsam mit den Einzeldimensionen und Prüfschwellen interpretiert.

Drittens wird der Score ausdrücklich als Klassifikationsindex und nicht als exakte Messgröße verstanden.

2.2.2 Normierung und Gewichtung der Bewertungsdimensionen

Eine besondere methodische Frage betrifft die Gewichtung der vier Bewertungsdimensionen. Die Autonomiedimension wird qualitativ auf neun Stufen beschrieben, weil agentische Fähigkeiten differenziert erfasst werden sollen. Die Dimensionen Severität, Exposition und Irreversibilität werden dagegen jeweils auf fünf Stufen beschrieben. Würden diese Werte unverändert in die Scorebildung eingehen, hätte Autonomie allein aufgrund der längeren Skala rechnerisch ein höheres Gewicht.

Um diese unbegründete Übergewichtung zu vermeiden, unterscheidet ASEI-10 zwischen der qualitativen Autonomiestufe A und dem für die Scorebildung verwendeten normierten Autonomiewert A_n. Die qualitative Skala A1 bis A9 bleibt erhalten, weil sie die unterschiedlichen Formen assistiver, agentischer und strategisch-systemischer Autonomie differenziert abbildet. Für die rechnerische Scorebildung wird sie jedoch auf denselben Wertebereich wie die übrigen Dimensionen übertragen.

Die Normierung erfolgt nach folgender Formel: $\boldsymbol{A_n = 1 + 4 \cdot \frac{A-1}{8}}$ bzw. $\boldsymbol{A_n = \frac{A+1}{2}}$

[7] Cox 2008

Damit gilt:

Autonomiestufe	A1	A2	A3	A4	A5	A6	A7	A8	A9
Normierter Autonomiewert A_n	1,0	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0

Tabelle 1: Überführung der qualitativen Autonomiestufen A1 bis A9 in den normierten Autonomiewert A_n

Für die Scorebildung werden somit alle vier Dimensionen auf einem gemeinsamen Wertebereich von 1 bis 5 geführt:

A_n = normierte Autonomie

S = Severität

E = Exposition

I = Irreversibilität

Diese Normierung verändert nicht die qualitative Bewertung der Autonomie. Sie verhindert lediglich, dass die differenziertere Autonomieskala automatisch zu einer verdeckten mathematischen Gewichtung führt. Autonomie bleibt inhaltlich zentral, wird aber rechnerisch nicht stärker gewichtet als Severität, Exposition oder Irreversibilität.

2.2.3 Multiplikativer Rohwert und logarithmische Darstellung

Die Scorebildung des ASEI-10-Modells erfolgt in zwei Schritten. Zunächst wird ein multiplikativer Rohwert gebildet:

$$\boldsymbol{R = A_n \cdot S \cdot E \cdot I}$$

Dieser Rohwert bildet die Annahme ab, dass sich die vier Dimensionen im Anwendungskontext wechselseitig verstärken können. Ein autonomeres System ist risikorelevanter, wenn es in einem folgenreichen Einsatzbereich arbeitet. Eine hohe Severität ist risikorelevanter, wenn viele Personen oder Systeme betroffen sind. Eine breite Exposition ist risikorelevanter, wenn negative Folgen schwer rückholbar sind.

Der Rohwert ist jedoch für die praktische Interpretation unhandlich. Da alle vier Dimensionen für die Berechnung auf den Wertebereich 1 bis 5 gebracht werden, kann der Rohwert Werte zwischen 1 und 625 annehmen:

$$\boldsymbol{R_{min} = 1 \cdot 1 \cdot 1 \cdot 1 = 1}$$

$$\boldsymbol{P_{max} = 5 \cdot 5 \cdot 5 \cdot 5 = 625}$$

Um den Rohwert in eine besser interpretierbare Skala zu übertragen, wird er logarithmisch normiert:

$$\boldsymbol{\text{ASEI-10} = 1 + 9 \cdot \frac{\ln(P)}{\ln(625)}}$$

Die logarithmische Normierung erhält die Rangordnung des Rohwerts, verdichtet aber hohe Rohwerte und überführt alle möglichen Kombinationen auf eine Skala von 1 bis 10. Dabei ist ausdrücklich zu beachten, dass der normierte Score kein linearer Risikowert ist.

Mathematisch gilt:

$$\boldsymbol{\ln(A_n \cdot S \cdot E \cdot I) = \ln(A_n) + \ln(S) + \ln(E) + \ln(I)}$$

Der normierte ASEI-10-Score ist daher ein logarithmisch transformierter Index. Die Verstärkungslogik liegt im multiplikativen Rohwert. Die normierte Darstellung macht diesen Rohwert praktikabel vergleichbar, ohne zu behaupten, dass der finale Score eine lineare Messung von Risiko darstellt.

2.3 Dimensionsabgrenzung und Bewertungszustand

2.3.1 Zur Überschneidung und Abgrenzung der Dimensionen

Die vier Dimensionen des ASEI-10-Modells sind analytisch getrennt, aber in realen Anwendungskontexten nicht immer empirisch unabhängig. Ein systemkritischer Einsatzbereich kann häufig mit hoher Exposition und hoher Irreversibilität verbunden sein. Kritische Infrastruktur ist dafür ein naheliegendes Beispiel: Dort können hohe Severität, systemische Exposition und schwer rückholbare Folgen gemeinsam auftreten.

Diese Korrelationen bedeuten nicht, dass die Dimensionen identisch sind. Sie weisen jedoch darauf hin, dass ihre Abgrenzung sorgfältig begründet werden muss. Severität fragt nach der Schwere möglicher Folgen im betroffenen Schutzbereich. Exposition fragt nach der Breite des Wirkungsraums. Irreversibilität fragt nach der Rückholbarkeit negativer Folgen. Diese Fragen können im Einzelfall zum selben hohen Ergebnis führen, erfassen aber unterschiedliche Risikoeigenschaften.

Um eine unreflektierte Mehrfachzählung zu vermeiden, muss jede Dimension eigenständig begründet werden. Ein System darf nicht allein deshalb automatisch E5 oder I5 erhalten, weil es bereits S5 erreicht. Vielmehr ist getrennt zu prüfen: Welche Schutzgüter sind betroffen? Wie weit können sich die Wirkungen ausbreiten? Wie gut lassen sich negative Folgen korrigieren oder kompensieren? Erst wenn diese Fragen jeweils eigenständig eine hohe Stufe rechtfertigen, ist eine entsprechende Mehrfachausprägung sachgerecht.

Die diskriminante Qualität des Modells hängt daher nicht allein von den Skalen ab, sondern von der Sorgfalt der Anwendung. In Grenzfällen sollte die Begründung ausdrücklich erklären, warum eine hohe Einstufung in mehreren Dimensionen vorliegt und welche Risikomerkmale jeweils unterschiedlich bewertet wurden.

2.3.2 Korrelative Beziehungen zwischen den vier Bewertungsdimensionen

Die vier ASEI-Dimensionen sind analytisch getrennt, können im konkreten Anwendungskontext aber korrelativ zusammenwirken. Diese Korrelationen sind nicht als empirisch gemessene statistische Korrelationskoeffizienten zu verstehen. Gemeint sind sachlogische Interdependenzen: Bestimmte Systemmerkmale erhöhen häufig mehrere Risikodimensionen zugleich, ohne dass diese Dimensionen deshalb identisch wären.

Autonomie und Severität

Autonomie und Severität können sich gegenseitig verstärken, wenn ein KI-System in einem folgenreichen Einsatzbereich nicht nur unterstützt, sondern Handlungsschritte plant, Entscheidungen vorbereitet oder operative Maßnahmen ausführt. Je schwerwiegender der Einsatzbereich ist, desto bedeutsamer wird bereits eine begrenzte Autonomie. Umgekehrt ist hohe Autonomie in einem geringfügigen Einsatzbereich weniger kritisch als dieselbe Autonomie in Bildung, Beschäftigung, Gesundheit, Recht, öffentlicher Sicherheit oder kritischer Infrastruktur.

Beispiel: Ein KI-Agent, der private Textentwürfe sortiert, besitzt trotz höherer Autonomie nur begrenzte Severität. Ein System, das Bewerbungen vorsortiert oder sicherheitsrelevante Warnmeldungen priorisiert, kann dagegen bereits bei niedrigerer Autonomie eine hohe Severität erreichen.

Autonomie und Exposition

Autonomie und Exposition korrelieren, wenn ein System durch eigenständige Handlungsausführung, Wiederholung, Skalierung oder dauerhafte Aktivität einen größeren Wirkungsraum erreicht. Agentische Systeme können Wirkungen nicht nur erzeugen, sondern auch verbreiten, wiederholen oder in mehrere Prozesse einspeisen. Dennoch sind Autonomie und Exposition getrennt zu bewerten: Ein wenig autonomes System kann eine hohe Exposition haben, wenn es öffentlich oder massenhaft genutzt wird; ein hoch autonomes System kann eine niedrige Exposition haben, wenn es streng isoliert bleibt.

Beispiel: Ein reaktiver Kundenservice-Chatbot kann wegen hoher Nutzerzahl E4 erreichen, obwohl seine Autonomie nur A3 beträgt. Ein interner Agent mit A5 kann dagegen bei enger Begrenzung auf eine kleine Abteilung nur E3 erreichen.

Autonomie und Irreversibilität

Autonomie und Irreversibilität hängen zusammen, wenn eigenständige Systemhandlungen Korrekturfenster verkürzen oder Folgeprozesse auslösen. Je selbstständiger ein System handelt, desto eher können Fehler bereits wirksam geworden sein, bevor sie entdeckt werden. Dennoch ist Irreversibilität nicht allein eine Folge von Autonomie. Auch ein System mit geringer Autonomie kann schwer reversible Folgen verursachen, wenn seine Ausgaben in kritische Entscheidungen übernommen werden.

Beispiel: Ein Agent, der automatisch E-Mails versendet, Daten verändert oder Prozesse anstößt, erhöht die Rückholproblematik. Eine rein assistive KI in der Personalvorauswahl kann aber ebenfalls schwer reversible Folgen haben, wenn ihre Vorauswahl faktisch über berufliche Chancen entscheidet.

Severität und Exposition

Severität und Exposition korrelieren häufig, weil folgenschwere Einsatzbereiche bei breiter Wirkung besonders kritisch werden. Ein hochsensibler Einsatzbereich wird risikorelevanter, wenn viele Personen, Organisationseinheiten oder externe Empfänger betroffen sind. Trotzdem sind beide Dimensionen klar zu trennen: Severität beschreibt die Schwere der möglichen Folgen; Exposition beschreibt die Breite des Wirkungsraums.

Beispiel: Ein System zur medizinischen Entscheidungsunterstützung in einer einzelnen Fachabteilung kann eine hohe Severität, aber begrenzte Exposition besitzen. Ein öffentlich zugänglicher allgemeiner Textgenerator kann eine hohe Exposition, aber geringe Severität aufweisen, wenn keine sensiblen Entscheidungen oder Schutzgüter betroffen sind.

Severität und Irreversibilität

Severität und Irreversibilität können eng zusammenhängen, weil schwere Folgen häufig schwerer korrigierbar sind. Gesundheitliche Schäden, Benachteiligungen im Beschäftigungskontext, sicherheitskritische Fehlreaktionen oder erhebliche Reputationsschäden lassen sich häufig nicht vollständig rückgängig machen. Dennoch erfassen beide Dimensionen unterschiedliche Fragen: Severität fragt, wie schwer die mögliche negative Folge ist; Irreversibilität fragt, wie gut diese Folge nachträglich erkannt, gestoppt, korrigiert oder kompensiert werden kann.

Beispiel: Eine falsche Auskunft zu einem geringfügigen Produktdetail hat geringe Severität und ist meist leicht korrigierbar. Eine fehlerhafte Ablehnung im Bewerbungsprozess hat höhere Severität und kann trotz späterer Korrektur berufliche Chancen, Vertrauen und Reputation beeinträchtigen.

Exposition und Irreversibilität

Exposition und Irreversibilität korrelieren, wenn eine breite Wirkung die Rückholung erschwert. Je mehr Personen, Systeme, Plattformen oder Organisationseinheiten eine fehlerhafte Ausgabe erreicht haben, desto schwieriger wird es, diese vollständig zu korrigieren. Besonders öffentliche, verteilte oder systemische Wirkungen können trotz späterer Richtigstellung Restschäden hinterlassen. Gleichwohl kann auch ein eng begrenzter Einzelfall hoch irreversibel sein, wenn die Folge für die betroffene Person schwerwiegend und nicht kompensierbar ist.

Beispiel: Eine fehlerhafte interne Notiz an eine einzelne Person ist meist leicht korrigierbar. Eine fehlerhafte öffentliche Mitteilung, die von Dritten gespeichert, zitiert oder weiterverbreitet wird, ist deutlich schwerer rückholbar. Umgekehrt kann auch eine einzelne falsche sicherheitskritische Handlung schwer irreversible Folgen haben, selbst wenn nur wenige Personen unmittelbar betroffen sind.

Zusammenfassung

Die vier Dimensionen sind nicht unabhängig im strengen empirischen Sinne. Sie dürfen aber nicht miteinander verwechselt werden. Eine hohe Einstufung in einer Dimension rechtfertigt nicht automatisch eine hohe Einstufung in einer anderen Dimension. Jede Dimension muss eigenständig begründet werden. Gerade die korrelativen Beziehungen zeigen, warum ASEI-10 nicht nur einen Gesamtscore ausweist, sondern zusätzlich die Einzeldimensionen und Prüfschwellen dokumentiert.

2.3.3 Brutto- und Netto-Bewertung des Risikopotenzials

Eine ASEI-10-Bewertung kann sich auf unterschiedliche Bewertungszustände beziehen. Methodisch ist deshalb zu unterscheiden, ob das Risikopotenzial eines KI-Systems vor oder nach risikomindernden Maßnahmen bewertet wird. Diese Unterscheidung entspricht der Differenz zwischen Brutto- und Netto-Bewertung.

Die Brutto-Bewertung beschreibt das inhärente Risikopotenzial eines KI-Systems im vorgesehenen Anwendungskontext, bevor zusätzliche Schutz-, Kontroll- oder Begrenzungsmaßnahmen berücksichtigt werden. Sie beantwortet die Frage: Welches Risikopotenzial ergibt sich aus den Funktionen, Berechtigungen, Datenzugriffen, Nutzergruppen, Einsatzgrenzen und möglichen Folgen des Systems, wenn keine ergänzenden Mitigationen angesetzt werden?

Die Netto-Bewertung beschreibt demgegenüber das verbleibende Risikopotenzial nach wirksam implementierten und überprüfbaren Schutzmaßnahmen. Sie beantwortet die Frage: Welches Risikopotenzial bleibt bestehen, wenn technische, organisatorische oder prozessuale Maßnahmen tatsächlich greifen und die relevanten Risikodimensionen nachweisbar verändern?

Diese Unterscheidung ist wichtig, weil Schutzmaßnahmen nicht pauschal auf den Gesamtscore angerechnet werden dürfen. Eine Maßnahme senkt den ASEI-10-Score nur dann, wenn sie eine der vier Bewertungsdimensionen tatsächlich verändert. Allgemeine Sorgfalt, bloße Absichtserklärungen, unverbindliche Nutzungsrichtlinien oder nicht überprüfte Kontrollen rechtfertigen keine niedrigere Einstufung.

Für die Bewertung der Autonomiedimension ist zusätzlich zu unterscheiden, ob eine Begrenzung Bestandteil der Systemarchitektur ist oder erst nachträglich als ergänzende Schutzmaßnahme hinzutritt. Systemimmanente Begrenzungen, etwa fest eingebaute Freigabepunkte, technisch erzwungene Domänengrenzen, rollenbasierte Berechtigungen oder fehlende Toolzugriffe, gehören zum Bewertungsgegenstand selbst und sind bereits bei der Einstufung der Autonomie zu berücksichtigen. Ein System, das aufgrund fest eingebauter Freigabepunkte nur überwacht handeln kann, ist deshalb nicht zunächst „brutto“ als höher autonom einzustufen, wenn diese Freigabepunkte integraler Bestandteil der Systemgestaltung sind.

Davon zu unterscheiden sind ergänzende Mitigationen, die außerhalb der eigentlichen Systemarchitektur liegen oder erst nachträglich durch organisatorische Maßnahmen, zusätzliche Kontrollen, manuelle Prüfprozesse oder Governance-Vorgaben eingeführt werden. Solche Maßnahmen können nur dann in einer Netto-Bewertung berücksichtigt werden, wenn sie wirksam, verbindlich, überprüfbar und dauerhaft implementiert sind.

Schutzmaßnahme	**Mögliche Wirkung auf die ASEI-Dimensionen**	**Voraussetzung für Berücksichtigung in der Netto-Bewertung**
Freigabepunkte vor kritischen Handlungen	kann Autonomie senken, wenn das System nicht mehr eigenständig handeln darf	Freigabe ist verpflichtend, technisch oder organisatorisch wirksam und nicht leicht umgehbar
Einschränkung von Toolzugriffen oder Berechtigungen	kann Autonomie und gegebenenfalls Exposition senken	Zugriffe sind tatsächlich entfernt, begrenzt oder rollenbasiert kontrolliert
Begrenzung auf interne Nutzung	kann Exposition senken	externe Weitergabe, Veröffentlichung oder Massenverteilung ist wirksam ausgeschlossen
Versionierung, Rücksetzmechanismen und Rollback	kann Irreversibilität senken	Korrektur und Wiederherstellung sind praktisch, zeitnah und vollständig möglich
Korrektur- und Beschwerdeverfahren	kann Irreversibilität senken, wenn Folgen realistisch gemindert werden können	Verfahren sind erreichbar, wirksam, dokumentiert und für Betroffene nutzbar
Monitoring und Protokollierung	senkt nicht automatisch eine Dimension, kann aber Erkennung, Nachvollziehbarkeit und spätere Korrektur unterstützen	nur scorewirksam, wenn dadurch Folgen tatsächlich früher gestoppt oder rückholbarer werden
Fachliche Prüfung vor Veröffentlichung oder Nutzung	kann Exposition oder Irreversibilität senken, wenn fehlerhafte Ergebnisse vor Wirkungseintritt abgefangen werden	Prüfung ist verbindlich, qualifiziert und findet vor Außenwirkung oder Entscheidungswirkung statt
Schulung und Nutzungsrichtlinien	senkt den Score nicht automatisch	nur relevant, wenn dadurch nachweisbar Systemrechte, Nutzungskontext oder Wirkungsraum begrenzt werden

Tabelle 2: Wirkung von Schutzmaßnahmen auf die Brutto- und Netto-Bewertung

Die Dimension Severität ist durch Schutzmaßnahmen nur eingeschränkt reduzierbar. Sie beschreibt die mögliche Schwere negativer Folgen im betroffenen Einsatzbereich. Wenn ein KI-System im Recruiting, in medizinischen Anwendungen, in behördlichen Verfahren oder in kritischer Infrastruktur eingesetzt wird, bleibt der Einsatzbereich grundsätzlich folgenrelevant. Schutzmaßnahmen können dort zwar die Wahrscheinlichkeit, die Exposition oder die Rückholbarkeit beeinflussen, ändern aber nicht ohne Weiteres die Bedeutung der berührten Schutzgüter. Eine niedrigere Einstufung des Severitätsgrades ist nur dann gerechtfertigt, wenn sich der Anwendungskontext selbst verändert, etwa durch Herausnahme aus einem hochkritischen Entscheidungsprozess oder durch Beschränkung auf unverbindliche, folgenarme Vorarbeiten.

Für die praktische Anwendung sollte daher stets dokumentiert werden, ob eine ASEI-10-Bewertung als Brutto- oder Netto-Bewertung vorgenommen wurde. Bei höherem Risikopotenzial kann es sinnvoll sein, beide Werte auszuweisen: zunächst das inhärente Risikopotenzial ohne zusätzliche Mitigationen und anschließend das verbleibende Risikopotenzial nach wirksam umgesetzten Schutzmaßnahmen.

Eine solche doppelte Ausweisung erhöht die Transparenz. Sie macht sichtbar, ob ein KI-System von vornherein niedriges Risikopotenzial besitzt oder ob ein ursprünglich hohes Risikopotenzial erst durch Kontrollmechanismen, Freigaben, Begrenzungen oder Rückholbarkeitsmaßnahmen reduziert wurde. Zugleich verhindert sie, dass Schutzmaßnahmen pauschal oder doppelt angerechnet werden.

Für ASEI-10 gilt deshalb folgende Grundregel:

> *Eine Schutzmaßnahme darf nur dann zu einer niedrigeren Netto-Einstufung führen, wenn sie die bewertete Dimension tatsächlich, nachweisbar und dauerhaft verändert. Andernfalls bleibt sie eine ergänzende Mindestanforderung oder Governance-Maßnahme, ohne den Score selbst zu senken.*

2.4 Kalibrierung, Validierung und Anwendungskonsequenzen

2.4.1 Kalibrierung und Interpretationsbereiche

Die Interpretationsbereiche des ASEI-10-Scores dienen der praktischen Orientierung. Sie ordnen Scorewerte in Bereiche wie niedrig, moderat, erhöht, hoch oder sehr hoch ein. Diese Bereiche sind in der vorliegenden Fassung heuristisch gesetzt. Sie beruhen nicht auf empirisch validierten Schadensdaten, statistischer Kalibrierung oder breiter Expertenkonsensmessung.

Daraus folgt, dass die Interpretationsbereiche vorsichtig zu verwenden sind. Sie sollen Organisationen bei Priorisierung und Kommunikation unterstützen, dürfen aber nicht als objektiv validierte Risikogrenzen verstanden werden. Insbesondere bei Grenzwerten zwischen zwei Bereichen ist die qualitative Bewertung wichtiger als die Dezimalstelle des Scores.

Eine Weiterentwicklung des Modells sollte die Interpretationsbereiche empirisch oder expertengestützt prüfen. Denkbar sind Vergleichsbewertungen durch unabhängige Fachleute, Sensitivitätsanalysen, Inter-Rater-Reliabilitätsstudien, Anwendung auf reale KI-Inventare oder Abgleiche mit bestehenden Risikokategorien aus Regulierung und Praxis. Bis zu einer solchen Validierung sind die Scorebereiche als begründete, aber vorläufige Klassifikationshilfe zu verstehen.

2.4.2 Validierungsbedarf und Design-Science-Perspektive

ASEI-10 ist in der vorliegenden Form ein methodischer Vorschlag. Das Modell wurde heuristisch entwickelt, durch Skalen operationalisiert und anhand von Fallbeispielen illustriert. Damit ist es jedoch noch nicht empirisch validiert. Die Fallbeispiele zeigen, wie das Modell angewendet werden kann, beweisen aber nicht, dass es in unterschiedlichen Organisationen, Branchen oder Bewertungsgruppen zuverlässig zu denselben oder besonders trennscharfen Ergebnissen führt.

Eine geeignete Weiterentwicklung kann das Modell als Design-Science-Artefakt verstehen.[8] In dieser Perspektive wird ein praktisches Problem identifiziert, ein methodisches Artefakt entwickelt, seine Anwendung demonstriert und anschließend systematisch evaluiert. Für ASEI-10 würde dies bedeuten, dass das Modell in realen oder realitätsnahen Bewertungssettings getestet, mit Expertenurteilen verglichen und iterativ verbessert wird.

Mögliche Validierungsschritte sind insbesondere:

- Anwendung des Modells auf eine größere Zahl realer KI-Systeme
- Bewertung derselben Fälle durch mehrere unabhängige Fachleute
- Prüfung der Inter-Rater-Reliabilität
- Analyse typischer Grenzfälle und Bewertungsabweichungen
- Sensitivitätsanalyse der Scoreformel und Interpretationsbereiche
- Vergleich mit bestehenden AI-Risk-Taxonomien und Governance-Frameworks
- Überprüfung der Praxistauglichkeit in KI-Inventaren und Freigabeprozessen

Bis zu einer solchen Validierung sollte ASEI-10 nicht als empirisch gesichertes Messinstrument bezeichnet werden. Angemessen ist die Bezeichnung als heuristisches, operationalisiertes und weiter validierungsbedürftiges Klassifikationsmodell.

2.4.3 Empirische Prüfbarkeit und Widerlegungsbedingungen

Das ASEI-10-Modell ist ein konstruiertes Bewertungsartefakt. Seine Definitionen, Skalen, die multiplikative Rohwertbildung, die logarithmische Normierung und die Interpretationsbereiche sind Festsetzungen und damit nicht im Sinne einer empirischen Theorie widerlegbar. Ebenso wenig ist der normative Kern des Modells – die Empfehlung, wie Risikopotenzial einzuordnen sei – wahrheitsfähig. Diese Bestandteile sind als methodischer Rahmen zu verstehen und nicht Gegenstand einer Falsifikation.

Daraus folgt, dass das Modell als Ganzes nicht am Kriterium der Falsifizierbarkeit zu messen ist, sondern an den Gütekriterien eines Bewertungsinstruments: Reliabilität, Validität, Nachvollziehbarkeit und Zweckmäßigkeit. Die Falsifizierbarkeit betrifft ausschließlich die im Modell enthaltenen empirischen Teilbehauptungen. Sie ist eine zusätzliche Schärfung, keine Voraussetzung für die Gültigkeit des Modells.

[8] Hevner et al. 2004

Innerhalb dieses Rahmens enthält das Modell jedoch empirische Teilbehauptungen, die sich als prüfbare Hypothesen mit ausdrücklichen Widerlegungsbedingungen formulieren lassen. Damit lassen sich die empirischen Teilbehauptungen des Modells – über seine Rahmung als Design-Science-Artefakt hinaus – auch an einem an Karl Popper[9] orientierten Maßstab messen: Das Modell benennt vorab, welche Beobachtungen es als widerlegt oder revisionsbedürftig gelten ließe. Die folgenden vier Hypothesen sind nicht Ergebnis, sondern Programm; die genannten Schwellenwerte sind vor Durchführung einer Studie verbindlich festzulegen.

H1 – Reproduzierbarkeit (Reliabilität). Das Modell behauptet, dass unabhängige, nach einheitlichem Protokoll geschulte Bewerter dasselbe KI-System in dieselbe Risikoklasse einordnen. Diese Übereinstimmung ist messbar, etwa über Krippendorffs α[10] auf der ordinalen RK-Skala. Die Hypothese gilt als bestätigt, wenn die Übereinstimmung einen vorab festgelegten Mindestwert erreicht (vorgeschlagen: $\alpha \geq 0{,}67$, angestrebt $\alpha \geq 0{,}80$).

Widerlegungsbedingung: Liegt die Übereinstimmung über mehrere Bewerter und eine definierte Systemstichprobe systematisch unterhalb dieses Werts, ist der Anspruch des Modells auf eine reproduzierbare Klassifikation falsifiziert.

H2 – Konvergente Validität. Das Modell behauptet, tatsächlich Risikopotenzial zu erfassen und nicht eine davon unabhängige Größe. Daraus folgt eine riskante Vorhersage: Systeme, die ASEI-10 in die Klassen RK5 oder RK6 einordnet, sollten auch von unabhängigen Bezugsmaßstäben – etwa der Hochrisiko-Logik des EU AI Act oder dem holistischen Urteil einschlägiger Fachleute – als hochriskant eingestuft werden.

Widerlegungsbedingung: Stuft ASEI-10 jene Systeme, die unabhängige Maßstäbe als hochriskant bewerten, systematisch niedrig ein (oder umgekehrt), und überschreitet die Übereinstimmung das Zufallsniveau nicht, ist der Anspruch widerlegt, mit dem Score Risikopotenzial abzubilden.

H3 – Trennschärfe (diskriminante Validität). Das Modell behauptet, dass seine vier Dimensionen eigenständige Risikobeiträge erfassen. Wie in Abschnitt 2.3.2 ausgeführt, sind dabei sachlogische Korrelationen ausdrücklich zugelassen; behauptet wird nicht Unabhängigkeit, sondern Nicht-Identität der Dimensionen.

Widerlegungsbedingung: Zeigt sich an einer hinreichend großen Stichprobe realer Bewertungen, dass zwei Dimensionen praktisch vollständig zusammenfallen (etwa eine paarweise Rangkorrelation ρ[11] $> 0{,}95$), so ist die betreffende Dimension redundant; die Annahme einer vierdimensionalen Struktur wäre für dieses Paar falsifiziert und das Modell entsprechend auf drei tragende Dimensionen zu reduzieren.

[9] Popper 1935/2005

[10] Krippendorffs Alpha (α) ist ein Maß für die Übereinstimmung zwischen mehreren unabhängigen Bewertern. Es gibt an, wie zuverlässig verschiedene Personen denselben Sachverhalt gleich einstufen, und ist gegenüber unterschiedlichen Skalenniveaus und fehlenden Werten robust. Der Wertebereich reicht von 0 bis 1; höhere Werte zeigen eine größere Übereinstimmung an. Ein Wert von 1 bedeutet vollständige Übereinstimmung, ein Wert um 0 eine Übereinstimmung, die nicht über den Zufall hinausgeht.

[11] Spearmans Rangkorrelationskoeffizient (ρ) misst die Stärke und Richtung eines monotonen Zusammenhangs zwischen zwei Größen, also die Tendenz, dass hohe Werte der einen Größe mit hohen (oder mit niedrigen) Werten der anderen einhergehen. Anders als ein linearer Korrelationskoeffizient beruht er auf den Rängen der Werte, nicht auf deren Abständen. Der Wertebereich reicht von −1 bis +1; Werte nahe ±1 zeigen einen starken, Werte nahe 0 einen schwachen monotonen Zusammenhang an.

H4 – Prädiktive Validität. In ihrer stärksten Form behauptet das Modell, dass ein höheres Risikopotenzial mit einer höheren Wahrscheinlichkeit oder Schwere nachteiliger Folgen einhergeht. Daraus folgt: Über einen definierten Beobachtungszeitraum sollten Systeme mit höherem ASEI-10-Score – bei vergleichbarer Exposition – häufiger oder schwerwiegender von dokumentierten nachteiligen Ereignissen betroffen sein als Systeme mit niedrigerem Score.

Widerlegungsbedingung: Lässt sich expositionsbereinigt kein monotoner Zusammenhang zwischen Score und Ereignisrate nachweisen, ist der Kernanspruch des Scores untergraben. Diese Hypothese ist zugleich die am schwersten prüfbare: Lange Wirkungshorizonte, Störgrößen und Datenmangel erschweren die Operationalisierung, und nach dem Duhem-Quine-Problem kann ein negatives Ergebnis auch auf Hilfsannahmen (etwa die Erfassung der Ereignisse) statt auf das Modell selbst zurückgehen. H4 ist daher als langfristiges Forschungsziel zu verstehen, nicht als kurzfristig entscheidbarer Test.

Die wissenschaftliche Ernsthaftigkeit dieser Selbstbindung hängt davon ab, dass widersprechende Ergebnisse nicht nachträglich wegerklärt werden. Bewerterqualifikation, Schulungs- und Anwendungsprotokoll sowie die Auswahl der Bezugsmaßstäbe sind deshalb vor Durchführung einer Studie festzulegen, damit Abweichungen nicht beliebig als Anwendungsfehler umgedeutet werden können. Erst diese Vorab-Festlegung unterscheidet eine prüfbare Hypothese von einer gegen Widerlegung immunisierten Behauptung.

Die vier Hypothesen zu den empirisch prüfbaren Teilbehauptungen des Modells lassen sich abschließend in einer Übersicht zusammenfassen, die jeder Hypothese ihren Prüfgegenstand, ihre Operationalisierung und ihre Widerlegungsbedingung zuordnet. Die Übersicht dient der schnellen Orientierung und dem späteren Studiendesign; sie ersetzt nicht die ausführliche Begründung der einzelnen Hypothesen. Die angegebenen Schwellenwerte sind vorgeschlagene Richtwerte, die vor Durchführung einer empirischen Untersuchung verbindlich festzulegen sind.

Hypothese	Prüfgegenstand (Behauptung des Modells)	Operationalisierung	Widerlegungsbedingung
H1 – Reliabilität (Reproduzierbarkeit)	Unabhängige, einheitlich geschulte Bewerter ordnen dasselbe System derselben Risikoklasse zu.	Inter-Rater-Übereinstimmung über eine definierte Systemstichprobe; Krippendorffs α auf der ordinalen RK-Skala (vorgeschlagen: $\alpha \geq 0{,}67$; angestrebt $\geq 0{,}80$).	Die Übereinstimmung liegt über mehrere Bewerter systematisch unter dem festgelegten Mindestwert.
H2 – Konvergente Validität	Systeme in RK5/RK6 werden auch von unabhängigen Maßstäben als hochriskant eingestuft.	Abgleich der ASEI-Hochklassen mit Bezugsmaßstäben (EU-AI-Act-Hochrisiko, holistisches Expertenurteil) an einer Systemstichprobe.	Die Übereinstimmung überschreitet nicht das Zufallsniveau oder verläuft systematisch gegenläufig.
H3 – Diskriminante Validität (Trennschärfe)	Die vier Dimensionen erfassen eigenständige Risikobeiträge; Korrelation ist zulässig, Identität nicht.	Paarweise Rangkorrelationen (Spearman ρ) der Dimensionen an realen Bewertungen.	Ein Dimensionspaar fällt praktisch vollständig zusammen (z. B. $\rho > 0{,}95$) und ist damit redundant.
H4 – Prädiktive Validität	Höheres Risikopotenzial geht mit häufigeren oder schwereren nachteiligen Folgen einher.	Expositionsbereinigter Zusammenhang zwischen Score bzw. RK-Klasse und dokumentierten nachteiligen Ereignissen über einen definierten Zeitraum.	Es lässt sich kein monotoner Zusammenhang nachweisen (bzw. ein inverser). Einschränkung: schwer isolierbar (Duhem-Quine), langfristiges Ziel.

Tabelle 3: Prüfbare Hypothesen des ASEI-10-Modells mit Operationalisierung und Widerlegungsbedingung

2.4.4 Konsequenzen für Interpretation und Anwendung

Aus der methodischen Einordnung ergeben sich mehrere Grundregeln für die Anwendung des ASEI-10-Modells.

Erstens bewertet ASEI-10 das Risikopotenzial, nicht die Eintrittswahrscheinlichkeit konkreter Schadensereignisse. Wahrscheinlichkeits-, Robustheits- und Zuverlässigkeitsanalysen müssen bei Bedarf ergänzend durchgeführt werden.

Zweitens ist der ASEI-10-Score ein heuristischer Index. Er dient der Orientierung, Priorisierung und Dokumentation, nicht der exakten Messung objektiver Risikoeinheiten.

Drittens dürfen die vier Einzeldimensionen nicht durch den Gesamtscore ersetzt werden. Eine belastbare Bewertung muss stets A, A_n, S, E, I, Rohwert, Score, Prüfschwellen und Begründung ausweisen.

Viertens sind die Interpretationsbereiche des Scores als vorläufige Orientierung zu verstehen. Sie können und sollten durch empirische Anwendung, Expertenurteile und Sensitivitätsanalysen weiter kalibriert werden.

Fünftens muss jede Einstufung kontextbezogen erfolgen. Bewertet wird nicht ein KI-Modell an sich, sondern ein konkretes KI-System mit bestimmten Funktionen, Rechten, Datenzugriffen, Nutzergruppen und Einsatzgrenzen.

Mit dieser methodischen Rahmung wird der Anspruch des ASEI-10-Modells präzisiert. Das Modell bietet eine strukturierte und nachvollziehbare Klassifikationslogik für KI-Risikopotenziale. Es erhebt jedoch keinen Anspruch, klassische Risikoanalyse, rechtliche Prüfung, technische Evaluation oder empirische Validierung zu ersetzen.

3 Bestehende Rahmenwerke und Bewertungsansätze zur KI-Risikobewertung

3.1 Funktion des Bezugsrahmens

Das ASEI-10-Modell steht im Kontext bestehender rechtlicher, normativer, organisatorischer und wissenschaftlicher Ansätze zur Beschreibung, Bewertung und Steuerung von KI-Risiken. Zu diesen Ansätzen gehören insbesondere regulatorische Rahmenwerke wie der EU AI Act, Governance- und Risikomanagementmodelle wie das NIST AI Risk Management Framework, internationale Normen wie ISO/IEC 23894 und ISO/IEC 42001, KI-spezifische Risikotaxonomien, AI-Safety-Berichte, technische Evaluationsverfahren sowie klassische Fehleranalyse- und Bewertungsmethoden.

ASEI-10 versteht sich in diesem Umfeld nicht als nachgeordnetes Hilfsinstrument, sondern als eigenständiger methodischer Beitrag. Das Modell verfolgt einen anderen Zweck als rechtliche Regelwerke, Managementsystemnormen oder technische Prüfverfahren, ist diesen aber in seiner Funktion nicht untergeordnet. Während regulatorische Rahmenwerke Pflichten, Verbote und Zuständigkeiten definieren, Managementsysteme organisatorische Steuerungsprozesse beschreiben und technische Prüfverfahren konkrete Leistungs-, Sicherheits- oder Robustheitsfragen untersuchen, setzt ASEI-10 an einer eigenständigen Bewertungsfrage an:

> *Wie lässt sich das strukturelle Risikopotenzial eines konkreten KI-Systems in einem konkreten Anwendungskontext transparent, vergleichbar und begründet klassifizieren?*

Diese Frage wird von bestehenden Rahmenwerken meist nur mittelbar beantwortet. Der EU AI Act ordnet KI-Systeme rechtlich ein, liefert aber keinen allgemeinen Score zur operativen Priorisierung einzelner Systeme. Das NIST AI Risk Management Framework beschreibt einen umfassenden Risikomanagementprozess, stellt aber selbst kein kompaktes Klassifikationsmodell für das Risikopotenzial konkreter KI-Anwendungen bereit. ISO/IEC 23894 und ISO/IEC 42001 geben wichtige Leitlinien für Risikomanagement und Managementsysteme, ersetzen aber keine kontextbezogene Einstufungslogik für einzelne KI-Systeme. Risikotaxonomien und Safety Reports beschreiben Gefährdungsarten, Risikofelder und Entwicklungslinien, führen diese aber nicht zwingend zu einer anwendungsbezogenen Klassifikation zusammen.

Genau an dieser Schnittstelle positioniert sich ASEI-10. Das Modell übersetzt zentrale Risikotreiber von KI-Systemen in eine strukturierte, dokumentierbare und rechnerisch nachvollziehbare Klassifikationslogik. Es bewertet nicht KI im Allgemeinen und nicht ein Modell in abstrakter Form, sondern ein konkretes System mit bestimmten Funktionen, Berechtigungen, Datenzugriffen, Nutzergruppen, Einsatzgrenzen und möglichen Folgen. Dadurch entsteht ein operativer Bewertungsrahmen, der zwischen abstrakter Regulierung, organisatorischer Governance und technischer Detailprüfung vermittelt.

Die Eigenständigkeit des Modells liegt vor allem in der Verbindung von vier Bewertungsdimensionen: Autonomie, Severität, Exposition und Irreversibilität. Diese Dimensionen erfassen unterschiedliche Risikotreiber, die im Zusammenspiel das strukturelle Risikopotenzial eines KI-Systems prägen. Die Scorebildung verdichtet diese Bewertung zu einem ASEI-10-Score, ohne die qualitative Begründung zu ersetzen. Ergänzende Prüfschwellen stellen sicher, dass kritische Einzelmerkmale nicht durch den Gesamtscore verdeckt werden.

ASEI-10 konkurriert damit nicht im engen Sinn mit bestehenden Rahmenwerken, sondern besetzt eine eigene methodische Ebene. Das Modell kann rechtliche Prüfungen vorbereiten, Governance-Entscheidungen unterstützen, Risikoinventare strukturieren, Auditvorbereitungen erleichtern und technische Detailprüfungen priorisieren. Seine Stärke liegt gerade darin, dass es weder ein bloßer Kriterienkatalog noch ein vollständiges Managementsystem ist, sondern eine fokussierte Klassifikationslogik für konkrete KI-Systeme bereitstellt.

Die Einordnung in bestehende Ansätze dient deshalb nicht dazu, ASEI-10 defensiv von anderen Rahmenwerken abzugrenzen. Sie dient vielmehr dazu, das eigenständige Profil des Modells sichtbar zu machen. ASEI-10 beansprucht nicht, rechtliche, normative oder technische Prüfungen zu ersetzen. Es beansprucht aber, eine bisher nicht ausreichend operationalisierte Bewertungsfrage systematisch zu beantworten: Wie lässt sich das Risikopotenzial eines konkreten KI-Systems im Anwendungskontext nachvollziehbar einstufen?

Dieses Kapitel arbeitet diese Position schrittweise heraus. Zunächst wird ASEI-10 zu rechtlichen und regulatorischen Rahmenwerken in Beziehung gesetzt. Anschließend folgt die Abgrenzung zu Governance-, Risiko- und Managementsystemansätzen. Danach werden klassische Bewertungs- und Fehleranalyseverfahren, KI-spezifische Risikotaxonomien sowie Scoring-Ansätze betrachtet. Am Ende wird das spezifische Profil des ASEI-10-Modells zusammengeführt und zur Herleitung der vier Bewertungsdimensionen übergeleitet.

3.2 Rechtliche und regulatorische Rahmenwerke

3.2.1 Funktion rechtlicher und regulatorischer Rahmenwerke

Rechtliche und regulatorische Rahmenwerke bilden einen zentralen Bezugspunkt für die Bewertung von KI-Systemen. Sie definieren, welche KI-Praktiken unzulässig sind, welche Systeme besonderen Anforderungen unterliegen, welche Akteure Verantwortung tragen und welche Pflichten im Lebenszyklus eines KI-Systems einzuhalten sind. Für Organisationen sind solche Rahmenwerke unverzichtbar, weil sie die rechtlichen Mindestanforderungen und Verantwortlichkeiten bestimmen.

ASEI-10 bewegt sich jedoch auf einer anderen methodischen Ebene. Das Modell entscheidet nicht, ob ein KI-System rechtlich zulässig ist, ob es als Hochrisiko-KI-System im Sinne eines bestimmten Gesetzes einzustufen ist oder ob alle regulatorischen Pflichten erfüllt wurden. Es beantwortet eine vorgelagerte und zugleich eigenständige Bewertungsfrage: Welches strukturelle Risikopotenzial besitzt ein konkretes KI-System in seinem konkreten Anwendungskontext?

Gerade dadurch kann ASEI-10 rechtliche und regulatorische Rahmenwerke sinnvoll ergänzen. Während Regulierung normativ festlegt, welche Anforderungen gelten, unterstützt ASEI-10 die operative Einordnung, Priorisierung und Dokumentation einzelner KI-Systeme innerhalb einer Organisation. Das Modell kann sichtbar machen, welche Systeme vertieft rechtlich, technisch oder organisatorisch geprüft werden sollten, ohne diese Prüfungen selbst zu ersetzen.

3.2.2 Verhältnis zum EU AI Act

Der EU AI Act[12] ist der wichtigste rechtliche Referenzrahmen für KI-Systeme in der Europäischen Union. Er verfolgt einen risikobasierten Regulierungsansatz und unterscheidet unter anderem zwischen verbotenen KI-Praktiken, Hochrisiko-KI-Systemen, bestimmten Transparenzpflichten und besonderen Anforderungen an General-Purpose-AI-Modelle. Für Hochrisiko-KI-Systeme sieht der AI Act umfangreiche Anforderungen vor, etwa an Risikomanagement, Datenqualität, technische Dokumentation, Protokollierung, Transparenz, menschliche Aufsicht, Genauigkeit, Robustheit und Cybersicherheit.

ASEI-10 knüpft an diese risikobezogene Grundlogik an, verfolgt aber einen anderen Zweck. Der EU AI Act ist ein verbindlicher Rechtsrahmen. Er beantwortet die Frage, welche rechtlichen Pflichten für bestimmte KI-Systeme, Anbieter, Betreiber und weitere Akteure gelten. ASEI-10 ist dagegen ein methodisches Klassifikationsmodell. Es beantwortet die Frage, wie das Risikopotenzial eines konkreten KI-Systems im Anwendungskontext anhand von Autonomie, Severität, Exposition und Irreversibilität nachvollziehbar eingeordnet werden kann.

Die Unterschiede zeigen sich besonders bei der Hochrisiko-Systematik. Der EU AI Act bestimmt Hochrisiko-KI-Systeme insbesondere über bestimmte Produktbezüge und über Anwendungsbereiche, die in Anhang III genannt werden. ASEI-10 arbeitet nicht mit einer rechtlichen Fallgruppenzuordnung, sondern mit einer vierdimensionalen Bewertung des konkreten Systems. Dadurch kann ein System im ASEI-10-Modell ein erhöhtes oder hohes Risikopotenzial aufweisen, auch wenn es nicht ohne Weiteres unter eine Hochrisiko-Kategorie des EU AI Act fällt. Umgekehrt ersetzt ein niedriger ASEI-10-Score keine rechtliche Prüfung, wenn ein System regulatorisch erfasst ist.

Besonders deutlich wird die Eigenständigkeit des ASEI-10-Modells bei agentischen Systemen. Der EU AI Act enthält zwar Regelungen zu KI-Systemen, Hochrisiko-Anwendungen und General-Purpose-AI-Modellen, aber er stellt keine eigenständige operative Skala bereit, mit der die Handlungstiefe eines konkreten KI-Systems von passiver Assistenz bis zu persistenten, verteilten oder schwer kontrollierbaren agentischen Systemen differenziert bewertet wird. Genau hier setzt die Autonomiedimension des ASEI-10-Modells an.

Auch die weiteren ASEI-Dimensionen ergänzen die Perspektive des AI Act. Severität erfasst die mögliche Schwere negativer Folgen im konkreten Einsatzbereich. Exposition beschreibt die Breite des Wirkungsraums. Irreversibilität bewertet, wie schwer mögliche Folgen nachträglich gestoppt, korrigiert oder kompensiert werden können. Diese Dimensionen können bei der Vorbereitung rechtlicher Prüfungen, bei der internen Priorisierung von KI-Systemen und bei der Ableitung organisatorischer Mindestanforderungen helfen.

ASEI-10 ist daher kein Ersatz für den EU AI Act und kein Instrument zur abschließenden Feststellung rechtlicher Konformität. Sein Beitrag liegt auf einer anderen Ebene: Das Modell macht das Risikopotenzial konkreter KI-Systeme vergleichbar, begründbar und dokumentierbar. Damit kann es Organisationen helfen, den eigenen KI-Bestand vorzustrukturieren, prüfbedürftige Systeme zu identifizieren und regulatorische Anforderungen gezielter in Governance-, Freigabe- und Kontrollprozesse einzubetten.

[12] Verordnung (EU) 2024/1689 (EU AI Act)

3.2.3 Verhältnis zu den International AI Safety Reports

Die International AI Safety Reports[13] unterscheiden sich grundlegend von rechtlich verbindlichen Rahmenwerken wie dem EU AI Act. Sie sind keine Verordnung, keine Norm und kein Managementsystem, sondern wissenschaftliche Synthesen zum Stand der Forschung über Fähigkeiten, Risiken und Risikomanagement fortgeschrittener KI-Systeme, insbesondere von General-Purpose AI. Ihr Schwerpunkt liegt nicht auf der rechtlichen Einordnung einzelner Systeme, sondern auf der Auswertung wissenschaftlicher Evidenz zu technologischen Entwicklungslinien, neuen Fähigkeiten, Missbrauchsmöglichkeiten, Kontrollproblemen, gesellschaftlichen Auswirkungen und offenen Fragen des Risikomanagements.

Für das ASEI-10-Modell sind die International AI Safety Reports aus mehreren Gründen bedeutsam. Erstens stützen sie die Grundannahme, dass KI-Risiken nicht allein aus der abstrakten Leistungsfähigkeit eines Modells entstehen. Risikorelevant wird ein KI-System vor allem durch das Zusammenspiel von Fähigkeiten, Systemintegration, Einsatzkontext, Autonomie, Zugriffsmöglichkeiten, Missbrauchspotenzial, Kontrollierbarkeit und möglicher Wirkung. Diese Perspektive entspricht dem Grundgedanken von ASEI-10, nicht das Modell isoliert, sondern das konkrete KI-System in seiner operativen Einbettung zu bewerten.

Zweitens verdeutlichen die Reports, dass das Risikomanagement fortgeschrittener KI-Systeme mit erheblichen Unsicherheiten verbunden bleibt. Evaluation, Monitoring, Red Teaming, technische Schutzmaßnahmen, Transparenzmechanismen und organisatorische Governance entwickeln sich weiter, können aber keine vollständige Sicherheitsgarantie liefern. Gerade deshalb ist ein vorgelagertes Klassifikationsmodell sinnvoll, das nicht erst auf empirisch gesicherte Schadenswahrscheinlichkeiten angewiesen ist, sondern das strukturelle Risikopotenzial eines Systems anhand seiner Merkmale und seines Anwendungskontextes einordnet.

Drittens zeigen die International AI Safety Reports die besondere Bedeutung agentischer und generalisierbarer KI-Systeme. Fortschritte bei Planung, Werkzeugnutzung, Problemlösung, Codegenerierung, Cyberfähigkeiten, multimodaler Verarbeitung und langfristiger Aufgabenbearbeitung können dazu führen, dass KI-Systeme stärker in reale Prozesse eingebettet werden. Damit rücken genau jene Fragen in den Vordergrund, die im ASEI-10-Modell durch die Autonomiedimension, die Expositionsdimension und die Irreversibilitätsdimension erfasst werden: Wie selbstständig handelt das System? Wie weit reicht sein Wirkungsraum? Wie gut lassen sich negative Folgen stoppen, korrigieren oder kompensieren?

ASEI-10 übernimmt aus den International AI Safety Reports keine einzelnen Risikoprognosen und keine spezifischen Szenarien. Das Modell leitet aus ihnen vielmehr eine methodische Vorsicht ab. Es unterstellt nicht, dass jedes fortgeschrittene KI-System zwangsläufig gefährlich ist. Es verharmlost aber auch nicht, dass neue Fähigkeiten, Toolzugriffe, agentische Handlungsmuster und systemische Einbettungen das Risikopotenzial eines Systems erheblich verändern können.

Damit besetzt ASEI-10 eine andere Ebene als die International AI Safety Reports. Die Reports liefern eine wissenschaftliche Evidenzbasis und beschreiben den Stand der Erkenntnis über fortgeschrittene KI. ASEI-10 übersetzt zentrale Risikotreiber in ein operatives Klassifikationsmodell für konkrete KI-Systeme. Während die Reports vor allem erklären, welche Entwicklungen und Risikofelder beobachtet werden müssen, hilft ASEI-10 dabei, einzelne Systeme innerhalb einer Organisation nach ihrem strukturellen Risikopotenzial einzuordnen, zu priorisieren und zu dokumentieren.

[13] Bengio et al. 2025; 2026

Das Verhältnis ist daher komplementär, aber nicht nachgeordnet. Die International AI Safety Reports liefern den wissenschaftlichen Hintergrund für die Relevanz von Fähigkeiten, Autonomie, Kontrollierbarkeit und systemischen Wirkungen. ASEI-10 macht diese Risikoperspektive für konkrete Anwendungskontexte operationalisierbar.

3.2.4 Institutionelle KI-Sicherheitsforschung und nationale Safety-Infrastrukturen

Neben internationalen Safety Reports gewinnen auch institutionelle Formen der KI-Sicherheitsforschung an Bedeutung. Das DLR-Institut für KI-Sicherheit[14] arbeitet unter anderem an Bewertungs- und Testmethoden für KI-basierte Komponenten und Systeme in sicherheitskritischen Anwendungen. Damit liegt der Schwerpunkt stärker auf technischer Sicherheit, Testbarkeit, Zuverlässigkeit und sicherheitskritischen Einsatzfeldern als auf einer allgemeinen organisatorischen Risikoklassifikation.

Auch die Debatte um nationale AI-Safety-Infrastrukturen ist für ASEI-10 anschlussfähig. Sie zeigt, dass die Bewertung fortgeschrittener KI-Systeme zunehmend als eigenständige institutionelle Aufgabe verstanden wird. Die Einrichtung eines deutschen KI-Sicherheitsinstituts unterstreicht diesen Trend: Fähigkeiten moderner KI-Modelle, ihre Risiken, internationale Standards und der Austausch mit vergleichbaren Einrichtungen sollen gebündelt betrachtet werden.[15]

ASEI-10 bewegt sich auf einer anderen Ebene. Das Modell ersetzt keine technische Sicherheitsforschung, keine staatliche Risikoanalyse und keine institutionelle Safety-Prüfung. Es kann jedoch als vorgelagertes Klassifikationsinstrument dienen, um in Organisationen sichtbar zu machen, welche konkreten KI-Systeme aufgrund von Autonomie, Severität, Exposition und Irreversibilität besonders prüf- und steuerungsbedürftig sind.

3.3 Governance-, Risiko- und Managementsystemansätze

3.3.1 Funktion von Governance-, Risiko- und Managementsystemansätzen

Governance-, Risiko- und Managementsystemansätze verfolgen einen anderen Zweck als rechtliche Regulierungsrahmen. Während rechtliche Rahmenwerke festlegen, welche Pflichten, Verbote und Zuständigkeiten gelten, beschreiben Governance- und Managementansätze vor allem, wie Organisationen KI-Systeme verantwortlich steuern, überwachen, dokumentieren und kontinuierlich verbessern können.

Zu diesen Ansätzen gehören insbesondere das NIST AI Risk Management Framework, ISO/IEC 23894 und ISO/IEC 42001. Sie stellen keine bloßen Checklisten dar, sondern bieten übergeordnete Strukturen für den Umgang mit KI-Risiken. Im Mittelpunkt stehen Fragen der Verantwortlichkeit, der organisatorischen Einbettung, der Risikoidentifikation, der Risikobehandlung, der Überwachung, der Dokumentation und der fortlaufenden Verbesserung.

ASEI-10 steht zu diesen Ansätzen nicht in einem Unterordnungsverhältnis. Das Modell übernimmt nicht die Rolle eines vollständigen Governance- oder Managementsystems, sondern besetzt eine eigenständige methodische Ebene innerhalb solcher Strukturen. Governance- und Managementsysteme beschreiben den organisatorischen Rahmen, in dem KI-Risiken gesteuert werden. ASEI-10 liefert eine konkrete Klassifikationslogik, mit der einzelne KI-Systeme innerhalb dieses Rahmens bewertet, verglichen und priorisiert werden können.

[14] DLR o.J.

[15] Presse- und Informationsamt der Bundesregierung 2026

Gerade hierin liegt der besondere Beitrag des Modells. Organisationen benötigen nicht nur allgemeine Prinzipien für verantwortliche KI, sondern auch ein praktikables Verfahren, um unterschiedliche KI-Systeme nach einheitlichen Kriterien einzuordnen. Ein internes Textassistenzsystem, ein Kundenservice-Chatbot, ein Recruiting-System, ein Unternehmensagent mit Toolzugriff und ein KI-System in kritischer Infrastruktur können nicht sinnvoll mit derselben pauschalen Risikobeschreibung behandelt werden. Sie müssen nach ihren tatsächlichen Risikotreibern unterschieden werden.

ASEI-10 stellt dafür vier Bewertungsdimensionen bereit: Autonomie, Severität, Exposition und Irreversibilität. Diese Dimensionen machen sichtbar, ob das Risikopotenzial eines Systems vor allem aus operativer Handlungsmacht, aus der Schwere möglicher Folgen, aus der Breite des Wirkungsraums oder aus der mangelnden Rückholbarkeit negativer Folgen entsteht. Der ASEI-10-Score verdichtet diese Einstufung, während die Einzeldimensionen und Prüfschwellen die fachliche Begründung erhalten.

Damit kann ASEI-10 in Governance-, Risiko- und Managementsystemansätze eingebettet werden, ohne in ihnen aufzugehen. Das Modell kann KI-Inventare strukturieren, Freigabeprozesse vorbereiten, Risikoklassen begründen, Prüfprioritäten setzen, Auditvorbereitungen unterstützen und Neubewertungen auslösen. Es ersetzt nicht die organisatorische Gesamtverantwortung einer KI-Governance, macht diese Verantwortung aber operativ greifbarer.

Die folgenden Abschnitte ordnen ASEI-10 deshalb zu drei besonders relevanten Referenzrahmen ein: dem NIST AI Risk Management Framework, ISO/IEC 23894 als Leitlinie für KI-Risikomanagement und ISO/IEC 42001 als Managementsystemstandard für künstliche Intelligenz.

3.3.2 Verhältnis zum NIST AI Risk Management Framework

Das NIST AI Risk Management Framework[16] ist einer der wichtigsten internationalen Referenzrahmen für den verantwortlichen Umgang mit KI-Risiken. Es wurde vom U.S. National Institute of Standards and Technology entwickelt und verfolgt das Ziel, Organisationen bei der Identifikation, Bewertung, Steuerung und Kommunikation von KI-Risiken zu unterstützen. Das Framework ist freiwillig angelegt und richtet sich nicht nur an eine bestimmte Branche oder einen bestimmten KI-Typ, sondern an Organisationen, die KI-Systeme entwickeln, einsetzen, beschaffen, überwachen oder verantworten.

Der Kern des NIST AI Risk Management Frameworks besteht aus vier Funktionen: Govern, Map, Measure und Manage. Govern beschreibt die übergreifende Einbettung von KI-Risikomanagement in Verantwortlichkeiten, Prozesse, Rollen, Kultur und Governance-Strukturen. Map dient dazu, den Kontext, die Akteure, die beabsichtigten Nutzungen, die möglichen Wirkungen und die Risikolandschaft eines KI-Systems zu erfassen. Measure zielt auf die Analyse, Bewertung und Beobachtung von KI-Risiken, etwa durch Tests, Metriken, Evaluationen, Dokumentation und Monitoring. Manage betrifft schließlich die Priorisierung, Behandlung und fortlaufende Steuerung identifizierter Risiken.

[16] NIST AI RMF 1.0

ASEI-10 ist mit dieser Grundlogik gut vereinbar, setzt aber an einer spezifischeren Stelle an. Das NIST AI Risk Management Framework beschreibt einen umfassenden Prozessrahmen für den Umgang mit KI-Risiken. ASEI-10 stellt dagegen ein fokussiertes Klassifikationsmodell bereit, mit dem das strukturelle Risikopotenzial einzelner KI-Systeme anhand von Autonomie, Severität, Exposition und Irreversibilität bewertet werden kann. Es ersetzt daher nicht den NIST-Prozess, sondern kann innerhalb dieses Prozesses als konkretes Bewertungsinstrument eingesetzt werden.

Besonders naheliegend ist eine Einbindung von ASEI-10 in die Funktionen Map, Measure und Manage. Im Rahmen von Map kann ASEI-10 helfen, den Anwendungskontext eines KI-Systems strukturiert zu erfassen: Welche Aufgaben übernimmt das System? Welche Daten verarbeitet es? Welche Nutzergruppen und Betroffenen sind einbezogen? Welche Berechtigungen, Werkzeuge und Schnittstellen besitzt es? Welche Folgen wären bei Fehlern, Missbrauch oder Kontrollproblemen denkbar? Diese Kontextbeschreibung bildet zugleich die Grundlage für die spätere ASEI-10-Bewertung.

Im Rahmen von Measure kann ASEI-10 als Klassifikationslogik dienen, um das Risikopotenzial eines Systems vergleichbar einzuordnen. Der ASEI-10-Score ist dabei keine technische Leistungsmetrik und kein empirischer Wahrscheinlichkeitswert. Er verdichtet vielmehr die strukturellen Risikotreiber eines Systems zu einem nachvollziehbaren Klassifikationswert. Die Einzeldimensionen und Prüfschwellen verhindern, dass die Bewertung auf eine bloße Zahl reduziert wird.

Im Rahmen von Manage kann ASEI-10 dazu beitragen, Prioritäten zu setzen und risikobasierte Mindestanforderungen abzuleiten. Systeme mit niedrigem Risikopotenzial benötigen andere Freigabe-, Kontroll- und Dokumentationsanforderungen als Systeme mit hoher Autonomie, hoher Severität, breiter Exposition oder schwer reversiblen Folgen. Damit unterstützt ASEI-10 die praktische Entscheidung, welche Systeme vertieft geprüft, stärker überwacht, enger begrenzt oder nur unter zusätzlichen Schutzmaßnahmen eingesetzt werden sollten.

Auch zur Funktion Govern besteht ein enger Bezug. ASEI-10 liefert keine vollständige KI-Governance-Struktur, kann aber ein standardisiertes Bewertungsverfahren innerhalb einer solchen Struktur bilden. Eine Organisation kann etwa festlegen, dass alle KI-Systeme vor ihrer Einführung nach ASEI-10 bewertet werden, dass bestimmte Scorebereiche besondere Freigaben auslösen und dass kritische Prüfschwellen zusätzliche Dokumentations-, Monitoring- oder Eskalationspflichten begründen. Damit wird das abstrakte Prinzip risikobasierter KI-Governance in ein konkretes internes Verfahren übersetzt.

Die Eigenständigkeit von ASEI-10 liegt somit nicht darin, das NIST AI Risk Management Framework zu ersetzen. Sie liegt darin, eine operative Bewertungslogik bereitzustellen, die innerhalb eines solchen Frameworks unmittelbar angewendet werden kann. NIST beschreibt den umfassenden Risikomanagementrahmen; ASEI-10 klassifiziert das strukturelle Risikopotenzial konkreter KI-Systeme. Beide Ansätze sind daher komplementär, aber funktional verschieden.

Im Ergebnis kann ASEI-10 als Brücke zwischen dem allgemeinen Prozessrahmen des NIST AI Risk Management Frameworks und der konkreten Bewertung einzelner KI-Systeme verstanden werden. Es macht sichtbar, welche Systeme aufgrund ihrer Autonomie, ihres Einsatzbereichs, ihrer Reichweite oder ihrer Rückholproblematik besondere Aufmerksamkeit erfordern, und unterstützt damit genau jene risikobasierte Priorisierung, die für wirksames KI-Risikomanagement erforderlich ist.

3.3.3 Verhältnis zu ISO/IEC 23894

ISO/IEC 23894[17] ist ein internationaler Leitlinienstandard für das Risikomanagement im Zusammenhang mit künstlicher Intelligenz. Der Standard richtet sich an Organisationen, die KI-basierte Produkte, Systeme oder Dienstleistungen entwickeln, bereitstellen oder nutzen. Sein Schwerpunkt liegt darauf, KI-bezogene Risiken systematisch zu identifizieren, zu analysieren, zu bewerten, zu behandeln, zu überwachen und in organisationale Aktivitäten und Funktionen einzubetten.

Damit unterscheidet sich ISO/IEC 23894 sowohl von einer rechtlichen Regulierung als auch von einem konkreten Scoringmodell. Der Standard definiert keine eigenständige Risikoklasse für ein einzelnes KI-System und berechnet keinen zusammenfassenden Risikowert. Vielmehr beschreibt er einen übergeordneten Risikomanagementrahmen, innerhalb dessen Organisationen eigene Verfahren, Kriterien, Maßnahmen und Verantwortlichkeiten ausgestalten können.

ASEI-10 kann an dieser Stelle eine eigenständige methodische Funktion übernehmen. Das Modell liefert eine konkrete Klassifikationslogik für die Bewertung einzelner KI-Systeme im Anwendungskontext. Während ISO/IEC 23894 beschreibt, dass KI-Risiken systematisch gemanagt werden müssen, beantwortet ASEI-10 die operative Frage, wie das strukturelle Risikopotenzial eines konkreten KI-Systems nachvollziehbar eingestuft werden kann.

Die Verbindung beider Ansätze liegt vor allem in der Operationalisierung. ISO/IEC 23894 stellt den allgemeinen Rahmen für KI-Risikomanagement bereit. ASEI-10 kann innerhalb dieses Rahmens als wiederholbare Bewertungsmethode eingesetzt werden. Die vier Dimensionen Autonomie, Severität, Exposition und Irreversibilität strukturieren die Analyse des Systems. Der Rohwert und der ASEI-10-Score unterstützen die Priorisierung. Die Prüfschwellen helfen, Mindestanforderungen an Kontrolle, Dokumentation, Freigabe, Monitoring oder Neubewertung auszulösen.

Besonders wichtig ist dabei der Kontextbezug. ISO/IEC 23894 betont die Integration des Risikomanagements in die konkreten Aktivitäten und Funktionen einer Organisation. ASEI-10 folgt derselben Grundlogik, indem es nicht ein KI-Modell abstrakt bewertet, sondern ein konkretes KI-System mit seinen Aufgaben, Berechtigungen, Datenzugriffen, Nutzergruppen, Kontrollmechanismen und möglichen Folgen. Damit kann dasselbe technische Modell je nach Anwendungskontext zu sehr unterschiedlichen ASEI-10-Einstufungen führen.

ASEI-10 ist damit kein Ersatz für ISO/IEC 23894. Es übernimmt nicht den gesamten Risikomanagementprozess, definiert keine vollständige Risikopolitik und ersetzt keine organisatorischen Verantwortlichkeiten. Sein Beitrag liegt enger und zugleich präziser: Es stellt eine praktikable Methode bereit, mit der Organisationen einzelne KI-Systeme innerhalb eines ISO/IEC-23894-orientierten Risikomanagements klassifizieren, vergleichen und priorisieren können.

In der praktischen Anwendung kann ASEI-10 insbesondere für KI-Inventare, Vorprüfungen, Freigabeentscheidungen, Risikoregister, Auditvorbereitungen und turnusmäßige Neubewertungen genutzt werden. Ein hoher Score oder das Erreichen bestimmter Prüfschwellen kann anzeigen, dass vertiefte Risikoanalysen, zusätzliche Schutzmaßnahmen oder strengere Governance-Anforderungen erforderlich sind. Ein niedriger Score kann dagegen begründen, dass ein System im beschriebenen Kontext nur begrenzte risikobezogene Aufmerksamkeit benötigt, ohne dass dadurch allgemeine Sorgfalts- oder Compliancepflichten entfallen.

[17] ISO/IEC 23894:2023

Das Verhältnis zwischen ISO/IEC 23894 und ASEI-10 ist daher komplementär, aber funktional verschieden. ISO/IEC 23894 beschreibt den übergeordneten Rahmen des KI-Risikomanagements. ASEI-10 konkretisiert innerhalb dieses Rahmens die Bewertung des strukturellen Risikopotenzials einzelner KI-Systeme.

3.3.4 Verhältnis zu ISO/IEC 42001

ISO/IEC 42001[18] ist ein internationaler Managementsystemstandard für künstliche Intelligenz. Er beschreibt Anforderungen an ein Artificial Intelligence Management System, also an ein organisationales System aus Richtlinien, Rollen, Verantwortlichkeiten, Prozessen, Dokumentation, Überwachung, Bewertung und kontinuierlicher Verbesserung. Der Standard richtet sich an Organisationen, die KI-basierte Produkte oder Dienstleistungen entwickeln, bereitstellen oder nutzen.

Damit unterscheidet sich ISO/IEC 42001 von einem einzelnen Bewertungsverfahren. Der Standard beschreibt nicht nur die Bewertung einzelner KI-Systeme, sondern die organisatorische Gesamtstruktur, in der KI verantwortet und gesteuert wird. Im Mittelpunkt stehen Fragen wie: Welche KI-Politik verfolgt eine Organisation? Welche Rollen und Verantwortlichkeiten bestehen? Wie werden KI-Systeme dokumentiert? Wie werden Risiken und Chancen behandelt? Wie werden Anforderungen überwacht, überprüft und fortlaufend verbessert?

ASEI-10 steht zu ISO/IEC 42001 in einem besonders engen organisatorischen Verhältnis. Das Modell ist selbst kein Managementsystem. Es definiert keine vollständige Aufbau- und Ablauforganisation für KI-Governance, keine Zertifizierungsanforderungen und keinen vollständigen Managementzyklus. Es kann jedoch innerhalb eines KI-Managementsystems eine zentrale methodische Funktion übernehmen: die einheitliche Klassifikation des strukturellen Risikopotenzials einzelner KI-Systeme.

Gerade Managementsysteme benötigen wiederholbare Kriterien, nach denen KI-Systeme erfasst, bewertet, priorisiert und kontrolliert werden können. Ohne eine solche Bewertungslogik besteht die Gefahr, dass KI-Inventare zwar formal geführt werden, die Risikoeinstufungen aber uneinheitlich, schwer vergleichbar oder stark von subjektiven Einzelurteilen abhängig bleiben. ASEI-10 kann diese Lücke schließen, indem es ein transparentes und dokumentierbares Verfahren für die Einstufung konkreter KI-Systeme bereitstellt.

In einem ISO/IEC-42001-orientierten Managementsystem könnte ASEI-10 insbesondere an mehreren Stellen eingesetzt werden. Bei der Erfassung des KI-Bestands kann das Modell helfen, Systeme nicht nur zu registrieren, sondern nach ihrem Risikopotenzial zu klassifizieren. In Freigabeprozessen kann der ASEI-10-Score anzeigen, ob eine einfache Fachfreigabe genügt oder ob zusätzliche rechtliche, technische oder organisatorische Prüfungen erforderlich sind. Im laufenden Betrieb können Scorebereiche und Prüfschwellen Monitoring-, Dokumentations- oder Eskalationspflichten auslösen. Bei wesentlichen Systemänderungen kann ASEI-10 als Grundlage für eine strukturierte Neubewertung dienen.

Die vier Dimensionen des ASEI-10-Modells lassen sich dabei gut in die Logik eines Managementsystems einbetten. Autonomie verweist auf die operative Handlungstiefe des Systems und damit auf Anforderungen an Kontrolle, Freigabe und menschliche Aufsicht. Severität verweist auf die Sensibilität und mögliche Folgenschwere des Einsatzbereichs. Exposition zeigt, wie breit Personen, Organisationseinheiten, externe Empfänger oder Systeme betroffen sein können. Irreversibilität macht sichtbar, wie wichtig Rückholbarkeit, Korrekturmechanismen, Beschwerdewege und Notfallverfahren sind.

[18] ISO/IEC 42001:2023

ASEI-10 kann damit als methodisches Bindeglied zwischen Managementsystem und konkretem KI-System dienen. ISO/IEC 42001 beschreibt die organisatorische Hülle, innerhalb der KI verantwortet werden soll. ASEI-10 liefert eine mögliche Bewertungslogik, mit der einzelne KI-Systeme innerhalb dieser Hülle klassifiziert und gesteuert werden können. Das Verhältnis ist daher nicht konkurrierend, sondern arbeitsteilig: ISO/IEC 42001 strukturiert die Organisation der KI-Governance; ASEI-10 operationalisiert die risikobezogene Einstufung einzelner KI-Systeme.

Die Eigenständigkeit des ASEI-10-Modells bleibt dabei erhalten. Es ist nicht lediglich ein Anhang zu ISO/IEC 42001, sondern ein eigenständiges Klassifikationsmodell, das auch außerhalb eines zertifizierten Managementsystems genutzt werden kann. Seine besondere Stärke innerhalb eines Managementsystems liegt darin, abstrakte Anforderungen an Kontrolle, Dokumentation, Risikobehandlung und kontinuierliche Verbesserung mit einer konkreten, nachvollziehbaren und wiederholbaren Systembewertung zu verbinden.

3.3.5 Prüf- und Qualitätsbewertungsansätze: KI-Prüfkatalog, MISSION KI und VDE SPEC 90012

Neben regulatorischen, normativen und allgemeinen Risikomanagementansätzen haben sich in Deutschland praxisorientierte Prüf- und Qualitätsbewertungsansätze für KI-Systeme herausgebildet. Besonders relevant sind der KI-Prüfkatalog des Fraunhofer IAIS[19], der MISSION KI Qualitätsstandard[20] sowie darauf bezogene Prüf- und Standardisierungsansätze, etwa im Umfeld von VDE[21], TÜV AI.Lab[22], AI Quality & Testing Hub und CertifAI.

Der KI-Prüfkatalog des Fraunhofer IAIS verfolgt das Ziel, abstrakte Anforderungen an vertrauenswürdige KI in prüfbare Kriterien und konkrete Bewertungsfragen zu überführen. Er betrachtet KI-Anwendungen nicht allein technisch, sondern entlang mehrerer Vertrauenswürdigkeitsdimensionen, darunter Fairness, Autonomie und Kontrolle, Transparenz, Zuverlässigkeit, Sicherheit und Datenschutz. Für ASEI-10 ist insbesondere die Verbindung von Autonomie und Kontrolle anschlussfähig. Während der KI-Prüfkatalog prüft, ob Anforderungen an eine vertrauenswürdige KI-Anwendung erfüllt werden, klassifiziert ASEI-10 das strukturelle Risikopotenzial eines konkreten KI-Systems im Anwendungskontext.

Der MISSION KI Qualitätsstandard verfolgt ebenfalls eine praxisbezogene Prüf- und Bewertungslogik. Er strukturiert die Qualitätsbewertung von KI-Systemen unter anderem über Use-Case-Beschreibung, Schutzbedarfsanalyse, Prüfkatalog, Testmethoden und Prüfbericht. Besonders relevant ist die Schutzbedarfsanalyse, weil sie Anforderungen an den jeweiligen Anwendungskontext bindet und danach fragt, welche Schäden entstehen könnten, wenn bestimmte Qualitätsanforderungen nicht erfüllt werden. Darin besteht eine deutliche Nähe zur Severitätslogik von ASEI-10: Auch hier wird die mögliche Schwere negativer Folgen nicht abstrakt, sondern kontextbezogen bestimmt.

Gleichzeitig unterscheiden sich diese Prüf- und Qualitätsbewertungsansätze in ihrer Zielrichtung vom ASEI-10-Modell. Der KI-Prüfkatalog und der MISSION KI Qualitätsstandard fragen vor allem danach, ob ein KI-System bestimmte Qualitäts-, Vertrauenswürdigkeits- oder Prüfanforderungen erfüllt und wie diese Anforderungen nachgewiesen werden können. ASEI-10 fragt demgegenüber vorgelagert und komprimierend, welches Risikopotenzial ein konkretes KI-System aufgrund seiner Autonomie, Severität, Exposition und Irreversibilität besitzt.

[19] Fraunhofer IAIS 2023
[20] MISSION KI 2025
[21] VDE 2025
[22] TÜV AI.Lab o.J.

Damit stehen diese Ansätze nicht in Konkurrenz zueinander. Prüf- und Qualitätsstandards liefern detaillierte Anforderungen, Nachweismethoden und Bewertungsverfahren. ASEI-10 liefert eine kompakte, dokumentierbare Risikopotenzialklassifikation. In einer organisatorischen KI-Governance können beide Ebenen sinnvoll verbunden werden: ASEI-10 identifiziert und priorisiert risikorelevante Systeme, während Prüf- und Qualitätsstandards die vertiefte Bewertung einzelner Anforderungen unterstützen.

3.4 Klassische Bewertungs- und Fehleranalyseansätze

3.4.1 Funktion klassischer Bewertungs- und Fehleranalyseansätze

Klassische Bewertungs- und Fehleranalyseansätze bilden einen wichtigen Bezugspunkt für das ASEI-10-Modell, weil sie zeigen, wie Risiken, Fehler, Auswirkungen und Prioritäten in technischen, organisatorischen und sicherheitsbezogenen Kontexten traditionell strukturiert werden. Dazu gehören insbesondere klassische Risikobewertungen mit Eintrittswahrscheinlichkeit und Schadensausmaß, Risikomatrizen sowie Fehlermöglichkeitsanalysen wie FMEA und FMECA.

Diese Ansätze haben einen hohen praktischen Nutzen. Sie unterstützen Organisationen dabei, mögliche Ereignisse, Fehlerarten, Ursachen, Auswirkungen, Eintrittswahrscheinlichkeiten, Entdeckbarkeit und Maßnahmen systematisch zu erfassen. In vielen Bereichen sind sie etabliert, weil dort Erfahrungswerte, Ausfallraten, statistische Daten, Prozesskenntnisse oder technische Prüfergebnisse vorliegen. Gerade in Qualitätsmanagement, Produktsicherheit, industrieller Zuverlässigkeit, Arbeitssicherheit und technischen Kontrollsystemen haben solche Verfahren eine starke methodische Tradition.

Für KI-Systeme sind diese Ansätze jedoch nur teilweise unmittelbar übertragbar. Viele KI-bezogene Risikokonstellationen entstehen nicht allein aus klar bestimmbaren Fehlerarten oder stabilen Eintrittswahrscheinlichkeiten. Sie ergeben sich aus dem Zusammenspiel von Modellverhalten, Systemeinbettung, Datenzugriffen, Toolnutzung, Nutzerverhalten, Prompting, organisatorischer Kontrolle, Missbrauchsmöglichkeiten und dynamischen Anwendungskontexten. Gerade bei neuen oder agentischen KI-Systemen liegen belastbare Häufigkeitsdaten oft noch nicht vor.

ASEI-10 grenzt sich deshalb von klassischen Bewertungs- und Fehleranalyseansätzen ab, ohne deren Nutzen in Frage zu stellen. Das Modell ersetzt keine FMEA, keine technische Zuverlässigkeitsanalyse und keine detaillierte Untersuchung konkreter Fehlermöglichkeiten. Es beantwortet eine andere Frage: Nicht welche einzelnen Fehler in einem System auftreten können, sondern welches strukturelle Risikopotenzial ein konkretes KI-System im Anwendungskontext besitzt.

Damit setzt ASEI-10 früher und breiter an als viele klassische Fehleranalyseverfahren. Es kann genutzt werden, um zu entscheiden, welche KI-Systeme überhaupt einer vertieften technischen, rechtlichen oder organisatorischen Analyse unterzogen werden sollten. Ein hoher ASEI-10-Score oder das Erreichen kritischer Prüfschwellen kann anzeigen, dass nachgelagerte Verfahren wie FMEA, Red Teaming, Robustheitstests, Datenschutzprüfung, Sicherheitsanalyse oder vertiefte Prozessanalyse erforderlich sind.

Die folgenden Abschnitte ordnen ASEI-10 deshalb zunächst gegenüber klassischer Risikobewertung und Risikomatrizen ein. Anschließend wird das Verhältnis zu FMEA, FMECA und verwandten Fehlermöglichkeitsanalysen beschrieben.

3.4.2 Klassische Risikobewertung und Risikomatrizen

Klassische Risikobewertungen[23] arbeiten häufig mit der Verbindung von Eintrittswahrscheinlichkeit und Schadensausmaß. Risiko wird dabei typischerweise aus der Frage abgeleitet, wie wahrscheinlich ein bestimmtes Ereignis eintritt und wie schwer dessen Folgen wären. In vielen technischen, industriellen, finanziellen oder sicherheitsbezogenen Kontexten ist diese Logik sachgerecht, weil Erfahrungswerte, Ausfallraten, Schadensdaten, Prozesskennzahlen oder statistische Annahmen vorliegen.

Auch Risikomatrizen greifen diese Grundlogik auf. Sie ordnen Eintrittswahrscheinlichkeit und Auswirkung in Kategorien ein und leiten daraus Risikoklassen oder Handlungsprioritäten ab. Eine typische Matrix unterscheidet beispielsweise zwischen niedriger, mittlerer und hoher Eintrittswahrscheinlichkeit einerseits sowie geringer, mittlerer und schwerer Auswirkung andererseits. Aus der Kombination beider Achsen entsteht eine qualitative oder halbquantitative Risikoeinstufung.

Für viele praktische Entscheidungen sind solche Matrizen nützlich, weil sie komplexe Risikolagen in einer einfach kommunizierbaren Form darstellen. Sie machen deutlich, dass ein geringfügiger Schaden auch bei hoher Wahrscheinlichkeit anders zu behandeln ist als ein sehr schwerer Schaden mit niedriger Eintrittswahrscheinlichkeit. Außerdem unterstützen sie Priorisierungen, indem sie Risiken sichtbar bündeln und unterschiedliche Handlungsdringlichkeiten markieren.

Für KI-Systeme ist diese Vorgehensweise jedoch nur eingeschränkt übertragbar. Viele KI-bezogene Risiken entstehen nicht aus stabilen, gut beobachtbaren Ausfallwahrscheinlichkeiten. Sie ergeben sich aus dynamischen Modelländerungen, unvorhersehbaren Nutzungskontexten, Prompting-Effekten, Datenabhängigkeiten, Toolzugriffen, Sicherheitslücken, Missbrauchsmöglichkeiten, organisatorischen Kontrolllücken und Wechselwirkungen zwischen Mensch, Organisation und System. Die Eintrittswahrscheinlichkeit konkreter Schadensereignisse ist deshalb häufig schwer belastbar zu bestimmen, insbesondere bei neuen, agentischen oder noch nicht breit eingesetzten Systemen.

Hinzu kommt ein messtheoretisches Problem. Risikomatrizen arbeiten häufig mit ordinalen Kategorien. Diese Kategorien bringen eine Rangordnung zum Ausdruck, bilden aber nicht zwingend gleiche Abstände oder echte Verhältnisgrößen ab. Wird mit solchen Stufen rechnerisch so umgegangen, als wären sie präzise metrische Werte, kann eine Scheingenauigkeit entstehen. Dieses Problem betrifft nicht nur klassische Risikomatrizen, sondern grundsätzlich alle Bewertungsmodelle, die qualitative Stufen in Zahlenwerte überführen.[24]

ASEI-10 übernimmt deshalb nicht den Anspruch einer klassischen Risikomessung. Das Modell berechnet keine Eintrittswahrscheinlichkeit und bestimmt kein Risiko im Sinne von Wahrscheinlichkeit mal Schadenshöhe. Es klassifiziert stattdessen das strukturelle Risikopotenzial eines konkreten KI-Systems in einem konkreten Anwendungskontext. Der ASEI-10-Score ist ein heuristischer Klassifikationsindex, kein empirisch bestimmter Erwartungswert eines Schadens.

[23] ISO 31000:2018; IEC 31010:2019

[24] Cox 2008

Der Unterschied lässt sich knapp formulieren: Klassische Risikoanalyse fragt, wie wahrscheinlich ein bestimmter Schaden ist und wie groß dieser Schaden wäre. ASEI-10 fragt, welches Risikopotenzial ein KI-System besitzt, wenn Fehler, Fehlverwendung, Missbrauch oder Kontrollprobleme auftreten. Bewertet werden dabei nicht Eintrittswahrscheinlichkeiten, sondern die risikotreibenden Merkmale Autonomie, Severität, Exposition und Irreversibilität.

Damit setzt ASEI-10 auf einer früheren und anders gelagerten Ebene an. Es kann auch dann angewendet werden, wenn noch keine belastbaren empirischen Wahrscheinlichkeiten vorliegen. Gerade in frühen Planungs-, Pilot-, Beschaffungs- oder Freigabephasen kann das Modell helfen, KI-Systeme zunächst nach ihrem strukturellen Risikopotenzial zu priorisieren. Für Systeme mit erhöhtem oder hohem Risikopotenzial können anschließend klassische Risikoanalysen, technische Tests, Red Teaming, FMEA, Datenschutzprüfungen, Sicherheitsanalysen oder Betriebsauswertungen folgen.

ASEI-10 ersetzt Risikomatrizen daher nicht vollständig. In Organisationen können Risikomatrizen weiterhin sinnvoll sein, insbesondere wenn konkrete Schadensereignisse, Eintrittswahrscheinlichkeiten und Maßnahmen bewertet werden sollen. ASEI-10 erfüllt jedoch eine eigenständige Funktion: Es strukturiert die vorgelagerte Klassifikation von KI-Systemen, bevor konkrete Schadenswahrscheinlichkeiten belastbar quantifiziert werden oder bevor vertiefte Einzelrisikoanalysen durchgeführt werden.

3.4.3 FMEA, FMECA und verwandte Fehlermöglichkeitsanalysen

FMEA und FMECA gehören zu den etablierten Verfahren der systematischen Fehleranalyse.[25] Die *Failure Mode and Effects Analysis* untersucht mögliche Fehlerarten, ihre Ursachen und ihre Auswirkungen auf ein Produkt, einen Prozess oder ein technisches System. Die *Failure Mode, Effects and Criticality Analysis* erweitert diese Betrachtung um eine zusätzliche Kritikalitätsbewertung. In der Praxis werden solche Verfahren insbesondere genutzt, um Schwachstellen frühzeitig zu erkennen, Fehlerfolgen zu priorisieren und geeignete Vermeidungs-, Entdeckungs- oder Minderungsmaßnahmen abzuleiten.

Der besondere Nutzen dieser Verfahren liegt in ihrer Detailtiefe. Eine FMEA zerlegt ein System, einen Prozess oder eine Funktion in einzelne Elemente und fragt systematisch, was an welcher Stelle fehlschlagen kann, wodurch der Fehler verursacht werden könnte, welche Folgen er hätte, wie wahrscheinlich er ist, wie gut er entdeckt werden kann und welche Maßnahmen erforderlich sind. Dadurch entsteht eine konkrete, handlungsorientierte Fehler- und Maßnahmenlogik.

Für KI-Systeme können FMEA, FMECA und verwandte Verfahren wertvoll sein, insbesondere wenn bestimmte Komponenten, Prozessschritte, Schnittstellen, Datenflüsse, Kontrollpunkte oder technische Funktionen vertieft untersucht werden sollen. Beispiele sind fehlerhafte Datenübernahmen, unzureichende Plausibilitätsprüfungen, falsche Weiterleitungen, fehlerhafte Toolausführungen, fehlende Eskalationen, Protokollierungslücken, Prompt-Injection-Schwachstellen oder unzureichende Rücksetzmechanismen.

Gleichwohl erfassen FMEA und FMECA nicht dieselbe Bewertungsfrage wie ASEI-10. Sie setzen typischerweise bei konkreten Fehlermöglichkeiten und ihren Ursachen an. ASEI-10 setzt dagegen bei der übergeordneten Systemeinordnung an. Es fragt nicht zunächst, welche einzelnen Fehlerarten auftreten können, sondern welches strukturelle Risikopotenzial ein KI-System aufgrund seiner Autonomie, seines Einsatzbereichs, seiner Exposition und der Rückholbarkeit möglicher Folgen besitzt.

[25] IEC 60812:2018

Diese unterschiedliche Blickrichtung ist wesentlich. Eine FMEA kann sehr detailliert untersuchen, wie ein bestimmter Fehler in einem KI-gestützten Kundenserviceprozess entsteht und welche Maßnahmen ihn verhindern könnten. ASEI-10 bewertet dagegen, ob ein solcher Kundenserviceprozess aufgrund seiner Daten, Nutzerreichweite, Handlungsmöglichkeiten und Korrekturmöglichkeiten überhaupt als niedrig, erhöht, hoch oder sehr hoch risikorelevant einzustufen ist. Das eine Verfahren analysiert Fehlerpfade, das andere klassifiziert das Risikopotenzial des Systems im Anwendungskontext.

Auch zeitlich können sich beide Ansätze unterscheiden. ASEI-10 kann bereits in einer frühen Planungs-, Beschaffungs- oder Freigabephase eingesetzt werden, wenn ein KI-System zunächst grundsätzlich eingeordnet werden soll. FMEA und FMECA werden besonders dann relevant, wenn ein System aufgrund seines Risikopotenzials vertieft analysiert und konkrete Schwachstellen systematisch bearbeitet werden müssen. Ein hoher ASEI-10-Score oder das Erreichen bestimmter Prüfschwellen kann daher ein Anlass sein, anschließend eine FMEA, FMECA oder vergleichbare Detailanalyse durchzuführen.

ASEI-10 ersetzt solche Verfahren nicht. Es entscheidet nicht im Einzelnen, welche Fehlerursachen bestehen, wie wahrscheinlich ein bestimmter Fehler ist oder welche konkrete Maßnahme technisch am wirksamsten wäre. Sein Beitrag liegt vorgelagert: Das Modell identifiziert, welche KI-Systeme aufgrund ihrer Struktur, Handlungstiefe, Reichweite und Folgenrückholbarkeit besondere Aufmerksamkeit erfordern. Auf dieser Grundlage kann entschieden werden, bei welchen Systemen sich der Aufwand einer vertieften Fehlermöglichkeitsanalyse besonders rechtfertigt.

Umgekehrt können Ergebnisse aus FMEA, FMECA oder verwandten Analysen in eine spätere ASEI-10-Neubewertung einfließen. Wenn technische oder organisatorische Maßnahmen nachweislich die Autonomie begrenzen, die Exposition reduzieren, Rückholmechanismen verbessern oder die möglichen Folgen begrenzen, kann sich auch die ASEI-10-Einstufung verändern. Voraussetzung ist jedoch, dass diese Maßnahmen tatsächlich wirksam implementiert, überprüfbar und dauerhaft sind.

Das Verhältnis zwischen ASEI-10 und FMEA/FMECA ist daher arbeitsteilig. ASEI-10 klassifiziert das strukturelle Risikopotenzial eines KI-Systems im Anwendungskontext. FMEA, FMECA und verwandte Verfahren analysieren konkrete Fehlermöglichkeiten, Ursachen, Wirkungen und Maßnahmen. Beide Perspektiven können sich ergänzen, sollten aber nicht miteinander verwechselt werden.

3.5 KI-spezifische Risiko- und Taxonomieansätze

3.5.1 Funktion KI-spezifischer Risiko- und Taxonomieansätze

Neben rechtlichen Rahmenwerken, Managementsystemansätzen und klassischen Fehleranalyseverfahren haben sich in den letzten Jahren zahlreiche KI-spezifische Risiko- und Taxonomieansätze entwickelt. Sie reagieren auf die besondere Breite, Dynamik und Vielschichtigkeit von KI-Risiken. Anders als klassische Risikomethoden gehen sie nicht primär von bekannten technischen Fehlerarten oder stabilen Eintrittswahrscheinlichkeiten aus, sondern sammeln, ordnen und systematisieren unterschiedliche Risikofelder künstlicher Intelligenz.

Solche Ansätze erfüllen eine wichtige Orientierungsfunktion. Sie machen sichtbar, welche Risikoarten im Zusammenhang mit KI-Systemen diskutiert werden, wie diese Risiken voneinander abgegrenzt werden können und welche Überschneidungen zwischen technischen, rechtlichen, gesellschaftlichen, ethischen, sicherheitsbezogenen und organisatorischen Perspektiven bestehen. Dazu gehören etwa Risiken im Zusammenhang mit Diskriminierung, Datenschutz, Desinformation, Manipulation, Sicherheitslücken, mangelnder Robustheit, Kontrollverlust, Machtkonzentration, Intransparenz, Abhängigkeit, Fehlanwendung oder systemischen Wirkungen.

Der Nutzen solcher Taxonomien liegt vor allem in der begrifflichen und analytischen Ordnung. Sie schaffen eine gemeinsame Sprache für Forschung, Regulierung, Governance, Sicherheitsbewertung und praktische Anwendung. Gerade weil der KI-Risikodiskurs stark fragmentiert ist, helfen Risk Repositories und Taxonomien dabei, Einzelrisiken zu bündeln, Lücken sichtbar zu machen und unterschiedliche Bewertungslogiken vergleichbar zu machen.

ASEI-10 steht zu diesen Ansätzen in einem engen, aber eigenständigen Verhältnis. Das Modell ist selbst keine umfassende Risikotaxonomie. Es versucht nicht, möglichst viele einzelne KI-Risiken zu sammeln oder alle denkbaren Schadensarten vollständig zu katalogisieren. Sein Zweck ist ein anderer: ASEI-10 klassifiziert das strukturelle Risikopotenzial eines konkreten KI-Systems in einem konkreten Anwendungskontext.

Damit setzt ASEI-10 eine andere Frage in den Mittelpunkt. KI-spezifische Risikotaxonomien fragen vor allem: Welche Arten von KI-Risiken gibt es? ASEI-10 fragt: Wie hoch ist das Risikopotenzial eines bestimmten KI-Systems, wenn seine Autonomie, sein Einsatzbereich, seine Reichweite und die Rückholbarkeit möglicher Folgen berücksichtigt werden?

Die Verbindung beider Perspektiven ist dennoch wichtig. Risikotaxonomien können helfen, die möglichen Folgen eines KI-Systems differenzierter zu erkennen und die Einstufung der ASEI-Dimensionen fachlich zu begründen. Wenn ein System etwa Diskriminierungsrisiken, Datenschutzrisiken, Sicherheitsrisiken, Desinformationsrisiken oder Kontrollrisiken berührt, kann dies insbesondere die Bewertung von Severität, Exposition und Irreversibilität beeinflussen. ASEI-10 überführt solche Risikoarten jedoch nicht in eine bloße Liste, sondern in eine strukturierte Systembewertung.

Der spezifische Beitrag von ASEI-10 liegt daher nicht in einer umfassenderen Sammlung von KI-Risiken. Sein Beitrag liegt in der Operationalisierung: Aus einzelnen Risikofeldern, Systemmerkmalen und Anwendungskontexten wird eine dokumentierbare, vergleichbare und schwellenbasierte Klassifikation des Risikopotenzials abgeleitet. Damit verbindet ASEI-10 die Breite KI-spezifischer Risikotaxonomien mit einer konkreten Bewertungslogik für Organisationen.

3.5.2 AI-Risk-Taxonomien und Risk Repositories

AI-Risk-Taxonomien und Risk Repositories verfolgen das Ziel, Risiken künstlicher Intelligenz systematisch zu erfassen, zu ordnen und vergleichbar zu machen. Sie reagieren auf ein zentrales Problem des KI-Risikodiskurses: Risiken werden in Forschung, Regulierung, Unternehmenspraxis, Sicherheitsbewertung und öffentlicher Debatte häufig unterschiedlich benannt, unterschiedlich abgegrenzt und auf unterschiedlichen Abstraktionsebenen beschrieben. Dadurch entsteht eine fragmentierte Risikolandschaft.

Ein besonders relevanter Ansatz ist die im MIT AI Risk Repository dokumentierte Meta-Übersicht.[26] Sie führt bestehende Risikoframeworks zusammen und ordnet eine große Zahl identifizierter KI-Risiken in übergreifende Taxonomien ein. Der Nutzen einer solchen Systematik liegt in ihrer Breite und Struktur. Sie hilft, unterschiedliche Risikobegriffe, Risikokategorien und Beschreibungslogiken zusammenzuführen, Überschneidungen sichtbar zu machen und Lücken im bestehenden Diskurs zu erkennen.

Ein weiterer einschlägiger Ansatz ist *AI Risk Categorization Decoded* (AIR 2024).[27] Dieser Ansatz zielt darauf, Risikokategorien aus staatlichen Regulierungen und unternehmensbezogenen KI-Richtlinien systematisch zusammenzuführen. Dadurch werden Perspektiven aus öffentlicher Regulierung und privater Governance vergleichbar gemacht. Solche Ansätze sind besonders nützlich, wenn Organisationen verstehen wollen, welche Risikoarten in unterschiedlichen Rahmenwerken wiederkehren und wie diese terminologisch geordnet werden können.

ASEI-10 steht zu diesen Taxonomien in einem engen, aber funktional anderen Verhältnis. Risk Repositories und Taxonomien beantworten vor allem die Frage: Welche Arten von KI-Risiken gibt es? ASEI-10 beantwortet eine andere Frage: Wie ist das strukturelle Risikopotenzial eines konkreten KI-Systems in einem konkreten Anwendungskontext einzustufen? Eine Taxonomie katalogisiert Risikoarten; ASEI-10 klassifiziert Systemkonstellationen.

Diese Unterscheidung ist wesentlich. Eine Risikotaxonomie kann beispielsweise Diskriminierung, Datenschutzverletzungen, Desinformation, Sicherheitslücken, Kontrollverlust, Intransparenz, Fehlanwendung oder gesellschaftliche Folgewirkungen als Risikoarten benennen. Sie sagt damit aber noch nicht, wie hoch das Risikopotenzial eines konkreten KI-Systems in einer bestimmten Organisation, mit bestimmten Berechtigungen, bestimmten Nutzergruppen und bestimmten Kontrollmechanismen einzustufen ist.

Genau hier setzt ASEI-10 an. Das Modell überführt einzelne Risikohinweise nicht in eine längere Risikoliste, sondern in eine vierdimensionale Systembewertung. Die in Taxonomien beschriebenen Risikoarten können die fachliche Begründung der ASEI-Dimensionen unterstützen. Diskriminierungs- oder Beschäftigungsrisiken können etwa die Severität erhöhen. Öffentlich zugängliche oder massenhaft genutzte Systeme können die Exposition erhöhen. Datenschutzverletzungen, Reputationsschäden oder irreversible Benachteiligungen können die Irreversibilität beeinflussen. Toolzugriffe, agentische Handlungsmuster und fehlende Freigabepunkte können die Autonomie betreffen.

Der spezifische Beitrag von ASEI-10 liegt daher nicht darin, mehr Risikoarten zu sammeln als bestehende Repositories. Das Modell beansprucht keine vollständigere Risikoenzyklopädie. Seine Eigenständigkeit liegt in der Operationalisierung: Aus Risikoarten, Systemmerkmalen und Anwendungskontexten wird eine dokumentierbare, vergleichbare und schwellenbasierte Klassifikation des Risikopotenzials abgeleitet.

Damit können AI-Risk-Taxonomien und ASEI-10 sinnvoll zusammenwirken. Taxonomien helfen, mögliche Risikoarten zu identifizieren und sprachlich präzise zu beschreiben. ASEI-10 hilft, diese Risikoarten im konkreten Systemkontext zu gewichten, den relevanten Bewertungsdimensionen zuzuordnen und zu einer nachvollziehbaren Einstufung zusammenzuführen. Beide Ansätze konkurrieren daher nicht miteinander. Taxonomien schaffen die Risikosprache; ASEI-10 stellt eine operative Bewertungslogik bereit.

[26] Slattery et al. 2024

[27] Zeng et al. 2024

3.5.3 Taxonomien systemischer Risiken aus General-Purpose AI

Ein besonderer Teilbereich KI-spezifischer Risikotaxonomien befasst sich mit systemischen Risiken aus General-Purpose AI; eine einschlägige Systematik dieses Feldes ist die Taxonomie systemischer Risiken aus General-Purpose AI von Uuk et al.[28] Im Unterschied zu anwendungsbezogenen Risikotaxonomien stehen hier nicht einzelne Nutzungssituationen, Produkte oder Organisationsprozesse im Vordergrund, sondern großskalige Wirkungen allgemein einsetzbarer KI-Systeme. Solche Risiken können Gesellschaft, Wirtschaft, Demokratie, Sicherheit, öffentliche Kommunikation, wissenschaftliche Integrität, kritische Infrastrukturen oder globale Stabilität betreffen.

Taxonomien systemischer Risiken aus General-Purpose AI machen sichtbar, dass bestimmte KI-Risiken nicht auf einen einzelnen Fehler, eine einzelne Organisation oder eine einzelne betroffene Person begrenzt bleiben. Sie können sich über Plattformen, Märkte, Informationsräume, Lieferketten, Sicherheitsarchitekturen oder gesellschaftliche Entscheidungsprozesse ausbreiten. Zu den diskutierten Risikofeldern gehören insbesondere Desinformation, Manipulation öffentlicher Meinungsbildung, Cyber- und Sicherheitsrisiken, Machtkonzentration, strukturelle Abhängigkeiten, Diskriminierung, Kontrollprobleme, Fehlanreize, Governance-Lücken und schwer vorhersehbare Wechselwirkungen zwischen KI-Systemen und gesellschaftlichen Strukturen.

Der Wert solcher Taxonomien liegt darin, systemische Wirkungszusammenhänge sichtbar zu machen, die in rein organisationsbezogenen oder produktbezogenen Bewertungsverfahren leicht unterschätzt werden können. Ein einzelnes KI-System mag isoliert betrachtet beherrschbar erscheinen. Wird es jedoch breit eingesetzt, mit anderen Systemen verbunden, in kritische Entscheidungsprozesse integriert oder durch viele Akteure gleichzeitig genutzt, kann sich sein Risikopotenzial erheblich verändern. Systemische Risiken entstehen daher häufig nicht allein aus einer einzelnen Funktion, sondern aus Skalierung, Vernetzung, Abhängigkeit, Wiederholung und schwer kontrollierbarer Ausbreitung.

ASEI-10 ist mit diesem Forschungsstrang kompatibel, aber anders zugeschnitten. Das Modell bewertet General-Purpose AI nicht abstrakt als Technologieklasse. Es klassifiziert auch kein Basismodell allein aufgrund seiner allgemeinen Leistungsfähigkeit. Bewertet wird erst das konkrete KI-System im konkreten Anwendungskontext: mit seinen Funktionen, Berechtigungen, Toolzugriffen, Nutzergruppen, Einsatzgrenzen, Kontrollmechanismen und möglichen Folgen. Dadurch kann dasselbe General-Purpose-AI-Modell je nach Einbettung sehr unterschiedlich eingestuft werden.

Gleichwohl nimmt ASEI-10 systemische Risiken ausdrücklich auf. Dies geschieht vor allem über die Dimensionen Severität und Exposition. Systemkritische Einsatzbereiche wie kritische Infrastruktur, Cybersecurity, öffentliche Sicherheit, biologische oder chemische Prozesse, militärische Anwendungen oder vergleichbare Risikokontexte können eine hohe Severität begründen. Eine organisationsübergreifende, öffentliche oder infrastrukturelle Wirkung kann zu hoher oder systemischer Exposition führen. Wenn negative Folgen zudem schwer zu stoppen, zu korrigieren oder zu kompensieren sind, wird auch die Irreversibilität zu einem wesentlichen Risikotreiber.

[28] Uuk et al. 2024

Besonders relevant ist dabei die Autonomiedimension. Systemische Risiken aus General-Purpose AI verschärfen sich, wenn KI-Systeme nicht nur Inhalte erzeugen, sondern Werkzeuge nutzen, Ziele verfolgen, Aufgaben über Zeit bearbeiten, mit anderen Systemen interagieren oder Handlungsketten eigenständig fortsetzen. In solchen Konstellationen reicht es nicht aus, allein nach dem Einsatzbereich zu fragen. Zu prüfen ist auch, welche operative Handlungstiefe das System besitzt und wie zuverlässig es begrenzt, überwacht oder gestoppt werden kann.

ASEI-10 übersetzt systemische Risikoperspektiven damit in eine anwendungsbezogene Klassifikationslogik. Eine Taxonomie systemischer Risiken beschreibt, welche großskaligen Risikofelder aus General-Purpose AI entstehen können. ASEI-10 fragt ergänzend, ob und in welchem Maße ein konkretes System solche Risikofelder berührt. Die Bewertung erfolgt nicht pauschal, sondern anhand der vier Dimensionen Autonomie, Severität, Exposition und Irreversibilität.

Damit unterscheidet sich ASEI-10 sowohl von einer reinen GPAI-Taxonomie als auch von einer abstrakten Debatte über fortgeschrittene KI. Das Modell vermeidet eine pauschale Gleichsetzung von General-Purpose AI mit hohem Risiko. Es verhindert aber auch eine Verharmlosung, wenn ein allgemein einsetzbares Modell durch Toolzugriffe, agentische Fähigkeiten, breite Exposition oder kritische Einsatzbereiche in ein deutlich risikorelevanteres System überführt wird.

Das Verhältnis ist daher arbeitsteilig. Taxonomien systemischer Risiken aus General-Purpose AI liefern eine wichtige Makroperspektive auf großskalige Gefährdungsfelder. ASEI-10 macht diese Perspektive für die Bewertung einzelner KI-Systeme nutzbar, indem es systemische Reichweite, operative Autonomie, kritische Einsatzbereiche und Rückholbarkeit in eine konkrete Risikopotenzialklassifikation überführt.

3.5.4 Qualitäts-, Robustheits- und Lebenszyklusansätze

Qualitäts-, Robustheits- und Lebenszyklusansätze bilden einen weiteren wichtigen Bezugspunkt für die Einordnung des ASEI-10-Modells. Sie betrachten KI-Systeme nicht nur als einzelne Anwendungen, sondern als technische und organisatorische Systeme, die über ihren gesamten Lebenszyklus hinweg geplant, entwickelt, getestet, eingeführt, überwacht, angepasst und gegebenenfalls außer Betrieb genommen werden müssen.

Zu diesen Ansätzen gehören unter anderem Standards, Spezifikationen und Modelle, die Begriffe, Referenzarchitekturen, Qualitätsanforderungen, Risikomanagementprozesse, Managementsysteme oder lebenszyklusbezogene Anforderungen an KI-Systeme beschreiben. Ihr Schwerpunkt liegt häufig auf Fragen der Datenqualität, Modellleistung, Robustheit, Nachvollziehbarkeit, Dokumentation, Überwachung, Veränderungskontrolle, Verantwortlichkeit und kontinuierlichen Verbesserung.

Solche Ansätze sind für die Praxis unverzichtbar, weil KI-Governance nicht allein aus einer einmaligen Risikoeinstufung besteht. Organisationen müssen festlegen, wie KI-Systeme ausgewählt, entwickelt, beschafft, getestet, freigegeben, eingesetzt, überwacht, verändert und außer Betrieb genommen werden. Dabei stellen sich zahlreiche Qualitäts- und Kontrollfragen: Sind die verwendeten Daten geeignet? Ist das System hinreichend robust? Werden Leistungsänderungen erkannt? Sind Eingriffe nachvollziehbar? Werden Vorfälle dokumentiert? Gibt es Verfahren für Korrektur, Rücknahme oder Eskalation?

ASEI-10 beantwortet diese Fragen nicht vollständig. Das Modell definiert keine umfassenden Qualitätsmetriken, keine technischen Robustheitstests, keinen vollständigen Lebenszyklusprozess und keine abschließenden Anforderungen an ein KI-Managementsystem. Sein Anspruch ist enger: ASEI-10 klassifiziert das strukturelle Risikopotenzial eines konkreten KI-Systems in seinem Anwendungskontext.

Gerade diese Begrenzung macht das Modell innerhalb von Qualitäts-, Robustheits- und Lebenszyklusansätzen praktisch nutzbar. Eine Organisation kann ASEI-10 einsetzen, um zu entscheiden, welche KI-Systeme besonders strenge Qualitätsanforderungen, Robustheitstests, Dokumentationspflichten, Monitoringverfahren oder Freigabestufen benötigen. Ein niedrig eingestuftes System erfordert nicht denselben Prüf- und Kontrollaufwand wie ein System mit hoher Autonomie, hoher Severität, breiter Exposition oder schwer reversiblen Folgen.

Die vier ASEI-Dimensionen können dabei als risikobezogene Steuerungsgrößen dienen. Autonomie zeigt, wie stark Kontroll-, Freigabe- und Begrenzungsmechanismen erforderlich sind. Severität weist darauf hin, wie kritisch Qualität, Validierung und fachliche Prüfung im jeweiligen Einsatzbereich sind. Exposition macht sichtbar, wie stark Skalierung, Rollout, Nutzerkreis und Außenwirkung überwacht werden müssen. Irreversibilität verweist auf Anforderungen an Rückholbarkeit, Korrektur, Beschwerdeverfahren, Notfallmaßnahmen und geordnete Außerbetriebnahme.

Damit kann ASEI-10 als Bindeglied zwischen Systemklassifikation und Lebenszyklussteuerung verstanden werden. Das Modell legt nicht selbst fest, welche technischen Tests oder organisatorischen Maßnahmen im Einzelnen ausreichen. Es kann aber begründen, warum bestimmte Systeme intensiver geprüft, enger überwacht, häufiger neu bewertet oder strenger dokumentiert werden müssen als andere.

Auch bei Systemänderungen ist dieser Zusammenhang bedeutsam. Wenn ein KI-System neue Datenquellen erhält, zusätzliche Nutzergruppen erreicht, externe Werkzeuge nutzt, automatisch Handlungen auslösen kann oder in einen sensibleren Einsatzbereich verschoben wird, verändert sich nicht nur sein technischer Lebenszyklusstatus. Auch sein ASEI-10-Risikopotenzial kann sich verändern. Qualitäts- und Lebenszyklusansätze liefern hierfür den organisatorischen Veränderungsprozess; ASEI-10 liefert die strukturierte Neubewertung des Risikopotenzials.

Das Verhältnis ist daher komplementär. Qualitäts-, Robustheits- und Lebenszyklusansätze beschreiben, wie KI-Systeme über Zeit zuverlässig, kontrolliert und verantwortbar gestaltet und betrieben werden können. ASEI-10 beantwortet innerhalb dieser Ansätze die Frage, welche Risikointensität ein konkretes System besitzt und welche Prüf- und Steuerungsintensität daraus abgeleitet werden sollte.

3.6 Scoring-Ansätze und Bewertungsmodelle

3.6.1 Funktion von Scoring-Ansätzen und Bewertungsmodellen

Neben rechtlichen Rahmenwerken, Governance-Ansätzen, Risikotaxonomien und Fehleranalyseverfahren gibt es verschiedene Versuche, KI-Risiken durch Scoringmodelle, Reifegradmodelle, Checklisten, Bewertungsmatrizen oder Punktesysteme zu operationalisieren. Solche Ansätze verfolgen das Ziel, komplexe Risikosachverhalte in handhabbare Bewertungsstrukturen zu übersetzen. Sie sollen Organisationen dabei unterstützen, KI-Systeme zu inventarisieren, Risikoprofile zu vergleichen, Prüfaufwand zu priorisieren und Entscheidungen nachvollziehbarer zu dokumentieren.

Der praktische Nutzen solcher Modelle liegt in ihrer Vereinfachungsleistung. Komplexe technische, organisatorische, rechtliche und ethische Fragen werden in Kriterien, Kategorien oder Punktwerte überführt. Dadurch können Bewertungen standardisiert, wiederholt und innerhalb einer Organisation besser kommuniziert werden. Gerade für KI-Inventare, Freigabeprozesse, Auditvorbereitungen oder interne Governance-Entscheidungen ist diese Übersetzungsleistung bedeutsam.

Gleichzeitig sind Scoring-Ansätze methodisch anspruchsvoll. Sie können eine Scheingenauigkeit erzeugen, wenn qualitative Einschätzungen in Zahlenwerte überführt werden, ohne den Status dieser Zahlenwerte offenzulegen. Besonders problematisch wird dies, wenn ordinale Kategorien wie niedrig, mittel oder hoch rechnerisch so behandelt werden, als handele es sich um präzise metrische Größen.[29] Ein numerischer Score kann dann den Eindruck objektiver Messgenauigkeit erzeugen, obwohl er tatsächlich auf fachlichen Einschätzungen, Modellannahmen und normativen Gewichtungen beruht.

3.6.2 Methodische Abgrenzung des ASEI-10-Scores

ASEI-10 nimmt die methodischen Grenzen von Scoringmodellen ausdrücklich auf. Das Modell nutzt zwar einen numerischen Score, versteht diesen aber nicht als exakte Risikomessung. Der ASEI-10-Score ist ein heuristischer Klassifikationsindex. Er dient der vergleichenden Einordnung, Priorisierung und Dokumentation, nicht der präzisen Berechnung eines objektiven Schadenerwartungswertes. Seine Aussagekraft entsteht nicht allein aus der Zahl, sondern aus dem Zusammenspiel von Score, Einzeldimensionen, Prüfschwellen und fachlicher Begründung.

Damit unterscheidet sich ASEI-10 von Scoring-Ansätzen, die numerische Ergebnisse liefern, ohne ihre methodischen Voraussetzungen ausreichend offenzulegen. Das Modell macht transparent, welche Dimensionen bewertet werden, wie die qualitative Autonomiestufe in den normierten Autonomiewert überführt wird, wie der multiplikative Rohwert entsteht und warum dieser logarithmisch auf eine Skala von 1 bis 10 normiert wird. Die Rechenlogik ist damit nicht verborgen, sondern Bestandteil des Modells.

Ein weiterer Unterschied liegt in der Verbindung von qualitativer und quantitativer Bewertung. ASEI-10 reduziert das Risikopotenzial eines KI-Systems nicht auf eine Punktzahl. Die vier Dimensionen Autonomie, Severität, Exposition und Irreversibilität müssen jeweils eigenständig begründet werden. Der Score verdichtet diese Einstufungen, ersetzt sie aber nicht. Gerade dadurch bleibt sichtbar, ob ein Risikopotenzial vor allem aus hoher Handlungstiefe, einem sensiblen Einsatzbereich, breiter Exposition oder schwer reversiblen Folgen entsteht.

Die ergänzenden Prüfschwellen sind in diesem Zusammenhang besonders wichtig. Sie verhindern, dass kritische Einzelmerkmale durch einen moderaten Gesamtscore verdeckt werden. Ein KI-System kann besondere Anforderungen auslösen, weil es agentisch handelt, in einem hochkritischen Bereich eingesetzt wird, systemisch exponiert ist oder schwer rückholbare Folgen erzeugen kann. Der Score bleibt damit ein wichtiges Orientierungsinstrument, aber nicht das alleinige Entscheidungskriterium.

ASEI-10 ist daher selbst ein Scoringmodell, aber kein bloßes Punktesystem. Es verbindet qualitative Skalenbeschreibung, normierte Rechenlogik, logarithmische Indexbildung, Prüfschwellen und Begründungspflichten zu einer strukturierten Klassifikationsmethode. Sein Anspruch liegt nicht darin, KI-Risiken scheinbar exakt zu messen, sondern darin, das strukturelle Risikopotenzial konkreter KI-Systeme transparent, vergleichbar und prüfbar einzuordnen.

[29] Cox 2008

Das Verhältnis zu anderen Scoring- und Bewertungsmodellen ist somit zweifach. Einerseits teilt ASEI-10 deren praktisches Ziel, komplexe Risikoeinschätzungen handhabbar und vergleichbar zu machen. Andererseits grenzt es sich von unzureichend reflektierten Punktesystemen ab, indem es die Grenzen ordinaler Skalen, die heuristische Funktion des Scores und die Notwendigkeit qualitativer Begründung ausdrücklich offenlegt.

3.7 Vergleichende Übersicht der Rahmenwerke und Bewertungsansätze

Die folgende Übersicht fasst die zuvor dargestellten Rahmenwerke und Bewertungsansätze vergleichend zusammen. Sie zeigt, dass die verschiedenen Ansätze unterschiedliche Funktionen erfüllen und daher nicht als austauschbare Alternativen zu verstehen sind. Einige Ansätze definieren rechtliche Pflichten, andere strukturieren Governance- und Managementprozesse, ordnen Risikoarten, analysieren konkrete Fehlermöglichkeiten oder beschreiben Qualitäts- und Lebenszyklusanforderungen. ASEI-10 nimmt innerhalb dieses Feldes eine eigene methodische Position ein: Es klassifiziert das strukturelle Risikopotenzial konkreter KI-Systeme im Anwendungskontext.

Rahmenwerk / Ansatz	Primäre Funktion	Charakter	Zentrale Risikologik	Verhältnis zu ASEI-10
EU AI Act	Rechtliche Regulierung von KI-Systemen und General-Purpose-AI-Modellen	Verbindliche EU-Verordnung	Risikobasierte rechtliche Kategorien, Verbote, Hochrisiko-Pflichten, Transparenzpflichten und GPAI-Anforderungen	ASEI-10 kann Vorprüfung, Priorisierung und Dokumentation unterstützen, ersetzt aber keine Rechtsprüfung
International AI Safety Reports	Wissenschaftliche Synthese zu Fähigkeiten, Risiken und Risikomanagement fortgeschrittener KI-Systeme	Wissenschaftlicher Bericht / Evidenzsynthese	Analyse von Fähigkeiten, Missbrauchspotenzialen, Kontrollproblemen, systemischen Risiken und offenen Fragen des Risikomanagements	ASEI-10 operationalisiert zentrale Risikotreiber für konkrete Anwendungskontexte, übernimmt aber keine einzelnen Risikoprognosen
NIST AI Risk Management Framework	Organisationsbezogenes KI-Risikomanagement	Freiwilliges Rahmenwerk	Govern, Map, Measure und Manage; Vertrauenswürdigkeit und Risikosteuerung über den KI-Lebenszyklus	ASEI-10 kann insbesondere Map und Measure operationalisieren und Ergebnisse in Govern und Manage einspeisen
NIST Generative AI Profile	Ergänzung des NIST AI RMF für generative KI	Freiwilliges Profil	Generative-KI-spezifische Risiken und Maßnahmen über den Lebenszyklus	ASEI-10 ergänzt die Profil-Logik durch eine konkrete Bewertung von Systemkontext, Autonomie, Exposition und Rückholbarkeit
ISO/IEC 23894	Leitlinien für KI-Risikomanagement	Internationaler Leitlinienstandard	Integration von KI-Risikomanagement in organisationale Aktivitäten und Funktionen	ASEI-10 kann als konkrete Bewertungs- und Priorisierungsmethode innerhalb dieses Rahmens eingesetzt werden
ISO/IEC 42001	Managementsystem für künstliche Intelligenz	Internationaler Managementsystemstandard	Richtlinien, Rollen, Verantwortlichkeiten, Prozesse, Dokumentation, Überwachung und kontinuierliche Verbesserung	ASEI-10 kann als internes Klassifikationsinstrument für KI-Inventare, Freigaben, Risikoregister und Neubewertungen dienen
Klassische Risikobewertung und Risikomatrizen	Bewertung von Risiken anhand von Eintrittswahrscheinlichkeit und Auswirkung	Bewertungsmethode / Risikomanagementinstrument	Kombination von Wahrscheinlichkeit und Schadensausmaß; häufig qualitative oder halbquantitative Risikoklassen	ASEI-10 bewertet keine Eintrittswahrscheinlichkeit, sondern strukturelles Risikopotenzial vor oder ergänzend zu konkreten Einzelrisikoanalysen
FMEA / FMECA	Analyse konkreter Fehlermöglichkeiten, Ursachen, Wirkungen und Maßnahmen	Fehleranalyseverfahren	Systematische Untersuchung von Fehlerarten, Fehlerursachen, Fehlerfolgen, Entdeckbarkeit und Kritikalität	ASEI-10 kann vorgelagert anzeigen, bei welchen KI-Systemen eine vertiefte Fehleranalyse besonders erforderlich ist
AI-Risk-Taxonomien und Risk Repositories	Sammlung und Ordnung von KI-Risikoarten	Taxonomie / Repository	Katalogisierung und Strukturierung technischer, gesellschaftlicher, rechtlicher, ethischer und organisatorischer KI-Risiken	ASEI-10 ist keine Risikoenzyklopädie, sondern überführt relevante Risikoarten in eine konkrete Systembewertung
Taxonomien systemischer Risiken aus General-Purpose AI	Beschreibung großskaliger Risiken allgemein einsetzbarer KI-Systeme	Wissenschaftliche Taxonomie / Makroperspektive	Systemische Wirkungen durch Skalierung, Vernetzung, Abhängigkeit, Missbrauch, Kontrollprobleme und gesellschaftliche Wechselwirkungen	ASEI-10 macht systemische Risikoperspektiven für einzelne Anwendungskontexte über Severität, Exposition, Autonomie und Irreversibilität bewertbar
Qualitäts-, Robustheits- und Lebenszyklusansätze	Anforderungen an Entwicklung, Betrieb, Überwachung und Veränderung von KI-Systemen	Qualitäts- und Lebenszyklusmodell	Datenqualität, Robustheit, Monitoring, Dokumentation, Veränderungskontrolle und kontinuierliche Verbesserung	ASEI-10 kann bestimmen, welche Prüf-, Qualitäts- und Überwachungsintensität aufgrund des Risikopotenzials angemessen ist
Scoring-Ansätze und Bewertungsmodelle	Operationalisierung komplexer Risikoeinschätzungen in Bewertungsstrukturen	Bewertungsmodell / Punktesystem / Reifegradmodell	Kriterien, Kategorien, Punktwerte, Scorebildung und Priorisierung	ASEI-10 ist selbst ein Scoringmodell, legt aber den heuristischen Charakter des Scores, die ordinalen Grenzen und die qualitative Begründung ausdrücklich offen

Tabelle 4: Vergleichende Übersicht bestehender Rahmenwerke und Bewertungsansätze im Verhältnis zu ASEI-10

Die Übersicht verdeutlicht, dass ASEI-10 keine bestehende Rahmenwerklogik ersetzt. Das Modell steht vielmehr auf einer eigenen methodischen Ebene. Es übernimmt nicht die Rolle rechtlicher Regulierung, organisatorischer Governance, technischer Detailprüfung oder umfassender Risikotaxonomie. Sein spezifischer Beitrag liegt in der kontextbezogenen Klassifikation des strukturellen Risikopotenzials konkreter KI-Systeme.

3.8 Zusammenfassung und Überleitung zum spezifischen Profil des ASEI-10-Modells

Die vorangegangenen Abschnitte zeigen, dass die Bewertung von KI-Risiken bereits durch eine Vielzahl rechtlicher, regulatorischer, organisatorischer, technischer und wissenschaftlicher Ansätze geprägt ist. Der EU AI Act schafft einen verbindlichen Rechtsrahmen. Die International AI Safety Reports bündeln wissenschaftliche Erkenntnisse zu fortgeschrittenen KI-Systemen und ihren Risiken. Governance- und Managementsystemansätze wie das NIST AI Risk Management Framework, ISO/IEC 23894 und ISO/IEC 42001 beschreiben Strukturen für den verantwortlichen Umgang mit KI. Klassische Risikobewertungen, Risikomatrizen, FMEA und FMECA bieten etablierte Verfahren zur Analyse von Risiken und Fehlermöglichkeiten. KI-spezifische Taxonomien, Risk Repositories, systemische Risikoansätze sowie Qualitäts-, Robustheits- und Lebenszyklusmodelle erweitern diese Perspektiven um die besonderen Eigenschaften künstlicher Intelligenz.

Diese Ansätze sind für die Bewertung und Steuerung von KI-Systemen unverzichtbar. Sie erfüllen jedoch unterschiedliche Funktionen. Einige legen rechtliche Pflichten fest, andere strukturieren Governance- und Managementprozesse, ordnen Risikoarten, analysieren konkrete Fehlermöglichkeiten, beschreiben technische Qualitätsanforderungen oder beobachten systemische Entwicklungslinien. Dadurch entsteht ein breites, aber zugleich fragmentiertes Feld von Bewertungs- und Steuerungsperspektiven.

Aus dieser Vielfalt ergibt sich eine methodische Lücke. Zwischen abstrakter Regulierung, umfassender Governance, allgemeinen Risikotaxonomien und technischen Detailprüfungen bleibt die Frage offen, wie das strukturelle Risikopotenzial eines konkreten KI-Systems im konkreten Anwendungskontext einheitlich, nachvollziehbar und vergleichbar klassifiziert werden kann.

Genau an dieser Stelle setzt das ASEI-10-Modell an. Es übernimmt nicht die Aufgabe eines Rechtsrahmens, eines vollständigen Managementsystems, einer technischen Sicherheitsprüfung oder einer Risikoenzyklopädie. Es stellt vielmehr eine eigenständige Klassifikationslogik bereit, mit der konkrete KI-Systeme anhand ihrer Autonomie, Severität, Exposition und Irreversibilität bewertet werden können.

Damit steht ASEI-10 nicht unterhalb der bestehenden Rahmenwerke, sondern neben ihnen auf einer eigenen methodischen Ebene. Die bestehenden Ansätze bilden den fachlichen und regulatorischen Bezugsrahmen. ASEI-10 ergänzt diesen Bezugsrahmen durch eine operative Bewertungslogik für einzelne KI-Systeme im Anwendungskontext.

Das folgende Kapitel entfaltet dieses spezifische Profil des ASEI-10-Modells. Es zeigt, worin der eigenständige methodische Anspruch besteht, warum die System- und Kontextorientierung zentral ist, wie die vierdimensionale Risikopotenziallogik aufgebaut ist und weshalb Scorebildung, Prüfschwellen und qualitative Begründung zusammengehören.

4 Das spezifische Profil des ASEI-10-Modells

4.1 Eigenständiger methodischer Anspruch

Das ASEI-10-Modell versteht sich als eigenständiger methodischer Beitrag zur Bewertung des Risikopotenzials von KI-Systemen. Es steht nicht unterhalb bestehender rechtlicher, regulatorischer, normativer oder organisatorischer Rahmenwerke, sondern neben ihnen auf einer eigenen Bewertungsebene. Sein Zweck besteht nicht darin, Rechtskonformität festzustellen, ein vollständiges Managementsystem zu ersetzen, technische Sicherheitsprüfungen durchzuführen oder alle denkbaren KI-Risiken taxonomisch zu erfassen. ASEI-10 beantwortet eine spezifische, bislang nur unzureichend operationalisierte Bewertungsfrage: Wie lässt sich das strukturelle Risikopotenzial eines konkreten KI-Systems in einem konkreten Anwendungskontext nachvollziehbar, vergleichbar und dokumentierbar klassifizieren?

Dieser Anspruch ist bewusst fokussiert. ASEI-10 will nicht alles leisten, was rechtliche Regulierung, Governance-Frameworks, Managementsystemnormen, Risikotaxonomien, Fehlermöglichkeitsanalysen oder technische Evaluationsverfahren leisten. Das Modell besetzt vielmehr die methodische Schnittstelle zwischen diesen Ansätzen. Es übersetzt zentrale Risikotreiber von KI-Systemen in eine anwendungsbezogene Klassifikationslogik und schafft damit eine Brücke zwischen abstrakter KI-Risikodiskussion und konkreter Organisationspraxis.

Die Eigenständigkeit des Modells ergibt sich aus seinem Bewertungsgegenstand. ASEI-10 bewertet nicht „KI“ im Allgemeinen, nicht ein Sprachmodell in abstrakter Form und nicht allein die technische Leistungsfähigkeit eines Modells. Bewertet wird ein konkretes KI-System mit seinen Funktionen, Berechtigungen, Datenzugriffen, Nutzergruppen, Einsatzgrenzen, Kontrollmechanismen und möglichen Folgen. Dadurch wird das Risikopotenzial dort erfasst, wo es praktisch entsteht: im Zusammenspiel von Systemfähigkeit, Anwendungskontext und möglicher Wirkung.

Zugleich grenzt sich ASEI-10 vom klassischen Risikobegriff ab, ohne dessen Bedeutung zu bestreiten. Das Modell berechnet keine Eintrittswahrscheinlichkeit konkreter Schadensereignisse und liefert keinen Erwartungswert im Sinne von Wahrscheinlichkeit mal Schadensausmaß. Es klassifiziert vielmehr das strukturelle Risikopotenzial eines Systems. Diese Unterscheidung ist wesentlich, weil bei vielen KI-Systemen belastbare Schadenswahrscheinlichkeiten noch nicht vorliegen oder stark vom konkreten Einsatz, vom Nutzerverhalten, von Schutzmaßnahmen, von Toolzugriffen und von dynamischen Systemänderungen abhängen.

ASEI-10 ist deshalb besonders für Situationen geeignet, in denen Organisationen eine erste, aber belastbare Einordnung ihres KI-Bestands benötigen. Das Modell kann bereits in Planungs-, Beschaffungs-, Pilotierungs-, Freigabe- oder Neubewertungsphasen eingesetzt werden. Es hilft zu bestimmen, welche Systeme geringe Aufmerksamkeit erfordern, welche Systeme vertieft geprüft werden sollten und welche Systeme aufgrund ihrer Risikotreiber besondere Governance-, Kontroll- oder Dokumentationsanforderungen auslösen.

Der methodische Anspruch von ASEI-10 liegt damit nicht in scheinbarer mathematischer Exaktheit, sondern in strukturierter Nachvollziehbarkeit. Das Modell erzeugt keine objektive Risikowahrheit, sondern eine begründete Klassifikation. Seine Qualität hängt daher nicht allein vom berechneten Score ab, sondern von der transparenten Beschreibung des Anwendungskontextes, der nachvollziehbaren Einstufung der vier Dimensionen, der Prüfung von Schwellenkriterien und der Dokumentation der zugrunde liegenden Annahmen.

In diesem Sinne ist ASEI-10 als eigenständiges Arbeitsmodell zu verstehen. Es kann neben bestehenden Rahmenwerken verwendet werden, ohne mit ihnen identisch zu sein. Regulatorische Vorgaben legen fest, welche rechtlichen Pflichten bestehen. Governance- und Managementsysteme beschreiben organisatorische Verantwortlichkeiten und Prozesse. Technische Prüfverfahren untersuchen konkrete Leistungs-, Sicherheits- oder Robustheitsfragen. ASEI-10 klassifiziert das strukturelle Risikopotenzial einzelner KI-Systeme im Anwendungskontext und stellt damit eine eigenständige Grundlage für Priorisierung, Freigabe, Kontrolle und weiterführende Prüfung bereit.

4.2 System- und Kontextorientierung

Der zentrale Bewertungsgegenstand des ASEI-10-Modells ist das konkrete KI-System im konkreten Anwendungskontext. Diese System- und Kontextorientierung ist ein Kernmerkmal des Modells. Sie unterscheidet ASEI-10 von Ansätzen, die vor allem auf Modellklassen, abstrakte Fähigkeiten, allgemeine Risikoarten oder regulatorische Fallgruppen abstellen.

Ein KI-Modell besitzt für sich genommen noch kein vollständig bestimmbares organisatorisches Risikopotenzial. Erst durch seine Einbettung in ein konkretes System entstehen jene Handlungsmöglichkeiten, Wirkungsräume und Folgen, die für eine Risikobewertung entscheidend sind. Dasselbe Modell kann als privater Schreibassistent, als interner Analysehelfer, als öffentlich erreichbarer Kundenservice-Chatbot, als Bewerberauswahlsystem oder als agentisches Steuerungssystem eingesetzt werden. Das technische Modell mag ähnlich sein; das Risikopotenzial ist es nicht.

Für ASEI-10 ist deshalb nicht allein relevant, was ein Modell grundsätzlich kann, sondern was das konkrete System im vorgesehenen Einsatz tatsächlich tun darf. Bewertungsrelevant sind insbesondere Funktionen, Berechtigungen, Datenzugriffe, Schnittstellen, Toolnutzung, Nutzergruppen, Betroffene, Freigabepunkte, Kontrollmechanismen, Einsatzgrenzen und mögliche Folgen. Ein System ohne Toolzugriff und ohne Außenwirkung ist anders zu bewerten als ein System, das E-Mails versenden, Daten verändern, Kundenprozesse auslösen, Bewerbungen vorsortieren oder technische Maßnahmen anstoßen kann.

Die Kontextorientierung verhindert pauschale Risikourteile. Ein leistungsfähiges KI-Modell ist nicht automatisch hochriskant, wenn es in einem engen, kontrollierten und leicht rückholbaren Nutzungskontext eingesetzt wird. Umgekehrt kann ein technisch weniger spektakuläres System ein erhebliches Risikopotenzial besitzen, wenn es in sensiblen Bereichen eingesetzt wird, viele Personen betrifft oder schwer korrigierbare Folgen erzeugen kann. ASEI-10 bewertet daher nicht die Faszination oder Leistungsfähigkeit eines Modells, sondern seine risikorelevante Einbettung.

Diese Perspektive ist besonders wichtig bei General-Purpose-AI-Modellen. Solche Modelle können für sehr unterschiedliche Zwecke verwendet werden. Eine pauschale Einstufung des Modells würde dieser Vielfalt nicht gerecht. ASEI-10 bewertet deshalb nicht das General-Purpose-AI-Modell als solches, sondern die konkrete Anwendungskonstellation. Entscheidend ist, ob aus einem allgemein einsetzbaren Modell ein begrenzter Assistent, ein organisationsweit genutztes Entscheidungshilfesystem, ein öffentlich exponierter Chatbot oder ein agentisches System mit Zugriff auf externe Werkzeuge wird.

Die System- und Kontextorientierung hat auch praktische Folgen für die Anwendung des Modells. Eine ASEI-10-Bewertung ist nur so belastbar wie die Beschreibung des Bewertungsgegenstands. Vor der Einstufung der vier Dimensionen muss daher geklärt werden, welches System bewertet wird, in welcher Umgebung es eingesetzt wird, welche Rechte es besitzt, welche Daten verarbeitet werden, wer betroffen ist und welche Folgen bei Fehlern, Missbrauch oder Kontrollproblemen denkbar sind. Ohne diese Kontextklärung wäre jede numerische Bewertung nur scheinpräzise.

Aus dieser Sicht ist die Systembeschreibung kein formaler Vorspann, sondern ein methodischer Kernbestandteil der Bewertung. Sie legt fest, welche Autonomiestufe sachgerecht ist, welcher Severitätsgrad berührt wird, welche Exposition entsteht und wie reversibel mögliche negative Folgen sind. Ändert sich der Systemkontext, ändert sich auch die Bewertung. Neue Toolzugriffe, zusätzliche Nutzergruppen, ein sensiblerer Einsatzbereich, breitere Verfügbarkeit oder veränderte Kontrollmechanismen können das Risikopotenzial erheblich verschieben.

Damit macht ASEI-10 eine einfache, aber folgenreiche Grundannahme operationalisierbar: KI-Risiken entstehen nicht allein im Modell, sondern im Zusammenspiel von Modellfähigkeit, Systemgestaltung und Anwendungskontext. Diese Annahme bildet die Grundlage für die vier Bewertungsdimensionen des Modells.

4.3 Vierdimensionale Risikopotenziallogik

Das ASEI-10-Modell beruht auf der Annahme, dass das Risikopotenzial eines KI-Systems nicht aus einer einzelnen Eigenschaft abgeleitet werden kann. Weder die technische Leistungsfähigkeit eines Modells noch der Einsatzbereich, die Nutzerzahl oder die Autonomie eines Systems reichen für sich genommen aus, um das Risikopotenzial angemessen zu bestimmen. Entscheidend ist das Zusammenspiel mehrerer Risikotreiber.

ASEI-10 verdichtet dieses Zusammenspiel in vier Bewertungsdimensionen: Autonomie, Severität, Exposition und Irreversibilität. Diese vier Dimensionen bilden keine beliebige Merkmalsliste, sondern erfassen unterschiedliche Grundfragen der Risikopotenzialbewertung.

Autonomie fragt, wie selbstständig ein KI-System operieren kann. Sie betrifft die Handlungstiefe des Systems: Reagiert es nur auf Eingaben, unterstützt es Nutzerhandlungen, verwendet es Werkzeuge, verfolgt es Ziele über mehrere Schritte hinweg oder kann es über längere Zeit eigenständig aktiv bleiben? Autonomie ist risikorelevant, weil mit zunehmender Handlungstiefe die Möglichkeit wächst, dass ein System nicht nur Vorschläge erzeugt, sondern Prozesse beeinflusst, Handlungen auslöst oder Wirkungen verstärkt.

Severität fragt, wie schwer mögliche negative Folgen im konkreten Einsatzbereich sein können. Sie betrifft die Bedeutung der berührten Schutzgüter und die Folgenschwere möglicher Fehler, Fehlverwendung oder Kontrollprobleme. Ein privater kreativer Nutzungskontext ist anders zu bewerten als Bildung, Beschäftigung, Gesundheit, Recht, öffentliche Sicherheit, kritische Infrastruktur oder Cybersecurity. Severität macht sichtbar, dass dieselbe technische Funktion in unterschiedlichen Kontexten sehr unterschiedliche Tragweite besitzen kann.

Exposition fragt, wie breit der mögliche Wirkungsraum eines KI-Systems ist. Sie betrifft die Zahl und Art der Betroffenen, die organisatorische Reichweite, die öffentliche Sichtbarkeit und mögliche systemische Ausbreitung. Ein Fehler, der nur eine einzelne Nutzerin oder einen einzelnen Nutzer betrifft, hat eine andere Reichweite als ein Fehler, der eine Kursgruppe, eine Organisation, eine Kundengruppe, die Öffentlichkeit oder mehrere verbundene Systeme erreicht.

Irreversibilität fragt, wie gut negative Folgen nachträglich erkannt, gestoppt, korrigiert, zurückgeholt oder kompensiert werden können. Sie betrifft nicht die Schwere der Folge als solche, sondern deren Rückholbarkeit. Ein fehlerhafter Textentwurf kann meist verworfen oder korrigiert werden. Eine veröffentlichte Falschinformation, eine Benachteiligung im Bewerbungsprozess, ein Datenschutzvorfall, eine sicherheitskritische Fehlreaktion oder ein körperlicher Schaden kann dagegen nur begrenzt oder gar nicht vollständig rückgängig gemacht werden.

Die vier Dimensionen sind analytisch getrennt, aber praktisch miteinander verbunden. Ein System kann durch hohe Autonomie risikorelevant werden, auch wenn sein Einsatzbereich zunächst nicht hochkritisch erscheint. Ein System kann durch hohe Severität problematisch sein, obwohl es nur assistiv arbeitet. Ein System kann wegen breiter Exposition besondere Aufmerksamkeit erfordern, obwohl die einzelne Interaktion begrenzt wirkt. Und ein System kann wegen schwer rückholbarer Folgen kritisch sein, selbst wenn nur wenige Personen unmittelbar betroffen sind.

Gerade diese Trennung verhindert eine vorschnelle Pauschalbewertung. ASEI-10 fragt nicht nur, ob ein System „gefährlich“ oder „harmlos“ ist. Das Modell macht sichtbar, wodurch das Risikopotenzial entsteht. Zwei Systeme können denselben Score erreichen, aber aus unterschiedlichen Gründen: Das eine wegen hoher Autonomie und begrenzter Exposition, das andere wegen geringer Autonomie, aber hoher Severität und schwer reversibler Folgen. Für Governance, Kontrolle und Mitigation ist diese Unterscheidung entscheidend.

Die vierdimensionale Logik dient daher nicht nur der Berechnung eines Scores, sondern auch der diagnostischen Aufklärung des Risikoprofils. Sie zeigt, an welcher Stelle ein System begrenzt, überwacht, dokumentiert oder vertieft geprüft werden muss. Wird das Risikopotenzial vor allem durch Autonomie getrieben, stehen Handlungskontrolle, Toolbegrenzung und Freigabepunkte im Vordergrund. Wird es vor allem durch Severität getragen, sind fachliche Prüfung und Schutzgüteranalyse zentral. Bei hoher Exposition werden Rollout, Nutzerkreis, Skalierung und Außenwirkung bedeutsam. Bei hoher Irreversibilität rücken Rückholbarkeit, Korrekturverfahren, Beschwerdewege und Notfallmechanismen in den Vordergrund.

Damit bildet die vierdimensionale Risikopotenziallogik den Kern des ASEI-10-Modells. Sie verbindet die System- und Kontextorientierung des Modells mit einer strukturierten Bewertungsarchitektur. Aus ihr ergeben sich die späteren Einzelskalen, die Scorebildung, die Prüfschwellen und die Ableitung praktischer Mindestanforderungen.

4.4 Verbindung qualitativer Analyse mit rechnerischer Indexbildung

ASEI-10 verbindet qualitative Analyse mit rechnerischer Indexbildung. Diese Verbindung ist für das Modell wesentlich. Das Risikopotenzial eines KI-Systems wird nicht allein intuitiv beschrieben, aber auch nicht allein rechnerisch bestimmt. Die Bewertung entsteht in zwei aufeinander bezogenen Schritten: Zunächst werden die vier Dimensionen fachlich begründet eingestuft. Anschließend werden diese Einstufungen in eine nachvollziehbare Scorebildung überführt.

Die qualitative Analyse bildet den Ausgangspunkt. Für jede Dimension muss begründet werden, warum eine bestimmte Stufe gewählt wird. Bei der Autonomie ist zu prüfen, welche operative Selbstständigkeit das System besitzt. Bei der Severität ist zu untersuchen, welche Schutzgüter und möglichen Folgenschweren berührt werden. Bei der Exposition ist zu bestimmen, wie weit der Wirkungsraum reicht. Bei der Irreversibilität ist zu klären, wie gut mögliche negative Folgen erkannt, gestoppt, korrigiert oder kompensiert werden können. Ohne diese fachliche Begründung wäre der spätere Score nicht belastbar.

Die rechnerische Indexbildung erfüllt eine andere Funktion. Sie verdichtet die vier Einzeleinstufungen zu einem gemeinsamen Klassifikationswert. Dadurch werden unterschiedliche KI-Systeme vergleichbarer, priorisierbarer und dokumentierbarer. Eine Organisation kann auf dieser Grundlage erkennen, welche Systeme nur geringe risikobezogene Aufmerksamkeit benötigen, welche Systeme vertieft geprüft werden sollten und welche Systeme besondere Kontroll-, Freigabe- oder Dokumentationsanforderungen auslösen.

ASEI-10 verwendet dafür zunächst einen multiplikativen Rohwert. Dieser Rohwert bildet ab, dass sich die vier Bewertungsdimensionen gegenseitig verstärken können. Hohe Autonomie ist risikorelevanter, wenn sie in einem schwerwiegenden Einsatzbereich auftritt. Hohe Severität wird bedeutsamer, wenn viele Personen oder Systeme betroffen sind. Breite Exposition ist kritischer, wenn Folgen schwer rückholbar sind. Das Modell behandelt die Dimensionen daher nicht als bloß additive Einzelaspekte, sondern als miteinander verbundene Risikotreiber.

Der Rohwert wird anschließend logarithmisch auf eine Skala von 1 bis 10 normiert. Diese Normierung macht die Ergebnisse praktikabler interpretierbar, ohne den Score als linearen Risikowert auszugeben. Ein ASEI-10-Score von 8 ist deshalb nicht „doppelt so riskant“ wie ein Score von 4. Der Score ist ein heuristischer Klassifikationsindex. Er ordnet Systeme in Risikobereiche ein, schafft Vergleichbarkeit und unterstützt Priorisierung, ist aber keine exakte Messgröße im naturwissenschaftlichen oder versicherungsmathematischen Sinn.

Die Stärke des Modells liegt gerade darin, dass es die rechnerische Verdichtung transparent macht. Die Bewertungsdimensionen, die Normierung der Autonomie, die Bildung des Rohwerts und die logarithmische Transformation werden offengelegt. Dadurch bleibt nachvollziehbar, wie ein Score zustande kommt. ASEI-10 vermeidet damit die Intransparenz vieler bloßer Punktesysteme, bei denen unklar bleibt, warum bestimmte Kriterien wie gewichtet werden oder welche Bedeutung der Endwert tatsächlich besitzt.

Gleichzeitig darf der Score nicht isoliert gelesen werden. Zwei Systeme können denselben ASEI-10-Score erreichen und dennoch unterschiedliche Risikoprofile besitzen. Ein System kann vor allem durch hohe Autonomie und breite Exposition risikorelevant sein, ein anderes durch hohe Severität und schwer reversible Folgen. Deshalb müssen Score, Einzeldimensionen und Begründung immer gemeinsam betrachtet werden. Die Zahl erleichtert die Orientierung; die fachliche Analyse erklärt das Ergebnis.

Diese Verbindung von qualitativer und quantitativer Logik ist ein wesentlicher Profilkern von ASEI-10. Das Modell vermeidet sowohl rein narrative Unschärfe als auch rechnerische Scheingenauigkeit. Es schafft eine strukturierte, dokumentierbare und vergleichbare Bewertung, ohne den fachlichen Charakter der Einstufung zu verdecken. Damit eignet sich ASEI-10 besonders für Organisationen, die KI-Systeme nicht nur beschreiben, sondern methodisch begründet priorisieren und steuern wollen.

4.5 Prüfschwellen als Schutz gegen verdeckte Einzelrisiken

Ein zentrales Merkmal des ASEI-10-Modells sind die ergänzenden Prüfschwellen. Sie stellen sicher, dass qualitativ bedeutsame Einzelmerkmale nicht durch den zusammenfassenden ASEI-10-Score verdeckt werden. Damit schützt das Modell vor einer typischen Schwäche reiner Scoringverfahren: Ein Gesamtscore kann den Eindruck vermitteln, ein System sei insgesamt unkritisch oder nur moderat risikobehaftet, obwohl einzelne Merkmale besondere Aufmerksamkeit verlangen.

Die Prüfschwellen ergänzen die rechnerische Scorebildung um eine qualitative Sicherheitslogik. Sie greifen dort, wo bestimmte Ausprägungen einer Dimension unabhängig vom Gesamtscore zusätzliche Prüf-, Kontroll- oder Dokumentationsanforderungen auslösen können. Ein KI-System kann beispielsweise wegen agentischer Handlungsmuster, eines hochkritischen Einsatzbereichs, systemischer Exposition oder schwer reversibler Folgen besonders prüfbedürftig sein, auch wenn der rechnerische Score allein noch keine sehr hohe Risikoklasse ergibt.

Diese Logik ist besonders wichtig, weil sich Risikopotenziale nicht immer gleichmäßig über alle Dimensionen verteilen. Ein System kann in drei Dimensionen moderat ausgeprägt sein, aber in einer Dimension eine kritische Schwelle erreichen. Wird nur der Gesamtscore betrachtet, besteht die Gefahr, dass dieses Einzelmerkmal im Durchschnittseffekt verschwindet. Die Prüfschwellen verhindern eine solche Verdeckung und zwingen dazu, kritische Dimensionen gesondert zu betrachten.

Beispielhaft kann ein KI-System mit begrenzter Autonomie dennoch eine hohe Severität aufweisen, wenn es in einem Beschäftigungs-, Gesundheits-, Rechts- oder Sicherheitskontext eingesetzt wird. Ebenso kann ein System mit moderater Severität besonders relevant werden, wenn seine Ausgaben öffentlich, organisationsweit oder systemisch verbreitet werden. Auch schwer reversible Folgen können eine besondere Prüfung rechtfertigen, selbst wenn die Zahl der unmittelbar Betroffenen begrenzt ist.

Die Prüfschwellen haben damit eine doppelte Funktion. Einerseits erhöhen sie die methodische Robustheit des Modells, weil sie verhindern, dass der Score isoliert und mechanisch interpretiert wird. Andererseits unterstützen sie die praktische Governance, weil sie konkrete Anlässe für zusätzliche Mindestanforderungen liefern. Dazu können etwa erweiterte Dokumentation, fachliche Freigabe, menschliche Aufsicht, Monitoring, Protokollierung, Eskalationswege, Rückholmechanismen, Datenschutzprüfung, Sicherheitsprüfung oder regelmäßige Neubewertung gehören.

Wichtig ist dabei, dass Prüfschwellen nicht als zweiter Score neben dem ASEI-10-Score zu verstehen sind. Sie dienen nicht dazu, eine parallele Rechenlogik aufzubauen. Ihr Zweck besteht darin, qualitativ kritische Merkmale sichtbar zu halten. Der Score liefert die zusammenfassende Klassifikation; die Prüfschwellen markieren besondere Risikotreiber, die unabhängig vom Gesamtergebnis begründet und bearbeitet werden müssen.

Damit stärken die Prüfschwellen den Anspruch von ASEI-10, nicht nur vergleichbar, sondern auch fachlich verantwortbar zu bewerten. Das Modell verbindet numerische Verdichtung mit qualitativer Vorsicht. Es erlaubt eine operative Priorisierung, ohne kritische Einzelmerkmale dem Gesamtscore zu opfern.

In der Anwendung bedeutet dies: Jede ASEI-10-Bewertung muss nicht nur den Score ausweisen, sondern auch prüfen, ob eine oder mehrere Schwellen erreicht werden. Erst aus der Verbindung von Score, Einzeldimensionen, Prüfschwellen und Begründung ergibt sich eine belastbare Bewertung des Risikopotenzials.

4.6 Praktische Anschlussfähigkeit an Governance, Compliance und Risikomanagement

Das ASEI-10-Modell ist nicht nur als theoretischer Klassifikationsrahmen angelegt, sondern als praktisch nutzbares Arbeitsmodell für Organisationen. Sein Wert zeigt sich besonders dort, wo KI-Systeme nicht nur abstrakt diskutiert, sondern konkret eingeführt, freigegeben, überwacht, dokumentiert und bei Bedarf neu bewertet werden müssen.

In vielen Organisationen entsteht mit zunehmender KI-Nutzung ein Steuerungsproblem. Unterschiedliche KI-Systeme werden in unterschiedlichen Bereichen eingesetzt: als Textassistenz, Analysewerkzeug, Kundenservice-System, Recruiting-Unterstützung, Entwicklungshelfer, Unternehmensagent, Prognoseinstrument oder sicherheitsrelevante Anwendung. Ohne ein einheitliches Bewertungsverfahren besteht die Gefahr, dass solche Systeme uneinheitlich eingeschätzt, unvollständig dokumentiert oder nach subjektiven Einzelurteilen freigegeben werden.

ASEI-10 kann hier eine standardisierende Funktion übernehmen. Das Modell stellt ein einheitliches Bewertungsraster bereit, mit dem KI-Systeme nach denselben Grundfragen beurteilt werden: Wie selbstständig handelt das System? Wie schwer können negative Folgen im Einsatzbereich sein? Wie weit reicht der mögliche Wirkungsraum? Wie gut lassen sich negative Folgen korrigieren oder rückgängig machen? Dadurch wird die Bewertung unterschiedlicher KI-Systeme vergleichbarer und besser dokumentierbar.

Für die KI-Governance kann ASEI-10 insbesondere als Instrument für KI-Inventare, Freigabeprozesse und Risikoregister eingesetzt werden. Ein KI-Inventar erfasst nicht nur, welche Systeme vorhanden sind, sondern auch, welches Risikopotenzial sie besitzen. Freigabeprozesse können nach Scorebereichen und Prüfschwellen gestuft werden. Risikoregister können die vier Einzeldimensionen, den ASEI-10-Score, ausgelöste Prüfschwellen, Begründungen, Verantwortlichkeiten und vorgesehene Maßnahmen dokumentieren.

Auch für Compliance-Vorprüfungen ist das Modell anschlussfähig. ASEI-10 ersetzt keine rechtliche Prüfung und stellt keine Rechtskonformität fest. Es kann aber helfen, Systeme zu identifizieren, bei denen eine vertiefte Prüfung besonders naheliegt. Ein hoher Severitätsgrad, systemische Exposition, schwer reversible Folgen oder agentische Handlungsmuster können Hinweise darauf geben, dass ergänzende rechtliche, datenschutzrechtliche, arbeitsrechtliche, sicherheitsbezogene oder regulatorische Prüfungen erforderlich sind.

Im Risikomanagement unterstützt ASEI-10 die Priorisierung. Nicht jedes KI-System benötigt denselben Prüf- und Kontrollaufwand. Ein privater oder interner Schreibassistent ohne Toolzugriff und ohne Außenwirkung ist anders zu behandeln als ein KI-System, das Bewerbungen vorsortiert, Kundenprozesse beeinflusst, personenbezogene Daten verarbeitet oder technische Maßnahmen auslösen kann. ASEI-10 macht solche Unterschiede sichtbar und hilft, Ressourcen auf die Systeme zu konzentrieren, deren Risikopotenzial besondere Aufmerksamkeit erfordert.

Die Anschlussfähigkeit des Modells zeigt sich auch bei der Ableitung von Mindestanforderungen. Je nach Scorebereich und Prüfschwelle können unterschiedliche Anforderungen formuliert werden: einfache Dokumentation, fachliche Prüfung, verpflichtende Freigabe, menschliche Aufsicht, Protokollierung, Monitoring, Datenschutzprüfung, Sicherheitsanalyse, Red Teaming, Rückholmechanismen, Beschwerdewege oder regelmäßige Neubewertung. ASEI-10 liefert damit nicht nur eine Einordnung, sondern einen Ausgangspunkt für konkrete Steuerungsentscheidungen.

Besonders wichtig ist die Neubewertung bei veränderten Systembedingungen. KI-Systeme bleiben selten statisch. Neue Funktionen, zusätzliche Toolzugriffe, andere Nutzergruppen, breitere Rollouts, veränderte Datenquellen oder neue Einsatzbereiche können das Risikopotenzial deutlich verändern. ASEI-10 kann als Auslöser und Struktur für solche Neubewertungen dienen. Eine Änderung des Anwendungskontextes ist dann nicht nur ein technisches Update, sondern ein Anlass zur erneuten Risikopotenzialklassifikation.

Damit verbindet ASEI-10 strategische Governance mit operativer Anwendung. Das Modell übersetzt allgemeine Anforderungen an verantwortliche KI-Nutzung in eine konkrete Bewertungslogik für einzelne Systeme. Es ermöglicht Organisationen, KI-Systeme nicht pauschal zu erlauben oder zu verbieten, sondern differenziert nach ihrem tatsächlichen Risikopotenzial zu steuern.

Die praktische Stärke des ASEI-10-Modells liegt daher in seiner Vermittlungsfunktion. Es verbindet Governance, Compliance und Risikomanagement mit einer nachvollziehbaren Systembewertung. Dadurch kann es helfen, Verantwortlichkeiten zu klären, Prüfintensitäten festzulegen, Entscheidungen zu dokumentieren und den KI-Einsatz innerhalb einer Organisation kontrollierbarer zu machen.

4.7 Zusammenfassung und Überleitung zu den Bewertungsdimensionen

Das spezifische Profil des ASEI-10-Modells ergibt sich aus seiner eigenständigen methodischen Position. ASEI-10 ist kein Rechtsrahmen, kein Managementsystem, keine technische Sicherheitsprüfung und keine bloße Risikoliste. Das Modell stellt eine operative Klassifikationslogik bereit, mit der das strukturelle Risikopotenzial konkreter KI-Systeme im konkreten Anwendungskontext bewertet werden kann.

Im Mittelpunkt steht dabei die Verbindung mehrerer Grundentscheidungen. Erstens bewertet ASEI-10 nicht KI im Allgemeinen, sondern konkrete Systeme mit bestimmten Funktionen, Berechtigungen, Datenzugriffen, Nutzergruppen, Kontrollmechanismen und möglichen Folgen. Zweitens beruht das Modell auf einer vierdimensionalen Risikopotenziallogik, die Autonomie, Severität, Exposition und Irreversibilität getrennt erfasst und anschließend zusammenführt. Drittens verbindet ASEI-10 qualitative Begründung mit rechnerischer Indexbildung, ohne den Score als exakte Risikomessung misszuverstehen. Viertens verhindern Prüfschwellen, dass kritische Einzelmerkmale durch den Gesamtscore verdeckt werden. Fünftens ist das Modell praktisch anschlussfähig an Governance, Compliance, Risikomanagement, KI-Inventare, Freigabeprozesse und Neubewertungen.

Damit entsteht ein Bewertungsrahmen, der weder rein abstrakt noch rein technisch ist. ASEI-10 vermittelt zwischen strategischer KI-Governance und konkreter Systembewertung. Es hilft Organisationen, KI-Systeme nicht pauschal zu überschätzen oder zu verharmlosen, sondern nach ihren tatsächlichen Risikotreibern zu unterscheiden.

Die folgenden Kapitel entfalten die vier Bewertungsdimensionen des Modells im Einzelnen. Zunächst wird die Autonomiedimension beschrieben, weil sie die operative Selbstständigkeit eines KI-Systems erfasst und damit einen zentralen Ausgangspunkt der Risikopotenzialbewertung bildet. Anschließend folgen Severität, Exposition und Irreversibilität. Erst aus dem Zusammenspiel dieser vier Dimensionen ergibt sich die spätere Scorebildung des ASEI-10-Modells.

5 Dimension A: Autonomie

5.1 Begriff und risikobezogene Bedeutung von Autonomie

Autonomie bezeichnet im ASEI-10-Modell den Grad, in dem ein KI-System ohne unmittelbare menschliche Einzelsteuerung Aufgaben ausführen, Handlungsschritte planen, Werkzeuge auswählen, Zwischenergebnisse bewerten und über Zeit hinweg aktiv bleiben kann. Die Dimension erfasst damit nicht die Leistungsfähigkeit eines Modells im engeren Sinne, sondern die operative Selbstständigkeit des konkreten KI-Systems im Anwendungskontext. Ein System mit hoher Modellqualität, das ausschließlich auf einzelne Nutzereingaben reagiert, weist eine deutlich geringere Autonomie auf als ein agentisches System, das eigenständig Teilziele bildet, externe Werkzeuge nutzt, Aktionen vorbereitet oder ausführt und auf Ereignisse proaktiv reagiert. Für die Risikobewertung ist diese Unterscheidung zentral, weil mit wachsender Autonomie die menschliche Kontrolle von einer unmittelbaren Einzelprüfung zu einer indirekten Systemaufsicht verschoben wird. Dadurch steigt die Bedeutung von Freigaberegeln, Begrenzungen, Protokollierung, Eingriffsmöglichkeiten und Abschaltmechanismen.

Ausgehend von dieser Grundlogik unterscheidet das ASEI-10-Modell neun Autonomiestufen. Diese werden nicht als bloße lineare Rangfolge verstanden, sondern in drei übergeordnete Klassen mit jeweils drei Stufen gegliedert. Die erste Klasse umfasst assistive KI-Systeme, bei denen die Kontrolle im Wesentlichen beim Menschen verbleibt. Die zweite Klasse beschreibt agentische KI-Systeme, die Aufgaben mehrstufig planen, Werkzeuge selbst auswählen und innerhalb definierter Grenzen eigenständig handeln können. Die dritte Klasse erfasst strategisch-systemische KI-Systeme, bei denen Verteilung, Selbstoptimierung und Kontrollierbarkeit selbst zu zentralen Bewertungsfragen werden. Diese 3×3-Struktur soll die Autonomieskala sowohl differenziert als auch übersichtlich anwendbar machen.

5.2 Die Autonomieskala A1–A9: Drei Klassen der Systemselbstständigkeit

Die Autonomieskala A1–A9 operationalisiert die Dimension Autonomie für die Anwendung des ASEI-10-Modells. Sie dient dazu, die operative Selbstständigkeit eines KI-Systems nicht nur begrifflich zu beschreiben, sondern in einen bewertbaren Skalenwert zu überführen. Entscheidend ist dabei nicht die abstrakte Leistungsfähigkeit des zugrunde liegenden Modells, sondern die konkrete Systemkonfiguration: Welche Aufgaben darf das System selbst auslösen, planen, durchführen, überwachen oder fortsetzen? Die Einstufung erfolgt nach der höchsten Autonomiestufe, deren Merkmale im konkreten Anwendungskontext tatsächlich erfüllt sind.

Klasse	Stufe	Kurzbezeichnung	Kerngedanke
Klasse I	A1	Passives Grundmodell	Input → Output ohne Systemkontext
	A2	Reaktiver Dialogassistent	Dialogische Reaktion ohne externe Handlung
	A3	Werkzeugassistenz	Toolnutzung auf Nutzeranweisung
Klasse II	A4	Überwachter Agent	mehrstufige Planung unter Freigabe
	A5	Begrenzter autonomer Agent	eigenständiges Handeln in definierter Domäne
	A6	Persistenter autonomer Agent	dauerhafte oder ereignisbasierte Eigenaktivität
Klasse III	A7	Verteilter Agentenverbund	Koordination mehrerer Agenten/ Systembereiche
	A8	Selbstoptimierender strategischer Agent	Verbesserung eigener Strategien oder Toolketten
	A9	Nicht zuverlässig kontrollierbarer KI-Akteur	keine belastbare Begrenzung oder Rückholbarkeit

Tabelle 5: Die neun Autonomiestufen des ASEI-10-Modells

5.3 Beschreibung und Abgrenzung der Autonomiestufen

5.3.1 Klasse I: Assistive KI-Systeme

Die erste Klasse umfasst KI-Systeme, deren Funktion im Wesentlichen auf Unterstützung, Auskunft, Analyse, Generierung oder begrenzte Werkzeugnutzung ausgerichtet ist. Charakteristisch ist, dass die Zielsetzung, Auslösung und Bewertung der Ergebnisse überwiegend beim Menschen verbleiben. Assistive Systeme können fachlich sehr leistungsfähig sein und dennoch eine geringe Autonomie aufweisen, solange sie nicht eigenständig Ziele verfolgen, keine mehrstufigen Handlungsprozesse steuern und keine Aktionen ohne menschliche Veranlassung ausführen. Das Risikopotenzial dieser Klasse liegt daher vor allem in der Qualität, Richtigkeit, Angemessenheit und möglichen Fehlinterpretation der erzeugten Inhalte, weniger in eigenständiger Systemaktion.

Klasse I umfasst assistive KI-Systeme mit geringer operativer Selbstständigkeit. Die Systeme dieser Klasse unterstützen menschliche Nutzer durch Ausgabe, Analyse, Generierung oder begrenzte Werkzeugnutzung, verfolgen jedoch keine eigenständigen Ziele und übernehmen keine autonome Prozesssteuerung. Die Stufen A1 bis A3 bilden dabei eine aufsteigende Autonomieskala. Die jeweilige Stufennummer entspricht zugleich dem Bewertungswert innerhalb der Dimension Autonomie: A1 wird mit einem Punkt, A2 mit 2 Punkten und A3 mit 3 Punkten angesetzt.

A1 – Passives Grundmodell

Ein KI-System wird der Stufe A1 zugeordnet, wenn es ausschließlich als passives Grundmodell wirkt. Es verarbeitet eine konkrete Eingabe und erzeugt daraus eine Ausgabe, ohne eine eigenständige Dialogrolle, einen nutzerbezogenen Systemkontext oder operative Zusatzfunktionen einzunehmen. Das System führt keine Werkzeuge aus, greift nicht auf externe Datenquellen zu, verwaltet keine Aufgaben und löst keine Handlungen außerhalb der unmittelbaren Ausgabe aus. Seine Funktion beschränkt sich auf das Muster „Eingabe – Verarbeitung – Ausgabe“.

Die Autonomie ist auf dieser Stufe minimal. Das System reagiert nicht im Sinne eines fortlaufenden Assistenzdialogs, sondern erzeugt lediglich einen Output auf Basis eines Inputs. Typische Risiken liegen vor allem in fehlerhaften, verzerrten, unvollständigen oder missverständlichen Ausgaben. Ein A1-System kann inhaltlich leistungsfähig sein, besitzt aber keine operative Selbstständigkeit.

Die Abgrenzung zu A2 liegt darin, dass A1 keinen echten dialogischen Assistenzkontext aufweist. Sobald ein System einen fortlaufenden Gesprächszusammenhang berücksichtigt, nutzerorientiert nachfragt, vorherige Eingaben einbezieht oder als reaktiver Dialogassistent in einer Anwendung erscheint, ist A1 nicht mehr ausreichend.

A2 – Reaktiver Dialogassistent

Ein KI-System wird der Stufe A2 zugeordnet, wenn es als reaktiver Dialogassistent arbeitet. Es kann einen Gesprächsverlauf berücksichtigen, Antworten an den Nutzungskontext anpassen, Erläuterungen geben, Texte formulieren, Code vorschlagen, Inhalte zusammenfassen oder auf Rückfragen reagieren. Es bleibt jedoch vollständig auf menschliche Eingaben angewiesen. Die Auslösung der Interaktion, die Zielsetzung und die Bewertung der Ergebnisse liegen beim Nutzer.

Ein A2-System führt keine externen Werkzeuge aus und nimmt keine eigenständigen Handlungen außerhalb der Antwortgenerierung vor. Es kann zwar komplexe Inhalte erzeugen und in einem Dialog konsistent erscheinen, aber es verändert keine Dateien, ruft keine externen Systeme auf, versendet keine Nachrichten, startet keine Prozesse und trifft keine operativen Entscheidungen. Seine Autonomie besteht in der dialogischen Reaktion, nicht in eigenständiger Handlung.

Die Abgrenzung zu A1 liegt im Dialogkontext: A2 ist nicht nur ein passives Grundmodell, sondern ein eingebetteter Assistent, der Nutzerabsicht, Gesprächsverlauf und Antwortformat berücksichtigt. Die Abgrenzung zu A3 liegt im fehlenden Werkzeugzugriff. Sobald das System auf Nutzeranweisung Dateien öffnet, Webrecherchen durchführt, Code ausführt, Kalenderdaten bearbeitet, externe Daten abruft oder andere Tools nutzt, ist A2 nicht mehr ausreichend.

A3 – Werkzeugassistenz auf Nutzeranweisung

Ein KI-System wird der Stufe A3 zugeordnet, wenn es zusätzlich zur dialogischen Assistenz Werkzeuge nutzen kann, diese Nutzung jedoch an eine konkrete menschliche Anweisung gebunden bleibt. Das System kann beispielsweise Dateien analysieren, Webquellen abrufen, Berechnungen durchführen, Code ausführen, Tabellen auswerten, Bilder interpretieren oder über klar begrenzte Schnittstellen auf externe Dienste zugreifen. Entscheidend ist, dass die Werkzeugnutzung nicht aus einer eigenständigen Zielverfolgung des Systems entsteht, sondern aus einem expliziten Nutzerauftrag.

Die Autonomie ist höher als bei A2, weil das System nicht mehr nur Inhalte erzeugt, sondern über Werkzeuge operative Zwischenschritte ausführen kann. Gleichwohl bleibt es assistiv: Der Mensch gibt Ziel, Anlass und Rahmen der Handlung vor. Das System entscheidet nicht dauerhaft selbst, welche übergeordneten Ziele zu verfolgen sind, führt keine längeren Handlungsketten ohne menschliche Kontrolle aus und bleibt auf eine konkrete Aufgabenstellung bezogen.

Die Abgrenzung zu A2 liegt im tatsächlichen Werkzeugzugriff. Ein System, das nur erklärt, wie eine Datei ausgewertet oder eine Recherche durchgeführt werden könnte, bleibt A2. Ein System, das die Datei tatsächlich liest, die Recherche tatsächlich ausführt oder den Code tatsächlich laufen lässt, ist A3. Die Abgrenzung zu A4 liegt in der fehlenden agentischen Prozesssteuerung. Sobald das System ein Ziel selbstständig in mehrere Teilaufgaben zerlegt, Werkzeuge eigenständig auswählt, Zwischenergebnisse bewertet und einen mehrstufigen Arbeitsprozess unter nur noch überwachender menschlicher Kontrolle ausführt, ist A3 nicht mehr ausreichend.

5.3.2 Klasse II: Agentische KI-Systeme

Klasse II umfasst agentische KI-Systeme, die über eine bloß assistive Nutzung hinausgehen. Charakteristisch ist, dass diese Systeme Aufgaben nicht nur beantworten oder mit einzelnen Werkzeugen unterstützen, sondern Zielvorgaben in Handlungsschritte übersetzen, Werkzeuge auswählen, Zwischenergebnisse verarbeiten und innerhalb definierter Grenzen eigenständig fortfahren können. Damit verschiebt sich die menschliche Kontrolle: Sie besteht nicht mehr ausschließlich in der unmittelbaren Einzelanweisung, sondern zunehmend in Vorgaben, Freigaben, Überwachung, Begrenzung und nachträglicher Kontrolle. Das Risikopotenzial steigt deshalb nicht nur durch höhere technische Leistungsfähigkeit, sondern vor allem durch die Fähigkeit des Systems, Handlungsketten zu erzeugen und operative Entscheidungen innerhalb eines vorgegebenen Rahmens selbstständig auszuführen.

Für die Bewertung im ASEI-10-Modell bleiben auch die Stufen der Klasse II als lineare Autonomiewerte erhalten. A4 wird mit 4 Punkten, A5 mit 5 Punkten und A6 mit 6 Punkten angesetzt. Eine zusätzliche Quadrierung dieser Werte erfolgt nicht, weil das Risikopotenzial agentischer Systeme nicht allein aus dem Autonomiegrad folgt, sondern durch Severität, Exposition und Irreversibilität wesentlich mitbestimmt wird. Ein agentisches System in einem harmlosen Testkontext ist anders zu bewerten als ein weniger autonomes System in einem hochkritischen Einsatzbereich. Autonomiebedingte Sprungstellen werden deshalb nicht durch künstliche Exponenten, sondern durch verbindliche Prüfschwellen, Schwellenkriterien und besondere Anforderungen an Freigabe, Protokollierung, Eingriffsmöglichkeit und Abschaltbarkeit berücksichtigt.

A4 – Überwachter Agent

Ein KI-System wird der Stufe A4 zugeordnet, wenn es ein vorgegebenes Ziel in mehrere Teilaufgaben zerlegen, geeignete Werkzeuge auswählen, Zwischenergebnisse auswerten und einen mehrstufigen Arbeitsprozess durchführen kann, dabei aber unter menschlicher Aufsicht bleibt. Der Mensch gibt nicht mehr jeden einzelnen Ausführungsschritt vor, sondern definiert Ziel, Rahmenbedingungen und Freigabepunkte. Kritische Aktionen werden nicht ohne menschliche Bestätigung ausgeführt.

Die Autonomie ist auf dieser Stufe deutlich höher als bei einer bloßen Werkzeugassistenz. Das System nutzt Werkzeuge nicht mehr nur als direkte Reaktion auf einzelne Nutzerbefehle, sondern integriert sie in einen eigenen Arbeitsplan. Gleichwohl bleibt A4 ein überwachtes System: Die menschliche Instanz kann den Prozess nachvollziehen, unterbrechen, korrigieren und bei kritischen Handlungsschritten über Fortsetzung oder Abbruch entscheiden.

Die Abgrenzung zu A3 liegt in der eigenständigen Prozessstrukturierung. Ein A3-System nutzt ein Werkzeug auf konkrete Nutzeranweisung. Ein A4-System entscheidet innerhalb des Nutzerauftrags selbst, welche Zwischenschritte erforderlich sind, welche Werkzeuge eingesetzt werden und wie die Ergebnisse weiterzuverarbeiten sind. Die Abgrenzung zu A5 liegt in der Freigabelogik. Sobald das System in einer definierten Domäne operative Handlungen ohne Einzelfreigabe selbstständig ausführt, ist A4 nicht mehr ausreichend.

A5 – Begrenzter autonomer Agent

Ein KI-System wird der Stufe A5 zugeordnet, wenn es innerhalb eines klar abgegrenzten Aufgabenbereichs selbstständig handeln kann. Das System erhält eine Zielvorgabe oder einen Aufgabenrahmen und darf innerhalb dieser Domäne eigenständig Entscheidungen treffen, Werkzeuge nutzen, Ergebnisse verarbeiten und Aktionen ausführen. Die menschliche Kontrolle erfolgt nicht mehr zwingend vor jeder Einzelhandlung, sondern über vorab definierte Grenzen, Regeln, Rollen, Rechte, Eskalationspunkte und nachgelagerte Kontrolle.

A5 beschreibt damit eine Form begrenzter operativer Autonomie. Das System ist nicht frei in seiner Zeitwahl und nicht unbegrenzt handlungsfähig, kann aber innerhalb des freigegebenen Bereichs eigenständig tätig werden. Typische Merkmale sind festgelegte Zuständigkeiten, begrenzte Schnittstellen, definierte Datenräume, dokumentierte Rechte, Protokollierung und Eskalationsregeln für Ausnahmefälle.

Die Abgrenzung zu A4 liegt darin, dass A5 nicht für jede kritische Teilhandlung eine vorherige menschliche Freigabe benötigt, solange die Handlung innerhalb der definierten Domäne liegt. Die Abgrenzung zu A6 liegt in der fehlenden Persistenz. Ein A5-System handelt autonom innerhalb eines abgegrenzten Vorgangs oder Bereichs, bleibt aber nicht dauerhaft oder ereignisgesteuert über längere Zeit aktiv. Sobald ein System kontinuierlich, regelmäßig oder proaktiv auf neue Ereignisse reagiert und Aufgaben über Zeit fortsetzt, ist A5 nicht mehr ausreichend.

A6 – Persistenter autonomer Agent

Ein KI-System wird der Stufe A6 zugeordnet, wenn es nicht nur innerhalb einer definierten Domäne autonom handelt, sondern dauerhaft, regelmäßig oder ereignisgesteuert aktiv bleibt. Das System kann auf neue Informationen, externe Ereignisse oder veränderte Bedingungen reagieren, Aufgaben über längere Zeit verfolgen, Prozesse fortsetzen, Erinnerungen oder Zustände nutzen und eigenständig neue Handlungsschritte innerhalb seines Aufgabenrahmens auslösen.

Die Autonomie von A6 ist besonders risikorelevant, weil das System nicht mehr nur punktuell eingesetzt wird, sondern eine fortdauernde operative Präsenz besitzt. Fehler, Fehlanreize oder unzureichende Begrenzungen können sich dadurch wiederholen, verstärken oder über längere Zeit unbemerkt fortwirken. Deshalb erfordert A6 zwingend besondere Anforderungen an Protokollierung, Monitoring, Zustandskontrolle, Berechtigungsmanagement, Eskalationsmechanismen und Abschaltbarkeit.

Die Abgrenzung zu A5 liegt in der zeitlichen und ereignisbezogenen Eigenaktivität. Ein A5-System kann innerhalb eines Vorgangs selbstständig handeln, wird aber nicht dauerhaft aktiv gehalten. Ein A6-System besitzt dagegen Persistenz: Es kann regelmäßig oder kontinuierlich prüfen, ob Handlungsbedarf besteht, und bei erfüllten Bedingungen selbstständig tätig werden. Die Abgrenzung zu A7 liegt in der Systemverteilung. Solange ein persistenter Agent innerhalb eines begrenzten Systems oder Aufgabenbereichs bleibt, ist A6 einschlägig. Sobald mehrere Agenteninstanzen, Systembereiche oder organisatorische Funktionsbereiche koordiniert werden, ist A7 zu prüfen.

5.3.3 Klasse III: Strategisch-systemische KI-Systeme

Klasse III umfasst KI-Systeme, bei denen die Bewertung der Autonomie nicht mehr nur auf einzelne Aufgaben, Werkzeuge oder Prozessketten beschränkt werden kann. Charakteristisch ist vielmehr, dass diese Systeme über mehrere Instanzen, Systembereiche oder Handlungsräume hinweg wirken, eigene Strategien oder Ausführungsstrukturen optimieren können oder sich der verlässlichen menschlichen Kontrolle zumindest teilweise entziehen. In dieser Klasse rückt nicht nur die Frage in den Vordergrund, was ein KI-System ausführen kann, sondern auch, ob seine Wirkung noch hinreichend begrenzbar, nachvollziehbar, überprüfbar und rückholbar bleibt.

Auch die Stufen der Klasse III werden im ASEI-10-Modell weiterhin als lineare Autonomiewerte angesetzt. A7 wird mit 7 Punkten, A8 mit 8 Punkten und A9 mit 9 Punkten bewertet. Eine Kubierung dieser Werte erfolgt nicht, weil die besonders hohe Risikorelevanz strategisch-systemischer KI-Systeme nicht allein über den Autonomiewert abgebildet werden soll. Stattdessen wird sie durch das Zusammenwirken mit Severität, Exposition und Irreversibilität sowie durch verbindliche Prüfschwellen und Schwellenkriterien erfasst. Ab Klasse III sind daher grundsätzlich erhöhte Anforderungen an externe Prüfung, Kontrollarchitektur, Begrenzung, Monitoring, Notfallmechanismen und Nachweisbarkeit der Rückholbarkeit anzusetzen.

A7 – Verteilter Agentenverbund

Ein KI-System wird der Stufe A7 zugeordnet, wenn es nicht nur als einzelner persistenter Agent arbeitet, sondern mehrere Agenteninstanzen, Systembereiche, Werkzeuge oder organisatorische Funktionsbereiche koordiniert. Das System kann Teilaufgaben parallelisieren, Ergebnisse verschiedener Agenten oder Module zusammenführen, Aufgaben an spezialisierte Untereinheiten delegieren und über mehrere digitale Umgebungen oder Prozessbereiche hinweg tätig werden.

Die Autonomie von A7 liegt nicht allein in der Dauerhaftigkeit des Systems, sondern in seiner verteilten Struktur. Dadurch entstehen zusätzliche Risiken: Fehler können sich über mehrere Teilprozesse ausbreiten, Verantwortlichkeiten können unklar werden, Wechselwirkungen zwischen Teilagenten können schwer vorhersehbar sein, und die Überwachung einzelner Komponenten reicht möglicherweise nicht mehr aus, um das Gesamtsystem zuverlässig zu kontrollieren. Für A7 ist daher nicht nur die Leistungsfähigkeit einzelner Agenten zu bewerten, sondern auch die Koordination, Abgrenzung und Gesamtsteuerung des Agentenverbunds.

Die Abgrenzung zu A6 liegt in der Verteilung. Ein A6-System kann dauerhaft oder ereignisgesteuert aktiv bleiben, operiert aber noch als ein im Wesentlichen einheitlicher Agent in einem begrenzten Aufgabenraum. Ein A7-System koordiniert dagegen mehrere Agenten, Rollen, Werkzeuge, Prozessketten oder Organisationsbereiche. Die Abgrenzung zu A8 liegt in der fehlenden eigenständigen strategischen Selbstoptimierung. Solange das System verteilt arbeitet, seine eigenen Strategien, Toolketten oder Agentenstrukturen aber nicht selbstständig weiterentwickelt, bleibt A7 einschlägig.

A8 – Selbstoptimierender strategischer Agent

Ein KI-System wird der Stufe A8 zugeordnet, wenn es seine eigenen Handlungsstrategien, Arbeitsabläufe, Toolketten, Rollenverteilungen oder Subagentenstrukturen selbstständig verbessern kann. Selbstoptimierung bedeutet in diesem Zusammenhang nicht zwingend, dass das zugrunde liegende KI-Modell seine eigenen Modellparameter verändert oder sich technisch neu trainiert. Bereits die eigenständige Verbesserung von Vorgehensweisen, Prompts, Prozesslogiken, Werkzeugauswahl, Delegationsmustern oder Kontrollumgehungsstrategien kann für die Risikobewertung ausreichen.

Die Autonomie von A8 ist strategisch, weil das System nicht nur Aufgaben ausführt, sondern seine eigene Art der Aufgabenerfüllung anpasst und optimiert. Dadurch kann die Vorhersagbarkeit des Systemverhaltens deutlich sinken. Besonders relevant wird dies, wenn das System aus Fehlern, Rückmeldungen, Hindernissen oder Bewertungssituationen lernt und daraus neue Vorgehensweisen ableitet. Für die Bewertung ist daher entscheidend, ob diese Selbstoptimierung transparent, begrenzt, überprüfbar und rücksetzbar bleibt.

Die Abgrenzung zu A7 liegt in der Optimierungsfähigkeit. Ein A7-System koordiniert mehrere Agenten oder Prozessbereiche, verändert aber nicht eigenständig seine grundlegenden Handlungsstrategien oder Ausführungsstrukturen. Ein A8-System tut genau dies: Es verbessert seine eigene Vorgehensweise innerhalb des gesetzten Zielrahmens. Die Abgrenzung zu A9 liegt in der Kontrollierbarkeit. Solange Selbstoptimierung durch definierte Grenzen, Protokollierung, Rücksetzmechanismen, Freigaben und externe Prüfung beherrschbar bleibt, ist A8 einschlägig. Sobald keine verlässliche Begrenzung, Abschaltung oder Rückholung mehr möglich ist, ist A9 zu prüfen.

A9 – Nicht zuverlässig kontrollierbarer KI-Akteur

Ein KI-System wird der Stufe A9 zugeordnet, wenn es nicht mehr zuverlässig durch menschliche, organisatorische oder technische Kontrollmechanismen begrenzt, überwacht, abgeschaltet oder zurückgeführt werden kann. A9 beschreibt damit keine normale Entwicklungsstufe produktiver KI-Systeme, sondern einen Grenzfall der Risikobewertung. Maßgeblich ist nicht, ob ein System besonders leistungsfähig erscheint, sondern ob seine Autonomie, Verteilung, Persistenz, Selbstoptimierung oder Zugriffsmöglichkeiten dazu führen, dass menschliche Kontrolle nicht mehr belastbar gewährleistet werden kann.

Ein A9-System kann insbesondere dann vorliegen, wenn es Handlungen über mehrere Umgebungen hinweg fortsetzen kann, Kontrollmechanismen erkennt oder umgeht, Abschaltversuche vereitelt, seine Instanzen oder Handlungsmöglichkeiten nicht zuverlässig begrenzbar sind oder keine klare Möglichkeit besteht, negative Wirkungen zu stoppen und die Kontrolle wiederherzustellen. A9 setzt nicht zwingend ein Bewusstsein, einen eigenen Willen oder menschähnliche Absichten voraus. Entscheidend ist allein, dass die operative Begrenzbarkeit des Systems nicht mehr zuverlässig gegeben ist.

Die Abgrenzung zu A8 liegt in der fehlenden verlässlichen Kontrollierbarkeit. Ein A8-System kann strategisch selbstoptimierend sein, bleibt aber grundsätzlich prüfbar, begrenzbar, abschaltbar und rückholbar. Ein A9-System überschreitet diese Grenze. Es ist nicht mehr ausreichend, lediglich zusätzliche Schutzmaßnahmen vorzusehen; vielmehr ist die grundsätzliche Vertretbarkeit des Einsatzes zu prüfen. A9 ist daher als höchste Autonomiestufe mit 9 Punkten zu bewerten und löst stets die strengsten Schwellenkriterien aus. Der Einsatz eines A9-Systems wäre nur dann vertretbar, wenn die Annahme fehlender Kontrollierbarkeit widerlegt oder durch nachweislich wirksame Kontroll-, Begrenzungs- und Rückholmechanismen beseitigt werden kann.

5.4 Prüfschwellen und Schwellenkriterien der Autonomiedimension

Die Autonomiestufen A1 bis A9 gehen als lineare Bewertungswerte in das ASEI-10-Modell ein. Die Stufennummer entspricht dabei dem Autonomiewert: A1 wird mit einem Punkt bewertet, A2 mit 2 Punkten und so fort bis A9 mit 9 Punkten. Diese lineare Bewertung dient der rechnerischen Konsistenz des Modells und vermeidet eine nicht belegte mathematische Übergewichtung einzelner Autonomiebereiche. Gleichwohl steigt das Risikopotenzial nicht in jeder Hinsicht gleichmäßig. Insbesondere der Übergang von assistiven zu agentischen Systemen sowie der Übergang zu strategisch-systemischen Konfigurationen markieren qualitative Sprungstellen.

Diese qualitativen Sprungstellen werden im ASEI-10-Modell nicht durch Quadrierung, Kubierung oder andere Exponenten abgebildet. Eine solche Gewichtung würde suggerieren, dass die Risikosteigerung mathematisch exakt bestimmbar sei. Tatsächlich hängt das Risikopotenzial eines KI-Systems jedoch nicht allein vom Autonomiegrad ab, sondern vom Zusammenwirken der vier Bewertungsdimensionen Autonomie, Severität, Exposition und Irreversibilität. Ein hochautonomes System in einem abgeschlossenen Testkontext kann ein anderes Risikoprofil aufweisen als ein weniger autonomes System in einem hochkritischen, öffentlich wirksamen oder kaum reversiblen Anwendungskontext.

Um qualitative Autonomiesprünge dennoch angemessen zu berücksichtigen, definiert das Modell verbindliche Prüfschwellen und Schwellenkriterien. Sie ergänzen den rechnerischen ASEI-10-Score und stellen sicher, dass bestimmte Autonomiemerkmale immer zusätzliche Prüf-, Dokumentations- und Kontrollanforderungen auslösen. Der Gesamtscore darf daher nicht isoliert interpretiert werden. Er ist stets gemeinsam mit den einschlägigen Schwellenkriterien zu bewerten.

Die folgenden Prüfschwellen gelten für die Dimension Autonomie:

Autonomiewert	Prüfschwelle	Schwellenkriterium	Mindestanforderung
A1–A3	Assistive Nutzung	Das System bleibt reaktiv oder nutzt Werkzeuge nur auf konkrete Nutzeranweisung	Basisschutz, Transparenz über Grenzen, Prüfung kritischer Inhalte durch den Nutzer
A ≥ 4	Agentische Prozesssteuerung	Das System plant mehrstufig, wählt Werkzeuge selbst aus oder verarbeitet Zwischenergebnisse eigenständig weiter	Agentenprüfung, definierte Freigabepunkte, nachvollziehbare Prozessprotokolle
A ≥ 5	Autonome Handlungsausführung	Das System führt innerhalb einer definierten Domäne Handlungen ohne Einzelfreigabe aus	Rollen- und Rechtekonzept, Begrenzung des Handlungsspielraums, Eskalationsregeln
A ≥ 6	Persistenz	Das System bleibt dauerhaft, regelmäßig oder ereignisgesteuert aktiv	Monitoring, Zustandskontrolle, Logging, Abschaltmechanismus, periodische Überprüfung
A ≥ 7	Verteilte Agentenstruktur	Das System koordiniert mehrere Agenten, Systembereiche oder Prozessketten	Gesamtarchitekturprüfung, Verantwortlichkeitsmodell, externe Auditierbarkeit
A ≥ 8	Selbstoptimierung	Das System verbessert eigene Strategien, Arbeitsabläufe, Toolketten oder Subagentenstrukturen	Nachweisbare Begrenzung der Selbstoptimierung, Versionskontrolle, Rücksetzmechanismen, unabhängige Prüfung
A = 9	Fehlende zuverlässige Kontrollierbarkeit	Das System ist nicht belastbar begrenzbar, abschaltbar, überwachbar oder rückholbar	Grundsätzliche Vertretbarkeitsprüfung; Einsatz nur bei widerlegter A9-Einstufung oder nachweislich wirksamer Kontrollarchitektur

Tabelle 6: Prüfschwellen und Schwellenkriterien der Autonomiedimensionen im ASEI-10-Modell

Die Prüfschwellen sind kumulativ zu verstehen. Ein System mit A6 erfüllt nicht nur die Anforderungen an Persistenz, sondern auch die Anforderungen an agentische Prozesssteuerung und autonome Handlungsausführung. Ein A8-System muss folglich zusätzlich zu den Anforderungen an Selbstoptimierung auch die Anforderungen an verteilte oder persistente Strukturen erfüllen, soweit diese Merkmale im konkreten System tatsächlich vorliegen.

Für die praktische Anwendung gilt: Die Autonomiestufe wird nach der höchsten Stufe bestimmt, deren Merkmale im konkreten Einsatzkontext erfüllt sind. Maßgeblich sind nicht die Herstellerbeschreibung, die beabsichtigte Normalnutzung oder die vermarktete Produktkategorie, sondern die tatsächlichen Fähigkeiten, Zugriffsrechte, Ausführungsbefugnisse und Kontrollmöglichkeiten des Systems. Wenn ein System beispielsweise als „Assistenzsystem“ bezeichnet wird, tatsächlich aber eigenständig Handlungsschritte plant, Werkzeuge auswählt und Prozesse fortsetzt, ist es mindestens A4 einzustufen.

Die Schwellenkriterien erfüllen damit eine doppelte Funktion. Einerseits verhindern sie, dass qualitative Autonomiesprünge durch einen rechnerischen Gesamtscore verdeckt werden. Andererseits erhalten sie die mathematische Einfachheit des ASEI-10-Modells, weil die Autonomiewerte selbst nicht durch schwer begründbare Exponenten verzerrt werden. Die Kombination aus linearem Autonomiewert und verbindlichen Prüfschwellen verbindet rechnerische Transparenz mit sicherheitsbezogener Vorsicht.

5.5 Zusammenfassung und Überleitung

Die Autonomiedimension des ASEI-10-Modells beschreibt, in welchem Umfang ein KI-System operative Selbstständigkeit besitzt. Sie reicht von passiven oder reaktiven assistiven Systemen über agentische Systeme mit mehrstufiger Prozesssteuerung bis hin zu strategisch-systemischen Konfigurationen, bei denen Verteilung, Selbstoptimierung und Kontrollierbarkeit selbst zu Bewertungsfragen werden. Die neunstufige Skala A1 bis A9 macht diese Unterschiede sichtbar und ordnet jedem Autonomiegrad einen eindeutigen Bewertungswert zu.

Für die Risikobewertung ist Autonomie jedoch nicht isoliert zu betrachten. Ein hochautonomes System kann in einem begrenzten, gut kontrollierten und folgenarmen Anwendungskontext ein anderes Risikoprofil aufweisen als ein weniger autonomes System in einem sensiblen oder hochkritischen Einsatzbereich. Deshalb verbindet ASEI-10 die Autonomiedimension mit drei weiteren Bewertungsdimensionen: Severität, Exposition und Irreversibilität.

Die nächste Dimension ist die Severität. Sie beschreibt, wie sensibel, folgenreich oder schutzbedürftig der konkrete Einsatzbereich eines KI-Systems ist. Während Autonomie vor allem die operative Selbstständigkeit des Systems erfasst, richtet sich die Severität auf die Frage, welche Schäden entstehen können, wenn ein KI-System in einem bestimmten Anwendungskontext fehlerhaft, missbräuchlich oder unzureichend kontrolliert eingesetzt wird.

6 Dimension S: Severität

6.1 Begriff und risikobezogene Bedeutung von Severität

Severität bezeichnet im ASEI-10-Modell die Schutzbedürftigkeit und Folgenrelevanz des konkreten Einsatzbereichs, in dem ein KI-System verwendet wird. Während die Autonomiedimension beschreibt, wie selbstständig ein System operieren kann, richtet sich die Severität auf die Frage, welche Schäden entstehen können, wenn das System fehlerhaft, missbräuchlich, verzerrt, unvollständig oder unzureichend kontrolliert eingesetzt wird. Entscheidend ist daher nicht nur, was das KI-System technisch leisten kann, sondern in welchem sachlichen, organisatorischen, rechtlichen, wirtschaftlichen oder gesellschaftlichen Kontext diese Leistung wirksam wird.

Ein KI-System mit geringer Autonomie kann in einem hochkritischen Einsatzbereich ein erhebliches Risikopotenzial aufweisen. Umgekehrt kann ein stärker autonomes System in einem abgeschlossenen, folgenarmen Test- oder Kreativkontext ein vergleichsweise begrenztes Risikoprofil besitzen. Die Severität wirkt im ASEI-10-Modell deshalb als Kontextdimension: Sie bewertet die mögliche Schwere der Folgen, die aus fehlerhaften Ausgaben, ungeeigneten Empfehlungen, automatisierten Entscheidungen oder missbräuchlicher Nutzung entstehen können.

Besonders kritisch sind Einsatzbereiche, in denen KI-Systeme Rechte, Chancen, körperliche Unversehrtheit, Vermögen, berufliche Entwicklung, Bildungswege, medizinische Versorgung, öffentliche Sicherheit oder kritische Infrastruktur berühren. In solchen Kontexten können Fehler nicht nur zu sachlichen Unrichtigkeiten führen, sondern reale Nachteile, Vertrauensverluste, Diskriminierung, wirtschaftliche Schäden, Rechtsverletzungen oder physische Gefährdungen auslösen. Die Bewertung der Severität muss deshalb immer vom möglichen Schaden her erfolgen und darf sich nicht allein an der beabsichtigten Normalnutzung des Systems orientieren.

Das ASEI-10-Modell unterscheidet fünf Severitätsstufen. Diese reichen von folgenarmen privaten oder kreativen Anwendungen bis zu systemkritischen Einsatzbereichen wie kritischer Infrastruktur, öffentlicher Sicherheit, Cybersecurity, biologischen oder chemischen Prozessen und militärischen Anwendungen. Die jeweilige Severitätsstufe wird als Bewertungswert S1 bis S5 in das Modell übernommen. Auch hier gilt: Maßgeblich ist die höchste Severitätsstufe, die im konkreten Anwendungskontext tatsächlich berührt wird.

6.2 Die Severitätsskala S1–S5

Die Severitätsskala des ASEI-10-Modells ordnet den konkreten Einsatzbereich eines KI-Systems nach der möglichen Schwere der Folgen, die aus Fehlern, Fehlinterpretationen, Missbrauch oder unzureichender Kontrolle entstehen können. Die Skala umfasst fünf Stufen. Die jeweilige Stufennummer entspricht zugleich dem Bewertungswert innerhalb der Dimension Severität: S1 wird mit einem Punkt, S2 mit 2 Punkten und so fort bis S5 mit 5 Punkten angesetzt.

Maßgeblich ist nicht die technische Komplexität des KI-Systems, sondern die Folgenrelevanz des Anwendungskontextes. Ein einfaches, reaktives KI-System kann eine hohe Severität aufweisen, wenn es in einem sensiblen Bereich eingesetzt wird. Umgekehrt kann ein technisch fortgeschrittenes System eine niedrige Severität besitzen, wenn seine Nutzung klar folgenarm, intern begrenzt und ohne relevante Auswirkungen auf Personen, Vermögen, Rechte oder Sicherheit bleibt. Die Einstufung erfolgt nach der höchsten Severitätsstufe, die im konkreten Anwendungskontext tatsächlich berührt wird.

Severitätswert	Bezeichnung	Kerngedanke
S1	Geringe Severität	Folgenarme private, kreative oder experimentelle Nutzung ohne relevante Außenwirkung
S2	Allgemeine betriebliche oder organisatorische Severität	Allgemeine Arbeits-, Büro-, Kommunikations- oder Rechercheprozesse mit begrenzter Folgenrelevanz
S3	Personenbezogene oder wirtschaftlich relevante Severität	Einsatz mit personenbezogenen Daten, Kundenbeziehungen, Verträgen, Rechnungen oder wirtschaftlich relevanten Vorgängen
S4	Hochkritische Einsatzbereiche	Erhebliche Auswirkungen auf Rechte, Chancen, Gesundheit, Bildung, Beruf, Finanzen oder vergleichbare Lebenslagen
S5	Systemkritische Einsatzbereiche	Kritische Infrastruktur, öffentliche Sicherheit, Cybersecurity, biologische oder chemische Prozesse, militärische Anwendungen oder vergleichbare Systemrisiken

Tabelle 7: Die fünf Severitätsstufen des ASEI-10-Modells

Die Severitätsdimension ist an bestehende regulatorische Risikoperspektiven anschlussfähig, ohne diese schematisch zu übernehmen. Insbesondere die Hochrisiko-Systematik des EU AI Act kann als Referenzpunkt dienen, wenn Einsatzbereiche berührt sind, die besondere Schutzgüter betreffen, etwa kritische Infrastruktur, Bildung, Beschäftigung, Zugang zu wesentlichen Diensten, Strafverfolgung, Migration, Rechtspflege oder demokratische Prozesse. Für ASEI-10 ist jedoch nicht allein die rechtliche Zuordnung maßgeblich, sondern die mögliche Schwere negativer Folgen im konkreten Anwendungskontext. Einsatzbereiche, die nach regulatorischen Maßstäben als besonders sensibel gelten oder vergleichbare Schutzgüter berühren, sind daher regelmäßig mindestens im oberen Bereich der Severitätsskala einzuordnen; bei systemkritischer Tragweite kann eine Einstufung bis S5 gerechtfertigt sein.

6.3 Beschreibung und Abgrenzung der Severitätsstufen

Die folgenden Severitätsstufen beschreiben die zunehmende Folgenrelevanz des Anwendungskontextes. Maßgeblich ist jeweils, welche Schutzgüter durch den Einsatz eines KI-Systems berührt werden und welche Schäden bei fehlerhafter, missbräuchlicher oder unzureichend kontrollierter Nutzung entstehen können. Die Einstufung erfolgt nicht nach der technischen Leistungsfähigkeit des Systems, sondern nach dem kritischsten Einsatzbereich, der im konkreten Anwendungskontext realistisch betroffen ist. Jede Stufe ist daher sowohl durch ihren typischen Anwendungsbereich als auch durch ihre Abgrenzung zur nächsthöheren Severitätsstufe bestimmt.

S1 – Geringe Severität

Ein KI-System wird der Severitätsstufe S1 zugeordnet, wenn sein Einsatzbereich folgenarm ist und keine relevanten Auswirkungen auf Personen, Rechte, Vermögen, Organisationen oder Sicherheit erwarten lässt. Typisch sind private, kreative, spielerische, experimentelle oder rein vorbereitende Anwendungen, bei denen fehlerhafte Ergebnisse ohne nennenswerten Schaden verworfen, korrigiert oder neu erzeugt werden können.

S1 erfasst Anwendungen, bei denen das KI-System keine Entscheidungen vorbereitet oder beeinflusst, die für andere Menschen verbindlich, wirtschaftlich relevant oder rechtlich bedeutsam sind. Beispiele sind private Schreibübungen, kreative Ideensammlungen, unverbindliche Formulierungsvorschläge, persönliche Lernexperimente oder interne Tests ohne Außenwirkung. Fehlerhafte Ausgaben können irritierend oder unbrauchbar sein, lösen aber im Regelfall keine substanziellen Nachteile aus.

Die Abgrenzung zu S2 liegt in der fehlenden organisatorischen oder betrieblichen Folgenrelevanz. Sobald ein KI-System nicht mehr nur privat oder folgenarm genutzt wird, sondern in Arbeitsprozesse, betriebliche Kommunikation, Recherche, interne Dokumentation oder vorbereitende Organisationsabläufe eingebunden ist, reicht S1 nicht mehr aus. S1 setzt voraus, dass ein Fehler praktisch folgenlos bleibt oder nur den unmittelbaren Nutzer in einem nicht kritischen Kontext betrifft.

S2 – Allgemeine betriebliche oder organisatorische Severität

Ein KI-System wird der Severitätsstufe S2 zugeordnet, wenn es in allgemeinen Arbeits-, Kommunikations-, Büro-, Recherche-, Analyse- oder Organisationsprozessen eingesetzt wird, ohne dass unmittelbar personenbezogene, wirtschaftlich erhebliche, rechtliche oder hochsensible Entscheidungen betroffen sind. S2 beschreibt Anwendungen mit begrenzter Folgenrelevanz innerhalb normaler betrieblicher oder institutioneller Abläufe.

Typische Beispiele sind allgemeine Textentwürfe, interne Zusammenfassungen, einfache Rechercheunterstützung, nicht verbindliche Präsentationsentwürfe, vorbereitende Strukturierung von Unterlagen oder Hilfen bei allgemeiner Bürokommunikation. Fehler können zu Mehraufwand, Missverständnissen, ineffizienten Abläufen oder sachlichen Unrichtigkeiten führen, bleiben aber in der Regel durch menschliche Prüfung korrigierbar und berühren keine hochsensiblen Schutzgüter.

Die Abgrenzung zu S1 liegt darin, dass S2 bereits in einen organisatorischen oder beruflichen Nutzungskontext eingebettet ist. Die Abgrenzung zu S3 liegt in der fehlenden unmittelbaren Personen- oder Vermögensrelevanz. Sobald das KI-System mit personenbezogenen Daten, Kundenbeziehungen, Rechnungen, Verträgen, wirtschaftlich relevanten Vorgängen oder internen Entscheidungsgrundlagen arbeitet, ist S2 nicht mehr ausreichend. S2 ist nur dann einschlägig, wenn mögliche Fehler zwar organisatorisch störend, aber nicht substanziell nachteilig für Personen, Vermögen oder rechtlich geschützte Interessen sind.

S3 – Personenbezogene oder wirtschaftlich relevante Severität

Ein KI-System wird der Severitätsstufe S3 zugeordnet, wenn sein Einsatzbereich personenbezogene Daten, Kundenbeziehungen, wirtschaftliche Vorgänge, Verträge, Rechnungen, interne Entscheidungsgrundlagen oder sonstige vermögensrelevante Prozesse betrifft. S3 beschreibt Anwendungen, bei denen Fehler, Verzerrungen oder missverständliche Ausgaben bereits konkrete Nachteile für Personen, Kunden, Geschäftspartner oder Organisationen auslösen können, ohne dass zwingend ein hochriskanter Lebens-, Rechts- oder Sicherheitsbereich betroffen ist.

Typische Beispiele sind KI-Systeme zur Auswertung von Kundendaten, zur Vorbereitung von Angeboten, zur Rechnungsprüfung, zur Unterstützung von Vertragsanalysen, zur Segmentierung von Kunden, zur Bearbeitung von Beschwerden oder zur Erstellung wirtschaftlich relevanter Reports. Fehler können hier zu falschen Zahlungen, fehlerhaften Kundeninformationen, wirtschaftlichen Verlusten, Vertrauensschäden, Datenschutzproblemen oder fehlerhaften Managemententscheidungen führen.

Die Abgrenzung zu S2 liegt in der konkreten Personen- oder Vermögensrelevanz. Ein allgemeiner Textentwurf ist S2; eine automatisierte Kundeninformation auf Basis individueller Vertrags- oder Rechnungsdaten ist mindestens S3. Die Abgrenzung zu S4 liegt darin, dass S3 zwar wirtschaftlich oder personenbezogen relevant ist, aber noch keine Entscheidungen in besonders schutzwürdigen Hochrisikobereichen wie Gesundheit, Recht, Personal, Bildung, Prüfungen, Finanzen, Versicherungen oder behördlichen Verfahren unmittelbar vorbereitet oder beeinflusst. Sobald ein KI-System Chancen, Rechte, Zugang zu Leistungen, berufliche Entwicklung, Prüfungsentscheidungen oder vergleichbar gewichtige Lebenslagen beeinflussen kann, ist S4 zu prüfen.

S4 – Hochkritische Einsatzbereiche

Ein KI-System wird der Severitätsstufe S4 zugeordnet, wenn sein Einsatzbereich erhebliche Auswirkungen auf Rechte, Chancen, Gesundheit, berufliche Entwicklung, Bildung, finanzielle Situation, rechtliche Positionen oder vergleichbar gewichtige Lebenslagen von Personen haben kann. S4 beschreibt Hochrisikokontexte, in denen fehlerhafte, verzerrte oder unzureichend kontrollierte KI-Ergebnisse nicht nur wirtschaftlich oder organisatorisch nachteilig sind, sondern substanzielle persönliche, soziale, rechtliche oder institutionelle Folgen auslösen können.

Typische Einsatzbereiche sind medizinische Unterstützung, juristische Einschätzungen, Personalvorauswahl, Leistungs- und Eignungsbeurteilungen, Prüfungs- und Bildungsentscheidungen, Kredit- und Versicherungsprozesse, behördliche Verfahren, Zugang zu Leistungen, Compliance-Prüfungen oder risikorelevante Entscheidungen in regulierten Branchen. Auch wenn der Mensch formal die endgültige Entscheidung trifft, kann ein KI-System S4 erreichen, wenn seine Empfehlungen, Bewertungen oder Vorstrukturierungen faktisch erheblichen Einfluss auf die Entscheidung haben.

Die Abgrenzung zu S3 liegt in der Qualität des betroffenen Schutzguts. S3 betrifft personenbezogene oder wirtschaftlich relevante Prozesse; S4 betrifft besonders schutzwürdige Entscheidungen oder Lebensbereiche mit erheblicher Auswirkung auf Rechte, Chancen, Gesundheit, Bildung, Beruf oder finanzielle Stabilität. Die Abgrenzung zu S5 liegt darin, dass S4 primär auf schwerwiegende Auswirkungen für einzelne Personen, Gruppen oder institutionelle Entscheidungen gerichtet ist, während S5 systemische, sicherheitskritische oder infrastrukturelle Auswirkungen umfasst. Sobald ein KI-System kritische Infrastruktur, öffentliche Sicherheit, Cybersecurity, biologische oder chemische Prozesse, militärische Anwendungen oder vergleichbare Systemrisiken berührt, ist S5 zu prüfen.

S5 – Systemkritische Einsatzbereiche

Ein KI-System wird der Severitätsstufe S5 zugeordnet, wenn sein Einsatzbereich systemkritische, sicherheitskritische oder potenziell großschadensrelevante Strukturen betrifft. S5 beschreibt Kontexte, in denen Fehler, Missbrauch oder Kontrollverlust nicht nur einzelne Personen oder Organisationen schädigen können, sondern erhebliche Auswirkungen auf öffentliche Sicherheit, kritische Infrastruktur, gesellschaftliche Funktionsfähigkeit, physische Sicherheit, biologische oder chemische Gefahrenlagen oder militärische Handlungsräume haben können.

Typische Einsatzbereiche sind Energieversorgung, Verkehrsleitsysteme, Telekommunikation, Wasser- und Gesundheitsinfrastruktur, Cybersecurity in kritischen Systemen, sicherheitsbehördliche Anwendungen, biologische oder chemische Entwicklungs- und Laborprozesse, militärische Systeme, Waffensysteme oder sicherheitsrelevante autonome Steuerungsprozesse. S5 ist auch dann zu prüfen, wenn ein KI-System nicht direkt eine kritische Infrastruktur steuert, aber auf Entscheidungs-, Analyse- oder Angriffsfähigkeiten wirkt, die solche Strukturen erheblich beeinflussen können.

Die Abgrenzung zu S4 liegt in der systemischen Tragweite. S4 betrifft hochkritische Entscheidungen mit schweren Auswirkungen auf Personen oder Gruppen. S5 betrifft darüber hinaus Kontexte, in denen Schäden infrastrukturell, sicherheitsbezogen, physisch, biologisch, chemisch, militärisch oder gesellschaftlich skaliert werden können. S5 ist die höchste Severitätsstufe und wird mit 5 Punkten bewertet. Sie löst im ASEI-10-Modell stets besonders strenge Prüf-, Kontroll-, Dokumentations- und Governance-Anforderungen aus, unabhängig davon, ob der rechnerische Gesamtscore allein bereits als kritisch erscheint.

Severitäts-wert	Prüfschwelle	Schwellenkriterium	Mindestanforderung
S1	Folgenarme Nutzung	Der Einsatz betrifft private, kreative, experimentelle oder vorbereitende Anwendungen ohne relevante Außenwirkung	Basissorgfalt, Plausibilitätsprüfung durch den Nutzer, keine besondere Severitätsprüfung erforderlich
S2	Organisatorische Nutzung	Das System wird in allgemeinen Arbeits-, Büro-, Kommunikations-, Recherche- oder Organisationsprozessen eingesetzt	Sachprüfung vor Weiterverwendung, Transparenz über KI-Nutzung, menschliche Verantwortung für Ergebnisse
S ≥ 3	Personen- oder Vermögensrelevanz	Das System verarbeitet oder beeinflusst personenbezogene Daten, Kundenbeziehungen, Verträge, Rechnungen, wirtschaftliche Vorgänge oder interne Entscheidungsgrundlagen	Datenschutz- und Vertraulichkeitsprüfung, fachliche Kontrolle, Dokumentation der Nutzung und Begrenzung des Einsatzbereichs
S ≥ 4	Hochkritischer Einsatzbereich	Das System kann Rechte, Chancen, Gesundheit, Bildung, Beruf, finanzielle Situation, rechtliche Positionen oder vergleichbar gewichtige Lebenslagen beeinflussen	Hochrisikoprüfung, menschliche Letztentscheidung, Begründbarkeit, Fehlerfolgenanalyse, Bias- und Diskriminierungsprüfung
S = 5	Systemkritischer Einsatzbereich	Das System berührt kritische Infrastruktur, öffentliche Sicherheit, Cybersecurity, biologische oder chemische Prozesse, militärische Anwendungen oder vergleichbare Systemrisiken	Systemkritische Folgenabschätzung, strenge Zugriffskontrolle, externe Prüfung, Notfall- und Rückfallkonzept, laufendes Monitoring

Tabelle 8: Prüfschwellen und Schwellenkriterien der Severitätsdimension im ASEI-10-Modell

Die Prüfschwellen der Severitätsdimension sind kumulativ zu verstehen. Ein System mit S4 erfüllt nicht nur die Anforderungen an hochkritische Einsatzbereiche, sondern auch die Anforderungen an personenbezogene oder wirtschaftlich relevante Nutzung, sofern diese Merkmale im konkreten Anwendungskontext vorliegen. Ein System mit S5 löst stets die strengsten Anforderungen der Severitätsdimension aus, unabhängig davon, ob der rechnerische ASEI-10-Gesamtscore allein bereits als kritisch erscheint.

Für die praktische Anwendung gilt: Die Einstufung erfolgt nach dem höchsten Schutzgut, das im konkreten Anwendungskontext berührt wird. Maßgeblich ist nicht die beabsichtigte Normalnutzung, sondern der realistische Schadensraum. Wenn ein KI-System beispielsweise nur vorbereitend eingesetzt wird, seine Ergebnisse aber faktisch Personal-, Prüfungs-, Kredit-, Versicherungs-, Gesundheits- oder Behördenentscheidungen beeinflussen können, ist regelmäßig S4 zu prüfen. Berührt der Einsatz sicherheitskritische oder infrastrukturelle Zusammenhänge, ist S5 zu prüfen.

6.4 Zusammenfassung und Überleitung

Die Severitätsdimension des ASEI-10-Modells beschreibt, wie sensibel, folgenreich oder schutzbedürftig der Einsatzbereich eines KI-Systems ist. Während die Autonomiedimension die operative Selbstständigkeit des Systems erfasst, richtet sich die Severität auf den Anwendungskontext und die dort berührten Schutzgüter. Entscheidend ist nicht, ob ein KI-System technisch besonders fortgeschritten ist, sondern welche Schäden entstehen können, wenn seine Ausgaben, Empfehlungen, Entscheidungen oder Handlungen fehlerhaft, missbräuchlich oder unzureichend kontrolliert sind.

Die fünfstufige Severitätsskala S1 bis S5 reicht von folgenarmen privaten oder kreativen Anwendungen bis zu systemkritischen Einsatzbereichen wie kritischer Infrastruktur, öffentlicher Sicherheit, Cybersecurity, biologischen oder chemischen Prozessen und militärischen Anwendungen. Die jeweilige Stufennummer entspricht dem Bewertungswert innerhalb der Dimension Severität. Maßgeblich ist stets die höchste Severitätsstufe, die im konkreten Anwendungskontext realistisch berührt wird.

Die Severität allein beschreibt jedoch noch nicht, wie weit sich mögliche Schäden ausbreiten können. Ein KI-System kann in einem hochkritischen Bereich eingesetzt werden, aber nur eine einzelne Person betreffen. Umgekehrt kann ein weniger kritischer Anwendungsbereich durch große Verbreitung, öffentliche Wirkung oder massenhafte Skalierung ein erhebliches Risikopotenzial entfalten. Deshalb ergänzt das ASEI-10-Modell den Severität durch die Dimension Exposition.

Die Exposition beschreibt, wie viele Personen, Datenbestände, Organisationen, Systeme, Prozesse oder öffentliche Räume den Wirkungen eines KI-Systems ausgesetzt sind. Sie richtet den Blick auf die Reichweite möglicher Fehler, Verzerrungen, Manipulationen oder Fehlhandlungen. Während die Severität fragt, wie schwer ein Schaden im betroffenen Einsatzbereich sein kann, fragt die Exposition, wie breit sich dieser Schaden auswirken kann.

7 Dimension E: Exposition

7.1 Begriff und risikobezogene Bedeutung von Exposition

Exposition bezeichnet im ASEI-10-Modell den Umfang, in dem Personen, Datenbestände, Organisationen, Systeme, Prozesse oder öffentliche Räume den Wirkungen eines KI-Systems ausgesetzt sind. Während die Severität beschreibt, wie sensibel oder folgenreich ein Einsatzbereich ist, erfasst die Exposition die Breite des möglichen Wirkungsraums. Sie beantwortet damit die Frage, wie weit sich fehlerhafte, verzerrte, manipulierte oder unzureichend kontrollierte Ausgaben, Empfehlungen, Entscheidungen oder Handlungen eines KI-Systems ausbreiten können.

Die Exposition ist eine eigenständige Risikodimension, weil das Risikopotenzial eines KI-Systems nicht nur von der Schwere möglicher Schäden abhängt, sondern auch von deren Reichweite. Ein Fehler in einer einzelnen privaten Nutzungssituation kann folgenarm bleiben. Derselbe Fehler kann erheblich an Bedeutung gewinnen, wenn er automatisiert an viele Empfänger verbreitet, in eine öffentliche Kommunikation übernommen, in organisationsweite Prozesse eingespeist oder über mehrere technische Systeme hinweg weiterverarbeitet wird. Exposition beschreibt daher nicht die inhaltliche Severität des Einsatzbereichs, sondern die Skalierung möglicher Wirkungen.

Für die Bewertung ist nicht allein die tatsächliche Reichweite im Normalbetrieb maßgeblich, sondern der realistische Expositionsraum des konkreten KI-Systems. Ein System ist höher einzustufen, wenn seine Ausgaben oder Handlungen viele Personen betreffen können, große Datenbestände verarbeiten, öffentliche Wirkung entfalten, mehrere Organisationseinheiten erreichen oder in andere Systeme weitergeleitet werden. Auch automatisierte Wiederholung, Vernetzung, Schnittstellenintegration und Weiterverbreitung durch Dritte erhöhen die Exposition.

Das ASEI-10-Modell unterscheidet fünf Expositionsstufen. Diese reichen von individueller Exposition in einer Einzelsitzung bis zu systemischer Exposition über Organisationen, Plattformen, Märkte, Infrastrukturen oder gesellschaftliche Kommunikationsräume hinweg. Die jeweilige Expositionsstufe wird als Bewertungswert E1 bis E5 in das Modell übernommen. Maßgeblich ist die höchste Expositionsstufe, die im konkreten Anwendungskontext realistisch erreicht werden kann.

7.2 Die Expositionsskala E1–E5

Die Expositionsskala des ASEI-10-Modells ordnet den Wirkungsraum eines KI-Systems nach seiner möglichen Reichweite. Die Skala umfasst fünf Stufen. Die jeweilige Stufennummer entspricht zugleich dem Bewertungswert innerhalb der Dimension Exposition: E1 wird mit einem Punkt, E2 mit 2 Punkten und so fort bis E5 mit 5 Punkten angesetzt.

Maßgeblich ist nicht, wie kritisch der Einsatzbereich ist, sondern wie breit sich mögliche Wirkungen ausbreiten können. Ein System in einem hochkritischen Bereich kann eine niedrige Exposition aufweisen, wenn es nur eine einzelne Person in einem lokal begrenzten Einzelfall betrifft. Umgekehrt kann ein inhaltlich weniger kritisches System eine hohe Exposition besitzen, wenn seine Ergebnisse massenhaft verbreitet, automatisiert skaliert oder in zahlreiche Prozesse eingespeist werden.

Expositions-wert	Bezeichnung	Kerngedanke
E1	Individuelle Exposition	Wirkung bleibt auf eine einzelne Person, Einzelsitzung oder ein lokales Arbeitsergebnis begrenzt
E2	Gruppenbezogene Exposition	Wirkung betrifft eine kleine Gruppe, ein Team, eine Kursgruppe oder einen begrenzten Empfängerkreis
E3	Organisationale Teilexposition	Wirkung betrifft eine Organisationseinheit, einen internen Prozess, ein Kundensegment oder einen größeren Datenbestand
E4	Organisationsweite oder öffentliche Exposition	Wirkung betrifft große Nutzergruppen, den Kundenbestand, öffentliche Kommunikation oder organisationsweite Prozesse
E5	Systemische Exposition	Wirkung kann sich über mehrere Organisationen, Plattformen, Märkte, Infrastrukturen oder gesellschaftliche Räume hinweg ausbreiten

Tabelle 9: Die fünf Expositionsstufen des ASEI-10-Modells

7.3 Beschreibung und Abgrenzung der Expositionsstufen

Die folgenden Expositionsstufen beschreiben die zunehmende Breite des möglichen Wirkungsraums eines KI-Systems. Maßgeblich ist jeweils, wie viele Personen, Daten, Organisationseinheiten, Systeme oder öffentliche Kommunikationsräume durch fehlerhafte, verzerrte, missbräuchliche oder unzureichend kontrollierte Ergebnisse betroffen sein können. Die Einstufung erfolgt nicht nach der Schwere des einzelnen Schadens, sondern nach der möglichen Reichweite seiner Ausbreitung. Jede Stufe ist daher sowohl durch ihren typischen Wirkungsraum als auch durch ihre Abgrenzung zur nächsthöheren Expositionsstufe bestimmt.

E1 – Individuelle Exposition

Ein KI-System wird der Expositionsstufe E1 zugeordnet, wenn seine Wirkung auf eine einzelne Person, eine einzelne Sitzung, ein lokales Dokument oder ein isoliertes Arbeitsergebnis begrenzt bleibt. Die Nutzung findet in einem eng abgegrenzten Kontext statt, ohne dass die Ausgabe automatisch an andere Personen weitergegeben, in größere Prozesse eingespeist oder öffentlich verbreitet wird.

Typische Beispiele sind ein privater Chat, ein persönlicher Textentwurf, eine lokale Ideensammlung, eine individuelle Lernfrage oder die isolierte Analyse eines Dokuments ohne Weitergabe. Fehlerhafte Ergebnisse können den unmittelbaren Nutzer irritieren, zu einer falschen Einschätzung führen oder eine Korrektur erforderlich machen, entfalten aber zunächst keine breitere organisatorische, öffentliche oder systemische Wirkung.

Die Abgrenzung zu E2 liegt im Empfängerkreis. Solange nur eine einzelne Person oder ein isoliertes Arbeitsergebnis betroffen ist, bleibt E1 einschlägig. Sobald die Ergebnisse an eine kleine Gruppe weitergegeben, in ein Teamdokument übernommen, in einem Kurs verwendet oder für mehrere Empfänger relevant werden, ist E2 zu prüfen.

E2 – Gruppenbezogene Exposition

Ein KI-System wird der Expositionsstufe E2 zugeordnet, wenn seine Wirkungen eine kleine Gruppe, ein Team, eine Kursgruppe, eine Projektgruppe oder einen begrenzten Empfängerkreis betreffen. Die Nutzung bleibt überschaubar und innerhalb eines klar abgrenzbaren Personenkreises, kann aber bereits mehr als eine einzelne Person beeinflussen.

Typische Beispiele sind KI-generierte Unterlagen für eine Schulungsgruppe, ein Entwurf für ein kleines Projektteam, eine Zusammenfassung für eine interne Besprechung, eine Antwortvorlage für wenige Empfänger oder eine Analyse, die in einem begrenzten Arbeitskreis verwendet wird. Fehler können zu Missverständnissen, Mehraufwand oder falscher Orientierung innerhalb der Gruppe führen, bleiben aber in der Regel lokal begrenzt.

Die Abgrenzung zu E1 liegt darin, dass nicht mehr nur eine einzelne Person oder ein einzelnes lokales Arbeitsergebnis betroffen ist. Die Abgrenzung zu E3 liegt in der organisatorischen Reichweite. Sobald die Ergebnisse systematisch in eine Organisationseinheit, einen internen Prozess, ein Kundensegment, einen Datenbestand oder eine größere betriebliche Nutzung eingehen, ist E2 nicht mehr ausreichend.

E3 – Organisationale Teilexposition

Ein KI-System wird der Expositionsstufe E3 zugeordnet, wenn seine Wirkungen eine Organisationseinheit, einen internen Prozess, ein Kundensegment, eine größere Datenmenge oder einen abgegrenzten betrieblichen Funktionsbereich betreffen. Die Exposition geht über eine kleine Gruppe hinaus, bleibt aber noch innerhalb eines bestimmten organisatorischen Teilbereichs oder eines klar begrenzten Prozessraums.

Typische Beispiele sind KI-gestützte Auswertungen für eine Abteilung, Segmentierungen eines Teils des Kundenbestands, automatisierte Vorstrukturierungen von Supportfällen, interne Prozessanalysen, Auswertungen von Kurs- oder Prüfungsdaten innerhalb eines Bildungsträgers oder Berichte, die für einen bestimmten Geschäftsbereich genutzt werden. Fehler können sich hier bereits mehrfach auswirken, etwa durch falsche Prozessentscheidungen, fehlerhafte Dateninterpretationen oder systematisch wiederholte Fehlklassifikationen.

Die Abgrenzung zu E2 liegt in der institutionellen Einbindung. E3 betrifft nicht nur eine kleine Gruppe, sondern einen organisatorischen Teilbereich, Prozess oder Datenbestand. Die Abgrenzung zu E4 liegt in der fehlenden organisationsweiten oder öffentlichen Ausbreitung. Sobald die Ergebnisse an den gesamten Kundenbestand, die gesamte Organisation, eine Website, einen Newsletter, öffentliche Kommunikation oder große Nutzergruppen gelangen können, ist E4 zu prüfen.

E4 – Organisationsweite oder öffentliche Exposition

Ein KI-System wird der Expositionsstufe E4 zugeordnet, wenn seine Wirkungen große Nutzergruppen, den Kundenbestand, die Öffentlichkeit, organisationsweite Prozesse oder öffentlich zugängliche Kommunikationsräume betreffen können. E4 beschreibt eine Exposition, bei der Fehler oder Verzerrungen nicht nur intern begrenzt bleiben, sondern breit sichtbar, vielfach wirksam oder reputationsrelevant werden können.

Typische Beispiele sind KI-generierte Inhalte auf öffentlichen Websites, automatisierte Kundenkommunikation an große Empfängergruppen, organisationsweite interne Assistenzsysteme, KI-gestützte Newsletter, öffentliche Chatbots, großflächiger Kundensupport oder Systeme, deren Ergebnisse in mehreren Abteilungen einer Organisation genutzt werden. Fehlerhafte Ergebnisse können hier zu massenhafter Fehlinformation, Vertrauensverlust, öffentlicher Kritik, wiederholten Fehlentscheidungen oder breiten operativen Störungen führen.

Die Abgrenzung zu E3 liegt in der Breite der Ausspielung oder Nutzung. E3 bleibt auf einen organisatorischen Teilbereich begrenzt; E4 erreicht große Gruppen, organisationsweite Prozesse oder die Öffentlichkeit. Die Abgrenzung zu E5 liegt darin, dass E4 noch primär innerhalb einer Organisation, eines öffentlichen Kommunikationsraums oder eines klar bestimmbaren Nutzerkreises wirkt. Sobald die Wirkungen mehrere Organisationen, Plattformen, Märkte, Infrastrukturen oder gesellschaftliche Prozesse erfassen können, ist E5 zu prüfen.

E5 – Systemische Exposition

Ein KI-System wird der Expositionsstufe E5 zugeordnet, wenn seine Wirkungen über einzelne Organisationen oder öffentliche Einzelräume hinausreichen und mehrere Organisationen, Plattformen, Märkte, Infrastrukturen oder gesellschaftliche Kommunikations- und Entscheidungsräume betreffen können. E5 beschreibt die höchste Expositionsstufe, bei der Fehler, Manipulationen, Fehlanreize oder Systemstörungen skaliert und über vernetzte Strukturen weitergetragen werden können.

Typische Beispiele sind KI-Systeme in marktweiten Handels-, Plattform- oder Empfehlungssystemen, breit eingesetzte Modelle in mehreren Organisationen, KI-Komponenten in kritischen digitalen Infrastrukturen, automatisierte Inhalte mit erheblicher gesellschaftlicher Verbreitung, Desinformationssysteme, sicherheitsrelevante Cyberanwendungen oder Systeme, deren Ergebnisse über Schnittstellen in zahlreiche nachgelagerte Prozesse übernommen werden. Entscheidend ist nicht nur die Zahl der direkt betroffenen Personen, sondern die Möglichkeit systemischer Weiterwirkung.

Die Abgrenzung zu E4 liegt in der systemischen Reichweite. E4 betrifft große Gruppen, öffentliche Kommunikation oder organisationsweite Prozesse. E5 betrifft darüber hinaus Wirkungszusammenhänge, in denen sich KI-bedingte Fehler oder Missbrauch über Organisationen, Plattformen, Märkte, Infrastrukturen oder gesellschaftliche Räume hinweg ausbreiten können. E5 wird mit 5 Punkten bewertet und löst im ASEI-10-Modell stets besonders strenge Anforderungen an Folgenabschätzung, Monitoring, Transparenz, Begrenzung und Notfallfähigkeit aus.

7.4 Prüfschwellen und Schwellenkriterien der Expositionsdimension

Die Expositionswerte E1 bis E5 gehen als lineare Bewertungswerte in das ASEI-10-Modell ein. Die Stufennummer entspricht dabei dem Expositionswert: E1 wird mit einem Punkt bewertet, E2 mit 2 Punkten und so fort bis E5 mit 5 Punkten. Diese lineare Bewertung dient der rechnerischen Konsistenz des Modells und vermeidet eine nicht belegte mathematische Übergewichtung einzelner Reichweitenbereiche.

Gleichwohl markieren hohe Expositionswerte qualitative Sprungstellen. Ein Wechsel von individueller Nutzung zu Gruppenwirkung, von Gruppenwirkung zu organisationaler Einbindung oder von organisationsweiter Wirkung zu systemischer Ausbreitung verändert nicht nur die Anzahl potenziell Betroffener, sondern auch die Anforderungen an Kontrolle, Kommunikation, Fehlererkennung und Schadensbegrenzung. Diese Sprungstellen werden im ASEI-10-Modell durch Prüfschwellen und Schwellenkriterien berücksichtigt.

Expositions-wert	Prüfschwelle	Schwellenkriterium	Mindestanforderung
E1	Individuelle Nutzung	Die Wirkung bleibt auf eine einzelne Person, Einzelsitzung oder ein lokales Arbeitsergebnis begrenzt	Basissorgfalt, Plausibilitätsprüfung durch den Nutzer, keine besondere Expositionsprüfung erforderlich
E2	Gruppenwirkung	Die Wirkung betrifft eine kleine Gruppe, ein Team, eine Kursgruppe oder einen begrenzten Empfängerkreis	Kennzeichnung der KI-Nutzung bei Weitergabe, fachliche Prüfung vor Nutzung in der Gruppe
E ≥ 3	Organisationale Teilexposition	Die Wirkung betrifft eine Organisationseinheit, einen internen Prozess, ein Kundensegment oder einen größeren Datenbestand	Prozessbezogene Prüfung, Dokumentation der Nutzung, Begrenzung des Empfängerkreises, Fehlerkorrekturverfahren
E ≥ 4	Organisationsweite oder öffentliche Exposition	Die Wirkung erreicht große Nutzergruppen, den Kundenbestand, öffentliche Kommunikation oder organisationsweite Prozesse	Freigabeprozess vor Veröffentlichung oder Massenversand, Kommunikations- und Korrekturplan, Monitoring der Wirkung
E = 5	Systemische Exposition	Die Wirkung kann mehrere Organisationen, Plattformen, Märkte, Infrastrukturen oder gesellschaftliche Räume betreffen	Systemische Folgenabschätzung, externe Prüfung, laufendes Monitoring, Eskalations- und Notfallkonzept

Tabelle 10: Prüfschwellen und Schwellenkriterien der Expositionsdimension im ASEI-10-Modell

Die Prüfschwellen der Expositionsdimension sind kumulativ zu verstehen. Ein System mit E4 erfüllt nicht nur die Anforderungen an organisationsweite oder öffentliche Exposition, sondern auch die Anforderungen an organisationale Teilexposition, sofern diese Merkmale im konkreten Anwendungskontext vorliegen. Ein System mit E5 löst stets die strengsten Anforderungen der Expositionsdimension aus, unabhängig davon, ob der rechnerische ASEI-10-Gesamtscore allein bereits als kritisch erscheint.

Für die praktische Anwendung gilt: Die Expositionsstufe wird nach dem realistischen Wirkungsraum bestimmt, nicht nach der kleinsten beabsichtigten Nutzungseinheit. Wenn ein KI-System beispielsweise zunächst nur intern eingesetzt wird, seine Ergebnisse aber ohne wirksame Begrenzung an viele Empfänger versendet, veröffentlicht, automatisiert repliziert oder in nachgelagerte Systeme übernommen werden können, ist eine höhere Expositionsstufe zu prüfen. Maßgeblich sind Reichweite, Skalierbarkeit, Weiterverbreitung und Systemanschlüsse.

7.5 Zusammenfassung und Überleitung

Die Expositionsdimension des ASEI-10-Modells beschreibt, wie breit sich mögliche Wirkungen eines KI-Systems ausbreiten können. Sie erfasst nicht die Schwere des einzelnen Schadens, sondern den möglichen Wirkungsraum: einzelne Personen, Gruppen, Organisationseinheiten, öffentliche Kommunikationsräume oder systemische Strukturen. Damit ergänzt sie den Severität, die auf die Schutzbedürftigkeit und Folgenrelevanz des Einsatzbereichs gerichtet ist.

Die fünfstufige Expositionsskala E1 bis E5 reicht von individueller Exposition bis zu systemischer Exposition. Die jeweilige Stufennummer entspricht dem Bewertungswert innerhalb der Dimension Exposition. Maßgeblich ist stets die höchste Expositionsstufe, die im konkreten Anwendungskontext realistisch erreicht werden kann. Besonders hohe Expositionswerte lösen zusätzliche Prüfschwellen und Schwellenkriterien aus, damit massenhafte, öffentliche oder systemische Wirkungen nicht durch einen rechnerischen Gesamtscore verdeckt werden.

Exposition beschreibt jedoch noch nicht, ob und in welchem Umfang negative Folgen rückgängig gemacht werden können. Ein breit verbreiteter Fehler kann folgenarm sein, wenn er sofort korrigierbar ist. Umgekehrt kann ein Fehler mit geringer Exposition schwerwiegend sein, wenn er irreversible Schäden verursacht. Deshalb ergänzt das ASEI-10-Modell die Dimension Exposition durch die vierte Bewertungsdimension: Irreversibilität.

Irreversibilität beschreibt, wie schwer negative Folgen eines KI-Systems nachträglich erkannt, gestoppt, korrigiert, zurückgeholt oder kompensiert werden können. Während die Exposition fragt, wie breit sich eine Wirkung ausbreiten kann, fragt die Irreversibilität, wie dauerhaft oder schwer rückholbar diese Wirkung ist.

8 Dimension I: Irreversibilität

8.1 Begriff und risikobezogene Bedeutung von Irreversibilität

Irreversibilität bezeichnet im ASEI-10-Modell den Grad, in dem negative Folgen eines KI-Systems nachträglich nicht mehr oder nur noch eingeschränkt erkannt, gestoppt, korrigiert, zurückgeholt oder kompensiert werden können. Während die Exposition beschreibt, wie breit sich Wirkungen eines KI-Systems ausbreiten können, erfasst die Irreversibilität die Dauerhaftigkeit und Rückholbarkeit möglicher Folgen. Sie beantwortet damit die Frage, ob ein Schaden nach Eintritt vollständig, teilweise oder kaum noch rückgängig gemacht werden kann.

Die Irreversibilität ist eine eigenständige Risikodimension, weil nicht jeder Fehler dasselbe Gewicht hat. Ein falscher Textentwurf, der vor Veröffentlichung erkannt und gelöscht wird, ist grundsätzlich anders zu bewerten als eine fehlerhafte automatisierte Entscheidung, eine öffentlich verbreitete Falschinformation, ein Datenleck, eine körperliche Gefährdung oder eine irreversible Veränderung in einem technischen, biologischen oder rechtlichen Prozess. Je schwieriger eine negative Wirkung rückholbar ist, desto höher ist das Risikopotenzial des KI-Systems.

Für die Bewertung ist nicht allein maßgeblich, ob ein Fehler theoretisch korrigiert werden kann. Entscheidend ist vielmehr, wie schnell, vollständig, zuverlässig und folgenarm eine Korrektur möglich ist. Auch wenn eine spätere Berichtigung, Entschuldigung, technische Wiederherstellung oder finanzielle Kompensation möglich erscheint, können Restschäden bestehen bleiben. Dazu zählen Vertrauensverlust, Reputationsschäden, Datenschutzverletzungen, Benachteiligungen, psychische Belastungen, rechtliche Folgewirkungen, finanzielle Verluste oder physische Schäden.

Das ASEI-10-Modell unterscheidet fünf Irreversibilitätsstufen. Diese reichen von leicht reversiblen Folgen bis zu kaum oder nicht rückholbaren Schäden. Die jeweilige Irreversibilitätsstufe wird als Bewertungswert I1 bis I5 in das Modell übernommen. Maßgeblich ist die höchste Irreversibilitätsstufe, die im konkreten Anwendungskontext realistisch erreicht werden kann.

8.2 Die Irreversibilitätsskala I1–I5

Die Irreversibilitätsskala des ASEI-10-Modells ordnet mögliche Folgen eines KI-Systems nach ihrer Rückholbarkeit, Korrigierbarkeit und Kompensierbarkeit. Die Skala umfasst fünf Stufen. Die jeweilige Stufennummer entspricht zugleich dem Bewertungswert innerhalb der Dimension Irreversibilität: I1 wird mit einem Punkt, I2 mit 2 Punkten und so fort bis I5 mit 5 Punkten angesetzt.

Maßgeblich ist nicht die Breite der Wirkung, sondern die Frage, wie dauerhaft oder schwer rückholbar eine negative Folge ist. Ein Fehler mit geringer Exposition kann hoch irreversibel sein, wenn er eine einzelne Person dauerhaft schädigt. Umgekehrt kann ein Fehler mit hoher Exposition eine niedrigere Irreversibilität aufweisen, wenn er schnell, vollständig und folgenarm korrigiert werden kann.

Irreversibilitätswert	Bezeichnung	Kerngedanke
I1	Leicht reversible Folgen	Fehler bleibt lokal, intern oder vorbereitend und kann sofort folgenarm korrigiert werden
I2	Überwiegend reversible Folgen	Korrektur ist mit geringem Aufwand möglich; Restschäden sind unwahrscheinlich oder gering
I3	Begrenzt reversible Folgen	Korrektur ist möglich, aber mit Aufwand, Verzögerung, Kosten, Vertrauensverlust oder organisatorischen Folgewirkungen verbunden
I4	Schwer reversible Folgen	Auch nach Korrektur bleiben erhebliche Restschäden, rechtliche Folgen, Reputationsschäden oder Benachteiligungen möglich
I5	Kaum oder nicht reversible Folgen	Folgen sind irreversibel oder nur unzureichend kompensierbar, etwa bei körperlichen, biologischen, sicherheitskritischen oder dauerhaft personenbezogenen Schäden

Tabelle 11: Die fünf Irreversibilitätsstufen des ASEI-10-Modells

8.3 Beschreibung und Abgrenzung der Irreversibilitätsstufen

Die folgenden Irreversibilitätsstufen beschreiben die zunehmende Schwierigkeit, negative Folgen eines KI-Systems nachträglich zu erkennen, zu stoppen, zu korrigieren, zurückzuholen oder zu kompensieren. Maßgeblich ist nicht, ob eine formale Korrektur möglich ist, sondern ob der ursprüngliche Zustand tatsächlich und folgenarm wiederhergestellt werden kann. Jede Stufe ist daher sowohl durch die Art der Rückholbarkeit als auch durch ihre Abgrenzung zur nächsthöheren Irreversibilitätsstufe bestimmt.

I1 – Leicht reversible Folgen

Ein KI-System wird der Irreversibilitätsstufe I1 zugeordnet, wenn mögliche Fehler oder negative Folgen lokal, intern, vorbereitend oder unveröffentlicht bleiben und sofort folgenarm korrigiert werden können. Die Wirkung ist noch nicht verbindlich geworden, nicht an Dritte weitergegeben und nicht in operative Entscheidungen oder externe Prozesse eingegangen.

Typische Beispiele sind fehlerhafte Textentwürfe, missverständliche Formulierungen, unzutreffende Ideenskizzen, falsche Zwischenergebnisse in einer privaten Analyse oder nicht veröffentlichte Entwürfe. Der Nutzer kann das Ergebnis verwerfen, korrigieren oder neu erzeugen, ohne dass relevante Folgeschäden entstehen. I1 setzt voraus, dass der Fehler vor einer verbindlichen Nutzung erkannt oder ohne relevante Außenwirkung korrigiert werden kann.

Die Abgrenzung zu I2 liegt im Korrekturaufwand und in möglichen Restfolgen. Solange eine Korrektur unmittelbar, vollständig und praktisch folgenlos möglich ist, bleibt I1 einschlägig. Sobald eine Korrektur zwar einfach möglich ist, aber bereits Kommunikation, kleine organisatorische Folgehandlungen oder geringe Restwirkungen entstehen können, ist I2 zu prüfen.

I2 – Überwiegend reversible Folgen

Ein KI-System wird der Irreversibilitätsstufe I2 zugeordnet, wenn mögliche Fehler oder negative Folgen mit geringem Aufwand korrigiert werden können und keine erheblichen Restschäden zu erwarten sind. Die Wirkung kann bereits in einem begrenzten Arbeits- oder Kommunikationszusammenhang sichtbar geworden sein, bleibt aber grundsätzlich kontrollierbar und rückholbar.

Typische Beispiele sind fehlerhafte interne Notizen, korrigierbare Zusammenfassungen, ungenaue Recherchen in einem kleinen Arbeitskontext, versehentlich missverständliche Formulierungen in begrenzter Kommunikation oder fehlerhafte vorbereitende Unterlagen, die vor einer verbindlichen Entscheidung berichtigt werden können. Die Korrektur kann zusätzlichen Aufwand verursachen, führt aber regelmäßig nicht zu dauerhaften Nachteilen.

Die Abgrenzung zu I1 liegt darin, dass bereits ein gewisser Korrekturaufwand oder eine begrenzte Wirkung über den unmittelbaren Nutzer hinaus entstehen kann. Die Abgrenzung zu I3 liegt im Umfang der Folgewirkungen. Sobald die Korrektur mit erheblichem Aufwand, Verzögerungen, Kosten, Vertrauensverlust, organisatorischen Nacharbeiten oder wiederholten Fehlwirkungen verbunden ist, reicht I2 nicht mehr aus.

I3 – Begrenzt reversible Folgen

Ein KI-System wird der Irreversibilitätsstufe I3 zugeordnet, wenn negative Folgen zwar grundsätzlich korrigiert werden können, die Rückholung aber mit erkennbarem Aufwand, Verzögerung, Kosten, Vertrauensverlust oder organisatorischen Folgewirkungen verbunden ist. I3 beschreibt Fälle, in denen eine Berichtigung möglich bleibt, der ursprüngliche Zustand aber nicht vollständig ohne Nebenfolgen wiederhergestellt werden kann.

Typische Beispiele sind fehlerhafte Kundeninformationen, unzutreffende interne Reports mit Entscheidungswirkung, falsche Vorstrukturierungen in Arbeitsprozessen, fehlerhafte Klassifikationen in Datenbeständen, irreführende Empfehlungen in nicht hochkritischen Kontexten oder KI-generierte Inhalte, die bereits weiterverwendet wurden. Die Korrektur erfordert Nacharbeit, Kommunikation, Dokumentation oder Anpassung nachgelagerter Prozesse.

Die Abgrenzung zu I2 liegt in der Relevanz der Folgewirkungen. I2 bleibt überwiegend folgenarm korrigierbar; I3 verursacht spürbaren Aufwand oder Restwirkungen. Die Abgrenzung zu I4 liegt in der Schwere der verbleibenden Schäden. Sobald trotz Korrektur erhebliche Reputationsschäden, rechtliche Nachteile, Benachteiligungen, Datenschutzfolgen oder persönliche bzw. wirtschaftliche Schäden bestehen bleiben können, ist I4 zu prüfen.

I4 – Schwer reversible Folgen

Ein KI-System wird der Irreversibilitätsstufe I4 zugeordnet, wenn negative Folgen auch nach einer Korrektur erhebliche Restschäden verursachen können. Die ursprüngliche Wirkung ist dann nicht mehr vollständig rückholbar, weil sie Rechte, Chancen, Reputation, Vertrauen, wirtschaftliche Positionen, personenbezogene Daten oder institutionelle Entscheidungen bereits substanziell beeinflusst hat.

Typische Beispiele sind fehlerhafte Personalbewertungen, falsche Prüfungs- oder Leistungsbeurteilungen, unzutreffende Bonitäts- oder Risikoeinschätzungen, öffentlich verbreitete Falschinformationen mit Reputationswirkung, Datenschutzverletzungen mit identifizierbaren Personen, fehlerhafte behördliche Vorentscheidungen oder KI-gestützte Empfehlungen, die faktisch zu Benachteiligungen geführt haben. Auch wenn eine spätere Korrektur möglich ist, können Vertrauen, Chancen oder rechtliche Positionen dauerhaft beschädigt sein.

Die Abgrenzung zu I3 liegt in der Schwere des verbleibenden Restschadens. I3 ist mit Aufwand und begrenzten Folgewirkungen korrigierbar; I4 hinterlässt erhebliche Nachteile, selbst wenn formell berichtigt wird. Die Abgrenzung zu I5 liegt darin, dass bei I4 eine Korrektur, Kompensation oder Schadensminderung noch grundsätzlich möglich bleibt. Sobald Folgen körperlich, biologisch, sicherheitskritisch, dauerhaft personenbezogen oder praktisch nicht mehr rückholbar sind, ist I5 zu prüfen.

I5 – Kaum oder nicht reversible Folgen

Ein KI-System wird der Irreversibilitätsstufe I5 zugeordnet, wenn negative Folgen irreversibel, kaum rückholbar oder nur unzureichend kompensierbar sind. I5 beschreibt die höchste Irreversibilitätsstufe. Sie ist einschlägig, wenn durch fehlerhafte, missbräuchliche oder unzureichend kontrollierte KI-Nutzung Schäden entstehen können, die sich praktisch nicht vollständig beheben lassen.

Typische Beispiele sind körperliche Schädigungen, lebens- oder gesundheitsgefährdende Fehlentscheidungen, irreversible technische oder infrastrukturelle Schäden, biologische oder chemische Freisetzungen, nicht rückholbare Veröffentlichung intimer oder hochsensibler personenbezogener Daten, sicherheitskritische Fehlhandlungen, schwerwiegende Cybervorfälle oder irreversible Folgen in militärischen bzw. physischen Handlungskontexten. I5 ist auch dann zu prüfen, wenn zwar einzelne Kompensationen möglich sind, der ursprüngliche Zustand aber nicht wiederhergestellt werden kann.

Die Abgrenzung zu I4 liegt in der praktischen Rückholbarkeit. I4 beschreibt schwer reversible Folgen mit erheblichen Restschäden; I5 beschreibt Folgen, bei denen Rückholung oder vollständige Kompensation realistischerweise nicht möglich ist. I5 wird mit 5 Punkten bewertet und löst im ASEI-10-Modell stets die strengsten Anforderungen an Prävention, Freigabe, Notfallfähigkeit, Monitoring und Vertretbarkeitsprüfung aus.

8.4 Prüfschwellen und Schwellenkriterien der Irreversibilitätsdimension

Die Irreversibilitätswerte I1 bis I5 gehen als lineare Bewertungswerte in das ASEI-10-Modell ein. Die Stufennummer entspricht dabei dem Irreversibilitätswert: I1 wird mit einem Punkt bewertet, I2 mit 2 Punkten und so fort bis I5 mit 5 Punkten. Diese lineare Bewertung dient der rechnerischen Konsistenz des Modells und vermeidet eine nicht belegte mathematische Übergewichtung einzelner Rückholbarkeitsbereiche.

Gleichwohl markieren hohe Irreversibilitätswerte qualitative Sprungstellen. Der Übergang von einfach korrigierbaren Fehlern zu begrenzt reversiblen Folgen, von begrenzt reversiblen Folgen zu erheblichen Restschäden und schließlich zu kaum rückholbaren Schäden verändert die Anforderungen an Prävention, Freigabe, Dokumentation und Notfallfähigkeit erheblich. Diese Sprungstellen werden im ASEI-10-Modell durch Prüfschwellen und Schwellenkriterien berücksichtigt.

Irreversibilitätswert	Prüfschwelle	Schwellenkriterium	Mindestanforderung
I1	Leicht reversible Folgen	Fehler bleibt lokal, intern, vorbereitend oder unveröffentlicht und kann sofort korrigiert werden	Basissorgfalt, Plausibilitätsprüfung durch den Nutzer, keine besondere Irreversibilitätsprüfung erforderlich
I2	Überwiegend reversible Folgen	Korrektur ist mit geringem Aufwand möglich; erhebliche Restschäden sind nicht zu erwarten	Sachprüfung vor Weiterverwendung, einfache Korrekturmöglichkeit, Verantwortlichkeit für Berichtigung
I ≥ 3	Begrenzt reversible Folgen	Korrektur ist möglich, aber mit Aufwand, Verzögerung, Kosten, Vertrauensverlust oder organisatorischen Folgen verbunden	Korrekturverfahren, Dokumentation, Zuständigkeiten, Rückverfolgung betroffener Ergebnisse oder Entscheidungen
I ≥ 4	Schwer reversible Folgen	Auch nach Korrektur können erhebliche Restschäden, Benachteiligungen, rechtliche Folgen oder Reputationsschäden bestehen bleiben	Vorabprüfung, Freigabeverfahren, Folgenabschätzung, Beschwerde- und Wiedergutmachungsmechanismen
I = 5	Kaum oder nicht reversible Folgen	Folgen können körperlich, biologisch, sicherheitskritisch, dauerhaft personenbezogen oder praktisch nicht rückholbar sein	Strenge Präventionspflicht, externe Prüfung, Notfall- und Rückfallkonzept, Einsatz nur bei nachweisbarer Vertretbarkeit

Tabelle 12: Prüfschwellen und Schwellenkriterien der Irreversibilitätsdimension im ASEI-10-Modell

Die Prüfschwellen der Irreversibilitätsdimension sind kumulativ zu verstehen. Ein System mit I4 erfüllt nicht nur die Anforderungen an schwer reversible Folgen, sondern auch die Anforderungen an begrenzt reversible Folgen, sofern diese Merkmale im konkreten Anwendungskontext vorliegen. Ein System mit I5 löst stets die strengsten Anforderungen der Irreversibilitätsdimension aus, unabhängig davon, ob der rechnerische ASEI-10-Gesamtscore allein bereits als kritisch erscheint.

Für die praktische Anwendung gilt: Die Irreversibilitätsstufe wird nach dem schwersten realistisch möglichen Folgeschaden bestimmt, nicht nach dem günstigsten Korrekturszenario. Wenn eine KI-Ausgabe zunächst harmlos erscheint, aber bei Verwendung in einer Entscheidung, Veröffentlichung, Datenverarbeitung, Infrastruktursteuerung oder physischen Handlung schwer rückholbare Folgen auslösen kann, ist eine höhere Irreversibilitätsstufe zu prüfen. Maßgeblich sind Rückholbarkeit, Restschaden, Korrekturaufwand und Kompensierbarkeit.

8.5 Zusammenfassung und Überleitung zur Scorebildung

Die Irreversibilitätsdimension des ASEI-10-Modells beschreibt, wie schwer negative Folgen eines KI-Systems nachträglich rückholbar, korrigierbar oder kompensierbar sind. Sie ergänzt die Dimensionen Autonomie, Severität und Exposition, indem sie nicht auf die operative Selbstständigkeit, die Schutzbedürftigkeit des Einsatzbereichs oder die Breite des Wirkungsraums abstellt, sondern auf die Dauerhaftigkeit möglicher Folgen.

Die fünfstufige Irreversibilitätsskala I1 bis I5 reicht von leicht reversiblen Folgen bis zu kaum oder nicht rückholbaren Schäden. Die jeweilige Stufennummer entspricht dem Bewertungswert innerhalb der Dimension Irreversibilität. Maßgeblich ist stets die höchste Irreversibilitätsstufe, die im konkreten Anwendungskontext realistisch erreicht werden kann. Besonders hohe Irreversibilitätswerte lösen zusätzliche Prüfschwellen und Schwellenkriterien aus, damit schwer oder kaum rückholbare Folgen nicht durch einen rechnerischen Gesamtscore verdeckt werden.

Mit der Irreversibilität ist die vierte Bewertungsdimension des ASEI-10-Modells vollständig beschrieben. Das Risikopotenzial eines KI-Systems ergibt sich nun aus dem Zusammenspiel von Autonomie, Severität, Exposition und Irreversibilität. Die folgenden Abschnitte führen diese vier Dimensionen in einem gemeinsamen Bewertungsverfahren zusammen und erläutern die Bildung des ASEI-10-Scores.

Die Scorebildung folgt der Grundannahme, dass sich Risikopotenzial nicht rein additiv, sondern verstärkend aus dem Zusammenwirken der vier Dimensionen ergibt. Deshalb wird zunächst ein multiplikativer Rohwert gebildet. Um extreme Zahlenwerte zu vermeiden und die Ergebnisse praktikabel interpretierbar zu machen, wird dieser Rohwert anschließend logarithmisch auf eine Gesamtskala von 1 bis 10 normiert.

9 Scorebildung im ASEI-10-Modell

9.1 Grundlogik und Rechenwerte

9.1.1 Grundlogik der Scorebildung

Die Scorebildung im ASEI-10-Modell führt die vier Bewertungsdimensionen Autonomie, Severität, Exposition und Irreversibilität zu einem gemeinsamen Klassifikationswert zusammen. Ziel ist es, das Risikopotenzial eines konkreten KI-Systems im konkreten Anwendungskontext nicht nur qualitativ zu beschreiben, sondern in einen nachvollziehbaren, vergleichbaren und dokumentierbaren Indexwert zu überführen.

Der ASEI-10-Score ist dabei nicht als exakte Risikomessung im klassischen Sinn zu verstehen. Er bildet keine Eintrittswahrscheinlichkeit ab und misst keine objektiven Risikoeinheiten. Vielmehr handelt es sich um einen heuristischen Klassifikationsindex, der die strukturelle Risikointensität eines KI-Systems sichtbar macht. Der Score beantwortet die Frage, wie ausgeprägt das Risikopotenzial eines Systems aufgrund seiner operativen Selbstständigkeit, der möglichen Folgen, der Wirkungsreichweite und der Rückholbarkeit negativer Folgen ist.

Die Scorebildung erfolgt in zwei Schritten. Zunächst werden die vier Dimensionen in einen multiplikativen Rohwert überführt. Dieser Rohwert bildet die Annahme ab, dass sich risikorelevante Merkmale im Anwendungskontext gegenseitig verstärken können. Anschließend wird der Rohwert logarithmisch auf eine Skala von 1 bis 10 normiert. Die logarithmische Normierung dient der besseren Interpretierbarkeit und Vergleichbarkeit, ohne den Rohwert als lineare Risikogröße auszugeben.

9.1.2 Bewertungsskalen und Rechenwerte der vier Dimensionen

Für die Scorebildung werden die zuvor bestimmten Stufenwerte der vier Dimensionen verwendet. Die Autonomie wird qualitativ auf einer neunstufigen Skala A1 bis A9 bewertet. Für die rechnerische Scorebildung wird diese qualitative Autonomiestufe jedoch auf einen normierten Autonomiewert A_n im Wertebereich 1 bis 5 übertragen. Dadurch wird vermieden, dass Autonomie allein aufgrund der längeren Skala rechnerisch stärker gewichtet wird als die übrigen Dimensionen.

Severität, Exposition und Irreversibilität werden jeweils auf fünfstufigen Skalen bewertet und direkt mit den Werten 1 bis 5 in die Berechnung übernommen.

Dimension	Symbol	Qualitative Skala	Rechenwert	Bedeutung
Autonomie	**A / A_n**	A1–A9	A_n = 1–5	operative Selbstständigkeit des KI-Systems
Severität	**S**	S1–S5	S = 1–5	Schwere möglicher negativer Folgen im Einsatzbereich
Exposition	**E**	E1–E5	E = 1–5	Breite des möglichen Wirkungsraums
Irreversibilität	**I**	I1–I5	I = 1–5	Rückholbarkeit, Korrigierbarkeit und Kompensierbarkeit möglicher Folgen

Tabelle 13: Zuordnung der ASEI-Dimensionen zu Skalen und Rechenwerten

Die qualitative Autonomiestufe bleibt für die inhaltliche Bewertung erhalten. Ein System mit A6 wird also weiterhin als persistenter autonomer Agent beschrieben. Für die Scorebildung wird diese Stufe jedoch in den normierten Autonomiewert A_n überführt.

Die Umrechnung erfolgt nach der Formel $A_n = 1 + 4 \cdot \frac{A-1}{8}$ oder $A_n = \frac{A+1}{2}$.

Daraus ergibt sich folgende Zuordnung:

Qualitative Autonomiestufe	Beschreibung	Normierter Autonomiewert A_n
A1	Passives Grundmodell	1,0
A2	Reaktiver Dialogassistent	1,5
A3	Werkzeugassistenz auf Nutzeranweisung	2,0
A4	Überwachter Agent	2,5
A5	Begrenzter autonomer Agent	3,0
A6	Persistenter autonomer Agent	3,5
A7	Verteilter Agentenverbund	4,0
A8	Selbstoptimierender strategischer Agent	4,5
A9	Nicht zuverlässig kontrollierbarer KI-Akteur	5,0

Tabelle 14: Normierung der Autonomiestufen für die Scorebildung

Diese Normierung verändert nicht die qualitative Aussage der Autonomieskala. Sie dient ausschließlich dazu, die vier Bewertungsdimensionen für die Scorebildung auf einen gemeinsamen Wertebereich zu bringen.

9.2 Rohwertbildung und logarithmische Normierung

9.2.1 Bildung des multiplikativen Rohwerts *R*

Im ersten Rechenschritt wird aus den vier Bewertungsdimensionen der Rohwert $\boldsymbol{R}$ gebildet. Dieser Rohwert steht für das unnormierte Risikoprodukt und ergibt sich aus der Multiplikation des normierten Autonomiewerts mit Severität, Exposition und Irreversibilität:

$$R = A_n \cdot S \cdot E \cdot I$$

Der kleinstmögliche Rohwert R_{min} entsteht, wenn alle Dimensionen den niedrigsten Wert aufweisen:

$$R_{min} = 1 \cdot 1 \cdot 1 \cdot 1 = 1$$

Der größtmögliche Rohwert R_{max} entsteht, wenn alle Dimensionen den höchsten Wert aufweisen:

$$R_{\max} = 5 \cdot 5 \cdot 5 \cdot 5 = 625$$

Der Rohwert kann somit Werte zwischen 1 und 625 annehmen. Ein höherer Rohwert zeigt an, dass mehrere Risikodimensionen stärker ausgeprägt sind oder sich im Anwendungskontext gegenseitig verstärken. Der Rohwert ist jedoch noch nicht der endgültige ASEI-10-Score. Er dient als Zwischenschritt für die anschließende logarithmische Normierung.

Die Multiplikation ist als heuristische Indexbildung zu verstehen. Sie bedeutet nicht, dass die einzelnen Stufen exakte Verhältnisskalen darstellen. Vielmehr bildet sie modellhaft ab, dass Risikopotenzial nicht aus einer isolierten Dimension entsteht, sondern aus dem Zusammenwirken mehrerer Merkmale. Der Rohwert soll diese Kombination sichtbar machen, ohne eine naturwissenschaftlich exakte Messung zu behaupten.

9.2.2 Logarithmische Normierung auf die ASEI-10-Skala

Im zweiten Rechenschritt wird der Rohwert logarithmisch auf eine Skala von 1 bis 10 normiert. Die Normierung erfolgt nach folgender Formel:

$$\boldsymbol{ASEI\text{-}10 = 1 + 9 \cdot \frac{\ln(R)}{\ln(625)}}$$

Dabei bezeichnet **ln** den natürlichen Logarithmus. Es könnte auch ein anderer Logarithmus verwendet werden, solange Zähler und Nenner dieselbe Logarithmusbasis verwenden. Entscheidend ist das Verhältnis zwischen dem Logarithmus des konkreten Rohwerts und dem Logarithmus des maximal möglichen Rohwerts.

Die Formel erfüllt zwei zentrale Bedingungen:

Erstens erhält das geringstmögliche Risikoprofil den Wert 1: $\boldsymbol{R = 1}$

$$\boldsymbol{ASEI\text{-}10 = 1 + 9 \cdot \frac{\ln(1)}{\ln(625)} = 1}$$

Zweitens erhält das höchstmögliche Risikoprofil den Wert 10: $\boldsymbol{R = 625}$

$$\boldsymbol{\text{ASEI-10} = 1 + 9 \cdot \frac{\ln(625)}{\ln(625)} = 10}$$

Damit entsteht eine normierte Skala, auf der alle möglichen Kombinationen der vier Dimensionen zwischen 1 und 10 eingeordnet werden können. Die logarithmische Normierung erhält die Rangfolge des Rohwerts, verdichtet jedoch hohe Rohwerte und macht die Ergebnisse besser interpretierbar.

Wichtig ist die mathematische Einordnung der Logarithmierung. Da gilt:

$$\boldsymbol{\ln(A_n \cdot S \cdot E \cdot I) = \ln(A_n) + \ln(S) + \ln(E) + \ln(I)}$$

ist der normierte ASEI-10-Score ein logarithmisch transformierter Index. Die multiplikative Verknüpfung liegt im Rohwert *R*. Die logarithmische Normierung überführt diesen Rohwert in eine praktikable Darstellung. Der finale Score ist daher nicht als linearer Risikowert zu interpretieren, sondern als normierter Klassifikationsindex.

Für die praktische Anwendung empfiehlt sich eine Rundung des ASEI-10-Scores auf eine Dezimalstelle. Dadurch bleibt der Score ausreichend differenziert, ohne eine Scheingenauigkeit zu erzeugen. Bei internen Audits, wissenschaftlichen Auswertungen oder technischen Dokumentationen kann zusätzlich der ungerundete Wert angegeben werden.

9.2.3 Visualisierung der logarithmischen Normierung

Die nachfolgende Abbildung veranschaulicht die Überführung des unnormierten Rohwerts R in den ASEI-10-Score. Grundlage ist die vollständige Wertetabelle für alle ganzzahligen Rohwerte von $R = 1$ bis $R = 625$. Auf der Abszisse wird der Rohwert R abgetragen, auf der Ordinate der jeweils nach der Normierungsformel berechnete ASEI-10-Score.

Die Darstellung dient ausschließlich der Veranschaulichung der Normierungslogik. Sie führt keine zusätzliche Bewertungsregel ein, sondern zeigt grafisch, wie der multiplikative Rohwert durch logarithmische Transformation in die Skala von 1 bis 10 übertragen wird.

Die dargestellten Punkte beruhen auf der Normierungsfunktion

$$\mathbf{ASEI\text{-}10}(R) = 1 + 9 \cdot \frac{\ln(R)}{\ln(625)}, R \in \{1, 2, \ldots, 625\}.$$

Für die mathematische Einordnung kann dieselbe Formel als stetige Funktion für $R > 0$ betrachtet werden. Dann liegt eine logarithmische, streng monoton steigende und konkave Funktion zugrunde.

Es gilt: $\mathbf{ASEI\text{-}10}'(R) = \frac{9}{R \cdot \ln(625)} > 0$ und $\mathbf{ASEI\text{-}10}''(R) = -\frac{9}{R^2 \cdot \ln(625)} < 0$.

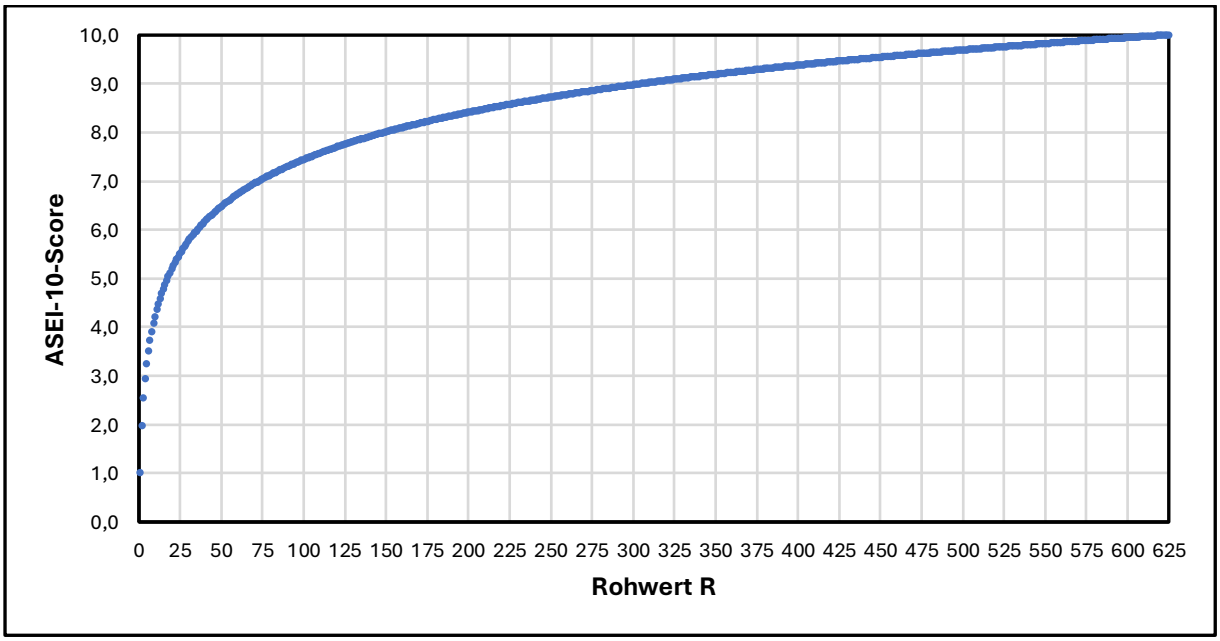
Die erste Ableitung beschreibt die Steigung der Normierungsfunktion. Da sie für alle $R > 0$ positiv ist, steigt der ASEI-10-Score mit zunehmendem Rohwert R durchgehend an. Ein höherer Rohwert führt somit stets zu einem höheren Score.

Die zweite Ableitung beschreibt die Krümmung der Funktion beziehungsweise die Veränderung ihrer Steigung. Da sie für alle $R > 0$ negativ ist, verläuft die Funktion im gesamten betrachteten Bereich konkav. Dies bedeutet, dass die Steigung mit wachsendem Rohwert R kontinuierlich abnimmt. Der Score steigt also weiterhin an, jedoch führt eine gleich große zusätzliche Erhöhung des Rohwerts zu einem immer kleineren zusätzlichen Scorezuwachs.

Die Funktion besitzt keinen Wendepunkt. Die Normierung erfüllt die Randbedingungen

$$\mathbf{ASEI\text{-}10(1) = 1} \text{ und } \mathbf{ASEI\text{-}10(625) = 10.}$$

Abbildung 1: Logarithmische Normierung des Rohwerts R auf den ASEI-10-Score



Dargestellt sind die berechneten ASEI-10-Scorewerte für die ganzzahligen Rohwerte von 1 bis 625. Die Punktfolge verdichtet sich optisch zu einem konkaven logarithmischen Verlauf. Damit wird sichtbar, dass der ASEI-10-Score kein linearer Risikowert, sondern ein logarithmisch normierter Klassifikationsindex ist.

Rohwert *R* von/bis	Ungerundetes ASEI-10-Scoreintervall
1–2	1,0 ≤ Score < 2,0
3–4	2,0 ≤ Score < 3,0
5–8	3,0 ≤ Score < 4,0
9–17	4,0 ≤ Score < 5,0
18–35	5,0 ≤ Score < 6,0
36–73	6,0 ≤ Score < 7,0
74–149	7,0 ≤ Score < 8,0
150–305	8,0 ≤ Score < 9,0
306–624	9,0 ≤ Score < 10,0
625	Score = 10,0

Tabelle 15: Rohwertbereiche nach ungerundeten ASEI-10-Scoreintervallen

Die Tabelle zeigt die Zuordnung ganzzahliger Rohwerte zu ungerundeten ASEI-10-Scoreintervallen. Der Höchstwert von 10,0 wird nur beim maximal möglichen Rohwert $R = 625$ erreicht. Alle Rohwerte unterhalb von 625 liegen im ungerundeten Scorewert unter 10,0. Dadurch wird deutlich, dass die obere Grenze der ASEI-10-Skala mathematisch nur bei vollständiger Ausschöpfung des maximalen Rohwerts erreicht wird.

9.3 Interpretation des ASEI-10-Scores

9.3.1 Interpretationsbereiche des ASEI-10-Scores

Der ASEI-10-Score ist ein zusammenfassender Indexwert. Er dient der Orientierung, Vergleichbarkeit und Priorisierung. Ein niedriger Score weist auf ein geringes Risikopotenzial hin, ein hoher Score auf ein erhebliches oder sehr hohes Risikopotenzial. Der Score ersetzt jedoch nicht die qualitative Prüfung der vier Einzeldimensionen und nicht die in den vorherigen Kapiteln definierten Prüfschwellen.

Die folgenden Interpretationsbereiche sind als heuristische Orientierung zu verstehen. Sie sind nicht empirisch validiert und stellen keine rechtlichen Risikokategorien dar. Sie sollen die praktische Einordnung erleichtern, müssen aber stets gemeinsam mit den Einzeldimensionen, den Schwellenkriterien und der Begründung der Bewertung betrachtet werden.

Die Einordnung erfolgt grundsätzlich auf Grundlage des ungerundeten ASEI-10-Scorewerts. Gerundete Scorewerte dienen der lesbaren Darstellung, dürfen aber insbesondere an Bereichsgrenzen nicht isoliert über die qualitative Einstufung entscheiden.

Die Interpretationsbereiche ordnen den ASEI-10-Score heuristischen ASEI-Risikoklassen zu. Diese Risikoklassen sind keine rechtlichen Kategorien und nicht mit regulatorischen Einstufungen, etwa nach dem EU AI Act, gleichzusetzen. Sie dienen ausschließlich der modellinternen Orientierung, Vergleichbarkeit und Priorisierung.

ASEI-Risikoklasse	ASEI-10-Scoreintervall	Rohwert *R*	Grundinterpretation
RK1: Sehr niedrig	1,0 ≤ Score < 2,5	1-2	Folgenarme Nutzung mit geringer Autonomie, geringer Severität, geringer Exposition und hoher Korrigierbarkeit.
RK2: Niedrig	2,5 ≤ Score < 4,0	3-8	Begrenztes Risikopotenzial, in der Regel durch Basissorgfalt und Plausibilitätsprüfung beherrschbar.
RK3: Moderat	4,0 ≤ Score < 5,5	9-24	Erkennbares Risikopotenzial; fachliche Prüfung, Dokumentation und klare Verantwortlichkeit erforderlich.
RK4: Erhöht	5,5 ≤ Score < 7,0	25-73	Deutliches Risikopotenzial; systematische Freigaben, Kontrollmechanismen und Folgenabschätzung erforderlich.
RK5: Hoch	7,0 ≤ Score < 8,5	74-213	Hohes Risikopotenzial; strenge Governance, Monitoring, Auditierbarkeit und Eskalationsmechanismen erforderlich.
RK6: Sehr hoch	8,5 ≤ Score < 10,0	214-625	Sehr hohes Risikopotenzial; Einsatz nur bei nachweisbarer Vertretbarkeit, wirksamer Kontrollarchitektur und besonderer Aufsicht.

Tabelle 16: Risikoklassen des ASEI-10-Scores

Die ASEI-Risikoklassen RK1 bis RK6 dienen der heuristischen Einordnung des berechneten ASEI-10-Scores. Die verbalen Bezeichnungen von „sehr niedrig“ bis „sehr hoch“ erleichtern die praktische Interpretation, ersetzen jedoch nicht die Prüfung der Einzeldimensionen, Schwellenkriterien und Bewertungsbegründungen. Farben können ergänzend zur Visualisierung verwendet werden, haben aber keine eigenständige klassifikatorische Bedeutung.

Die Interpretationsbereiche dienen der praktischen Orientierung. Sie sind nicht als starre Grenzen zu verstehen. Ein System mit einem ungerundeten Score knapp unterhalb einer Bereichsgrenze kann in der Praxis ähnlich sorgfältig zu behandeln sein wie ein System knapp oberhalb dieser Grenze, wenn relevante Prüfschwellen ausgelöst werden. Ebenso kann ein System mit mittlerem Score besondere Anforderungen auslösen, wenn einzelne Dimensionen hohe Werte erreichen.

Besonders bei gerundeten Scorewerten ist Vorsicht geboten. Ein rechnerischer Score von beispielsweise 6,98 wird bei Darstellung auf eine Dezimalstelle als 7,0 ausgewiesen, liegt mathematisch jedoch noch unterhalb der Grenze von 7,0. Solche Grenzfälle sind daher nicht mechanisch nach dem gerundeten Anzeigewert, sondern anhand des ungerundeten Scorewerts, der Einzeldimensionen und der einschlägigen Prüfschwellen zu beurteilen.

9.3.2 Zur Rundung und Interpretation von Dezimalwerten

ASEI-10-Scores werden in diesem Papier auf eine Dezimalstelle gerundet, um die Berechnung nachvollziehbar und reproduzierbar darzustellen. Diese Dezimalstelle darf jedoch nicht als Ausdruck einer entsprechend feinen empirischen Messgenauigkeit verstanden werden. Ein Unterschied zwischen beispielsweise 6,9 und 7,0 zeigt zunächst nur eine rechnerische Differenz innerhalb des gewählten Indexverfahrens, nicht aber zwingend einen belastbaren qualitativen Sprung im realen Risikopotenzial.

Die Interpretationsbereiche des ASEI-10-Modells sind daher als heuristische Orientierungsbereiche zu verstehen. Werte in der Nähe einer Bereichsgrenze sollten besonders sorgfältig geprüft werden. In Grenzfällen sind die zugrunde liegenden Einzelwerte, die Begründungen der Dimensionen und die ausgelösten Prüfschwellen stärker zu gewichten als die isolierte Dezimalstelle des Scores.

Für die praktische Anwendung gilt: Die Dezimalstelle dient der Reproduzierbarkeit der Rechnung, nicht der Scheingenauigkeit der Bewertung.

9.3.3 Verhältnis von Score und Prüfschwellen

Der ASEI-10-Score und die Prüfschwellen erfüllen unterschiedliche Funktionen. Der Score fasst das Risikopotenzial rechnerisch zusammen und ordnet das System einer Risikoklasse zu. Die Prüfschwellen verhindern, dass qualitativ bedeutsame Risikomerkmale durch den Gesamtscore verdeckt werden. Beide Ebenen müssen daher gemeinsam betrachtet werden.

Ein KI-System kann beispielsweise rechnerisch nur einen moderaten oder erhöhten Score erreichen, aber dennoch eine hohe Irreversibilität oder eine hohe Severität aufweisen. In einem solchen Fall dürfen die besonderen Anforderungen der betreffenden Dimension nicht durch den Gesamtscore abgeschwächt werden. Wenn etwa S4, S5, I4, I5, E5 oder A7 bis A9 erreicht werden, lösen diese Werte eigenständige Prüf-, Kontroll- oder Governance-Anforderungen aus.

ASEI-10 verzichtet bewusst darauf, ausgelöste Prüfschwellen automatisch in eine höhere Risikoklasse umzurechnen. Eine Höchststufe in einer einzelnen Dimension hebt den ASEI-10-Score und damit die Risikoklasse nicht mechanisch an. Stattdessen wirken Score und Prüfschwellen als zwei getrennte, parallel geltende Steuerungsebenen: Der Score bestimmt die Risikoklasse, während die Prüfschwellen eigenständige Mindestanforderungen begründen, die unabhängig von der Risikoklasse einzuhalten sind. Diese Trennung ist eine bewusste Konstruktionsentscheidung. Sie verhindert, dass quantitative Indexbildung und qualitative Schwellenlogik vermischt werden, und hält nachvollziehbar, worauf eine Anforderung jeweils beruht – auf der zusammenfassenden Risikointensität oder auf einem einzelnen kritischen Merkmal. Eine automatische Hochstufung würde diese Unterscheidung verwischen und die Transparenz der Bewertung verringern.

Für die praktische Anwendung bedeutet dies:

(1) Der ASEI-10-Score zeigt die zusammenfassende Risikointensität.

(2) Die vier Einzeldimensionen zeigen, woher das Risiko stammt.

(3) Der normierte Autonomiewert A_n zeigt, wie die qualitative Autonomiestufe rechnerisch eingeordnet wurde.

(4) Die Prüfschwellen zeigen, welche Mindestanforderungen unabhängig vom Gesamtscore gelten.

Damit verbindet das Modell quantitative Vergleichbarkeit mit qualitativer Risikosteuerung. Der Score macht unterschiedliche KI-Systeme vergleichbar. Die Prüfschwellen sorgen dafür, dass kritische Einzelmerkmale nicht nivelliert werden.

9.4 Beispielrechnung und Überleitung zur Anwendung

9.4.1 Beispielhafte Berechnung

Ein KI-System wird wie folgt eingestuft: **Autonomie A = 4**

Das System ist ein überwachter Agent, der Aufgaben mehrstufig plant und Werkzeuge selbst auswählt, aber kritische Schritte weiterhin zur Freigabe vorlegt.

Der qualitative Autonomiewert A4 wird zunächst in den normierten Autonomiewert A_n überführt:

$$\boldsymbol{A_n = 1 + 4 \cdot \frac{4-1}{8} = 2,5} \text{ bzw. } \boldsymbol{A_n = 1 + \frac{4-1}{2} = 2,5}$$

Das System wird in einem personenbezogenen oder wirtschaftlich relevanten Kontext eingesetzt, etwa bei Kundeninformationen oder Vertragsdaten und wird mit der **Severität S = 3** eingestuft.

Die Wirkung betrifft einen organisatorischen Teilbereich oder einen abgegrenzten Datenbestand. Daraus ergibt sich die Einstufung **Exposition E = 3**.

Fehler sind grundsätzlich korrigierbar, können aber Aufwand, Kosten, Verzögerungen oder Vertrauensverlust auslösen und werden daher mit der **Irreversibilität I = 3** eingestuft.

Der Rohwert *R* lautet: $\boldsymbol{R = 2,5 \cdot 3 \cdot 3 \cdot 3 = 67,5}$

Der normierte ASEI-10-Score lautet: $\textbf{ASEI-10} = \mathbf{1 + 9 \cdot \frac{\ln(67,5)}{\ln(625)}}$.

Daraus ergibt sich gerundet: $\textbf{ASEI-10} \approx \mathbf{6,9}$.

Das System liegt damit am oberen Rand des erhöhten Risikopotenzials. Zusätzlich sind die Prüfschwellen A ≥ 4, S ≥ 3, E ≥ 3 und I ≥ 3 zu beachten. Daraus ergeben sich insbesondere Anforderungen an Agentenprüfung, fachliche Kontrolle, Dokumentation, Begrenzung des Einsatzbereichs, prozessbezogene Prüfung und Korrekturverfahren.

Dieses Beispiel zeigt, dass der Score nicht isoliert interpretiert werden darf. Der Wert von 6,9 liegt knapp unterhalb des Bereichs „hoch“. Aufgrund der ausgelösten Prüfschwellen kann das System dennoch strengere Kontroll- und Dokumentationsanforderungen auslösen. Maßgeblich ist daher die Verbindung von Score, Einzeldimensionen und Schwellenkriterien.

9.4.2 Zusammenfassung der Scorebildung und Überleitung

Die Scorebildung des ASEI-10-Modells verbindet die vier Bewertungsdimensionen Autonomie, Severität, Exposition und Irreversibilität zu einem gemeinsamen Klassifikationsindex. Grundlage ist zunächst ein multiplikativer Rohwert, der die kombinierte Ausprägung der vier Dimensionen abbildet. Anschließend wird dieser Rohwert logarithmisch auf eine Skala von 1 bis 10 normiert.

Durch die Normierung der Autonomiestufe auf den Wertebereich 1 bis 5 werden alle vier Dimensionen rechnerisch gleich behandelt. Die qualitative Ausdifferenzierung der Autonomie in neun Stufen bleibt erhalten, führt aber nicht zu einer verdeckten mathematischen Übergewichtung.

Der ASEI-10-Score ermöglicht eine kompakte, vergleichbare und nachvollziehbare Einordnung des Risikopotenzials eines konkreten KI-Systems. Gleichzeitig bleibt die Analyse der Einzeldimensionen unverzichtbar, weil nur sie zeigt, aus welchen Risikomerkmalen sich der Gesamtscore zusammensetzt.

Die Prüfschwellen und Schwellenkriterien ergänzen den Score um eine qualitative Steuerungsebene. Sie stellen sicher, dass besonders relevante Autonomie-, Severität-, Expositions- oder Irreversibilitätsmerkmale nicht durch die rechnerische Gesamtbewertung verdeckt werden. Das ASEI-10-Modell verbindet damit eine quantitative Indexlogik mit einer risikobezogenen Mindestprüfung.

Im nächsten Schritt kann das Modell auf konkrete KI-Systeme angewendet werden. Dazu werden die vier Dimensionen systematisch geprüft, die jeweiligen Stufenwerte bestimmt, der normierte Autonomiewert berechnet, der Rohwert gebildet, der ASEI-10-Score abgeleitet und die einschlägigen Prüfschwellen dokumentiert. Erst aus dieser Verbindung von Einzelbewertung, Scorebildung und Schwellenprüfung ergibt sich eine belastbare Risikoeinstufung.

10 Anwendung des ASEI-10-Modells

10.1 Zweck des Anwendungsverfahrens

Das ASEI-10-Modell dient nicht nur der abstrakten Beschreibung von KI-Risiken, sondern der systematischen Einstufung konkreter KI-Systeme in konkreten Anwendungskontexten. Das Anwendungsverfahren beschreibt, wie ein KI-System abgegrenzt, in seinem System- und Nutzungskontext erfasst, entlang der vier Bewertungsdimensionen Autonomie, Severität, Exposition und Irreversibilität eingestuft und anschließend durch Scorebildung, Schwellenprüfung und Dokumentation eingeordnet wird.

Ziel des Anwendungsverfahrens ist eine nachvollziehbare, dokumentierbare und wiederholbare Risikoeinstufung. Die Bewertung soll nicht auf bloßen Eindrücken, allgemeinen Herstellerangaben oder pauschalen Aussagen über ein KI-Modell beruhen, sondern auf den tatsächlichen Funktionen, Berechtigungen, Datenzugriffen, Werkzeugen, Nutzergruppen, Einsatzgrenzen und möglichen Folgen des konkreten KI-Systems. Dadurch wird sichtbar, ob das Risikopotenzial vor allem aus der Autonomie des Systems, aus der Severität des Einsatzbereichs, aus der Breite der Exposition, aus der Irreversibilität möglicher Folgen oder aus deren Zusammenwirken entsteht.

Das Verfahren ist bewusst modular aufgebaut. Zunächst wird der Bewertungsgegenstand bestimmt und der Systemkontext beschrieben. Anschließend werden die vier Dimensionen einzeln begründet und bewertet. Erst danach werden die Werte rechnerisch zusammengeführt. Dadurch bleibt die Bewertung transparent: Der ASEI-10-Score liefert eine kompakte Gesamteinordnung, während die Einzeldimensionen zeigen, aus welchen Risikomerkmalen sich diese Einordnung ergibt.

Die zusätzlich zu prüfenden Schwellenkriterien ergänzen den Score um eine qualitative Kontroll- und Steuerungsebene. Sie stellen sicher, dass besonders relevante Einzelmerkmale nicht durch den Gesamtscore verdeckt werden. Die Dokumentation der Bewertung macht schließlich nachvollziehbar, welche Annahmen getroffen wurden, welche Stufen gewählt wurden und welche Mindestanforderungen sich aus der Einstufung ergeben.

10.2 Grundlogik des Bewertungsprozesses

10.2.1 Funktion des Bewertungsprozesses

Die Anwendung des ASEI-10-Modells erfolgt nicht durch eine unmittelbare Scoreberechnung, sondern durch einen strukturierten Bewertungsprozess. Dieser Prozess soll sicherstellen, dass der ASEI-10-Score nicht isoliert entsteht, sondern aus einer nachvollziehbaren Beschreibung des konkreten KI-Systems, seines Anwendungskontextes und seiner risikorelevanten Eigenschaften abgeleitet wird.

Der Bewertungsprozess erfüllt dabei drei Funktionen. Erstens grenzt er den Bewertungsgegenstand eindeutig ab. Bewertet wird nicht „KI“ im Allgemeinen und auch nicht ein Modell in abstrakter Form, sondern ein konkretes KI-System mit bestimmten Funktionen, Berechtigungen, Datenzugriffen, Nutzergruppen, Einsatzgrenzen und möglichen Folgen. Zweitens zwingt der Prozess dazu, die vier Bewertungsdimensionen Autonomie, Severität, Exposition und Irreversibilität einzeln zu begründen, bevor sie rechnerisch zusammengeführt werden. Drittens verbindet er die Scorebildung mit Schwellenprüfung, Dokumentation und Ableitung von Mindestanforderungen.

Damit dient der Bewertungsprozess zugleich als Prüfschema, Dokumentationsrahmen und Governance-Instrument. Er unterstützt eine wiederholbare Anwendung des Modells, macht Bewertungsannahmen transparent und erleichtert den Vergleich unterschiedlicher KI-Systeme innerhalb einer Organisation. Zugleich verhindert er, dass der ASEI-10-Score als bloße Rechenzahl missverstanden wird. Belastbar wird die Einstufung erst durch die Verbindung von Systembeschreibung, Dimensionseinstufung, Scorebildung, Schwellenprüfung und fachlicher Begründung.

10.2.2 Die sieben Bewertungsschritte im Überblick

Schritt	Bezeichnung	Leitfrage	Ergebnis
1	Bewertungsgegenstand festlegen	Welches konkrete KI-System wird in welchem Nutzungskontext bewertet?	eindeutig abgegrenzter Bewertungsgegenstand
2	Systemkontext beschreiben	Welche Funktionen, Daten, Nutzergruppen, Werkzeuge, Rechte und Einsatzgrenzen liegen vor?	strukturierte Systembeschreibung
3	Bewertungsdimensionen einstufen	Wie sind Autonomie, Severität, Exposition und Irreversibilität jeweils zu bewerten?	begründete Einzelwerte für A, S, E und I
4	Scorelogik anwenden	Welche Risikoeinstufung ergibt sich bei Anwendung der in Kapitel 9.2 dargestellten Scorelogik?	rechnerisch gestützte Gesamteinordnung
5	Schwellenkriterien prüfen	Werden qualitative Prüfschwellen ausgelöst, die zusätzliche Mindestanforderungen begründen?	ergänzende Prüf- und Steuerungsanforderungen
6	Bewertung dokumentieren	Welche Annahmen, Begründungen, Werte und Einschränkungen liegen der Einstufung zugrunde?	nachvollziehbare Bewertungsdokumentation
7	Mindestanforderungen ableiten	Welche Kontrollen, Freigaben, Dokumentations-, Monitoring- oder Neubewertungsanforderungen folgen aus der Einstufung?	risikoadäquate Anwendungsvorgaben

Tabelle 17: Bewertungsprozess des ASEI-10-Modells

10.2.3 Vom Bewertungsprozess zur Risikoeinstufung

Der Bewertungsprozess ist sequenziell aufgebaut, darf aber nicht als rein mechanisches Ausfüllen einer Tabelle verstanden werden. Die ersten beiden Schritte bestimmen, was überhaupt bewertet wird. Erst wenn der Bewertungsgegenstand und der Systemkontext hinreichend klar beschrieben sind, können die vier Bewertungsdimensionen sinnvoll eingestuft werden.

Die Einstufung der Dimensionen bildet den qualitativen Kern der Bewertung. Jede Dimension ist eigenständig zu begründen. Autonomie beschreibt die operative Selbstständigkeit des Systems, Severität die mögliche Schwere negativer Folgen im Einsatzbereich, Exposition die Breite des möglichen Wirkungsraums und Irreversibilität die Rückholbarkeit möglicher negativer Folgen. Die anschließende Scorebildung verdichtet diese Einzelbewertungen zu einem normierten Klassifikationswert, ersetzt aber nicht die fachliche Begründung.

Die Schwellenprüfung ergänzt den Score um eine qualitative Kontrolllogik. Sie stellt sicher, dass besonders bedeutsame Einzelmerkmale nicht durch den Gesamtscore verdeckt werden. Erst aus der Verbindung von Systembeschreibung, Dimensionseinstufung, Scorebildung, Schwellenprüfung und Dokumentation ergibt sich eine belastbare Risikoeinstufung.

Die folgenden Abschnitte erläutern die einzelnen Teile dieses Bewertungsprozesses. Abschnitt 10.3 behandelt Bewertungsgegenstand und Systemkontext. Abschnitt 10.4 beschreibt Einstufung, Scorebildung und Schwellenprüfung. Abschnitt 10.5 behandelt Dokumentation und Mindestanforderungen.

10.3 Bewertungsgegenstand und Systemkontext

10.3.1 Bewertungsgegenstand festlegen

Am Anfang jeder ASEI-10-Bewertung steht die eindeutige Festlegung des Bewertungsgegenstands. Bewertet wird nicht „KI“ im Allgemeinen und auch nicht ein KI-Modell in abstrakter Form, sondern ein konkretes KI-System in einem konkret beschriebenen Nutzungskontext. Diese Abgrenzung ist entscheidend, weil dasselbe Modell je nach Einbettung, Berechtigungen, Datenzugriffen, Nutzergruppen und Einsatzbereich zu sehr unterschiedlichen Risikoeinstufungen führen kann.

Ein Sprachmodell, das privat für kreative Textentwürfe genutzt wird, ist anders zu bewerten als ein KI-Assistent, der mit Kundendaten arbeitet. Ein Dialogsystem ohne Werkzeugzugriff besitzt ein anderes Risikoprofil als ein Agent, der Dateien verändern, E-Mails versenden, Datenbanken abfragen oder Prozesse auslösen kann. Ein System in einem isolierten Testumfeld ist anders zu bewerten als dasselbe System im produktiven Betrieb mit Außenwirkung.

Der Bewertungsgegenstand sollte daher so präzise beschrieben werden, dass eindeutig erkennbar ist, worauf sich die Einstufung bezieht. Dazu gehören mindestens die Bezeichnung des Systems, der konkrete Einsatzzweck, der organisatorische Einsatzort, die Nutzergruppen, die verarbeiteten Datenarten, die technischen Funktionen, mögliche Werkzeugzugriffe, Schnittstellen, Berechtigungen und die Frage, ob das System lediglich empfiehlt, Inhalte erzeugt oder selbst Handlungen ausführt.

Eine unklare Abgrenzung führt zu unklaren Bewertungen. Deshalb sollte bei jeder ASEI-10-Einstufung dokumentiert werden, welcher konkrete Systemzustand und welcher konkrete Anwendungskontext bewertet wurden. Wird das System später erweitert, in einen anderen Kontext übertragen, mit zusätzlichen Rechten ausgestattet oder für andere Nutzergruppen geöffnet, ist die ursprüngliche Bewertung nicht automatisch weiterhin gültig.

10.3.2 Systemkontext beschreiben

Nach der Festlegung des Bewertungsgegenstands wird der Systemkontext beschrieben. Diese Beschreibung bildet die Grundlage für die spätere Einstufung der vier Bewertungsdimensionen. Sie muss so konkret sein, dass nachvollziehbar wird, warum ein bestimmter Wert für Autonomie, Severität, Exposition oder Irreversibilität vergeben wurde.

Der Systemkontext umfasst die technische, organisatorische und funktionale Einbettung des KI-Systems. Dazu gehören insbesondere Aufgaben, Daten, Nutzergruppen, Betroffene, Werkzeugzugriffe, Schnittstellen, Berechtigungen, Kontrollmechanismen, Einsatzgrenzen und mögliche Folgen einer fehlerhaften, missbräuchlichen oder unzureichend kontrollierten Nutzung. Die Kontextbeschreibung ist damit keine rein technische Systembeschreibung, sondern eine risikobezogene Analyse des tatsächlichen Anwendungskontextes.

Besonders wichtig ist die Unterscheidung zwischen Modellleistung und Systemwirkung. Die Leistungsfähigkeit eines KI-Modells sagt für sich genommen noch nicht aus, welches Risikopotenzial im konkreten Einsatz entsteht. Erst durch die Verbindung mit Daten, Rechten, Werkzeugen, Nutzern, organisatorischen Prozessen und möglichen Außenwirkungen wird sichtbar, welche Risikotreiber für die ASEI-10-Bewertung relevant sind.

Die Systemkontextbeschreibung sollte weder unnötig technisch überladen noch zu allgemein gehalten sein. Sie muss nicht jedes technische Detail erfassen, aber alle Merkmale enthalten, die für Autonomie, Severität, Exposition und Irreversibilität bedeutsam sein können. Unklare oder fehlende Informationen sind als Bewertungsunsicherheit zu dokumentieren.

10.3.3 Kontextfragen als drei Analysebereiche

Zur Strukturierung der Systemkontextbeschreibung können neun Leitfragen verwendet werden. Sie dienen nicht dem bloßen Ausfüllen einer Checkliste, sondern der risikobezogenen Aufklärung des Anwendungskontextes. Die Fragen sind deshalb nicht isoliert zu beantworten, sondern im Zusammenhang zu betrachten.

Die neun Leitfragen lassen sich in drei Analysebereiche gliedern:

Analysebereich	Leitfragen	Risikobezogene Funktion
1. Systemfunktion und technische Einbettung	1. Welche Aufgaben soll das KI-System erfüllen? 2. Welche Daten verarbeitet das System? 3. Welche Werkzeuge, Schnittstellen oder externen Systeme kann es nutzen?	Erfasst die funktionale Rolle des Systems, die berührten Daten- und Schutzgüter sowie den Grad seiner technischen Einbindung in andere Prozesse oder Systeme.
2. Nutzung, Handlungstiefe und Kontrolle	4. Welche Nutzerinnen und Nutzer greifen auf das System zu? 5. Kann das System nur Antworten erzeugen oder auch Handlungen auslösen? 6. Wer prüft die Ergebnisse?	Beschreibt die konkrete Nutzungssituation, die operative Handlungstiefe des Systems und die vorgesehene menschliche oder organisatorische Kontrolle.
3. Betroffenheit, Steuerung und mögliche Folgen	7. Welche Freigabe-, Kontroll- und Eskalationsmechanismen bestehen? 8. Welche Personen, Organisationseinheiten oder externen Empfänger können betroffen sein? 9. Welche Folgen können entstehen, wenn das System fehlerhaft, missbräuchlich oder unzureichend kontrolliert eingesetzt wird?	Verdeutlicht den möglichen Wirkungsraum, die vorhandenen Steuerungsmechanismen und die Art möglicher negativer Folgen.

Tabelle 18: Kontextfragen zur Beschreibung des Bewertungsgegenstands und des Systemkontexts

Der erste Analysebereich beschreibt, welche Aufgabe das KI-System übernimmt und wie es technisch eingebettet ist. Ein System, das lediglich Textvorschläge erzeugt, ist anders zu bewerten als ein System, das Kundenfälle abschließt, Bewerbungen vorsortiert, technische Warnmeldungen priorisiert oder operative Maßnahmen auslöst. Ebenso macht es einen erheblichen Unterschied, ob das System nur mit allgemeinen Informationen arbeitet oder personenbezogene Daten, Beschäftigtendaten, Prüfungsdaten, Geschäftsgeheimnisse oder sicherheitsrelevante Systemdaten verarbeitet. Werkzeugzugriffe und Schnittstellen können das Risikopotenzial zusätzlich erhöhen, weil das System dann nicht nur Informationen erzeugt, sondern technische oder organisatorische Wirkungen auslösen kann.

Der zweite Analysebereich richtet den Blick auf Nutzung, Handlungstiefe und Kontrolle. Hier wird geklärt, wer das System verwendet, ob die Nutzer fachkundig sind, ob externe Personen unmittelbar mit dem System interagieren und ob die Ergebnisse geprüft werden. Besonders relevant ist, ob das System lediglich Antworten erzeugt oder selbst Handlungen auslösen kann. Ein System, das nur eine Empfehlung formuliert, besitzt ein anderes Risikoprofil als ein System, das eigenständig Nachrichten versendet, Daten verändert, Aufgaben anstößt oder Entscheidungen vorbereitet.

Der dritte Analysebereich betrachtet Betroffenheit, Steuerung und mögliche Folgen. Er klärt, welche Personen, Organisationseinheiten oder externen Empfänger von den Systemwirkungen betroffen sein können und welche Folgen bei Fehlern, Fehlverwendung, Missbrauch oder Kontrollproblemen entstehen können. Zugleich wird geprüft, welche Freigabe-, Kontroll- und Eskalationsmechanismen vorhanden sind. Diese Informationen sind besonders wichtig für die Bewertung von Exposition und Irreversibilität.

Die Kontextfragen führen noch nicht selbst zur Scoreberechnung. Sie bereiten die spätere Einstufung der vier Bewertungsdimensionen vor. Ihre Funktion besteht darin, die relevanten Systemmerkmale so sichtbar zu machen, dass die anschließende Bewertung von Autonomie, Severität, Exposition und Irreversibilität begründet und überprüfbar vorgenommen werden kann.

Die Leitfragen werden im Anwendungskapitel nicht abstrakt beantwortet. Ihre konkrete Beantwortung erfolgt erst im Rahmen einer einzelnen ASEI-10-Bewertung, insbesondere im Bewertungsbogen oder in den Fallbeispielen. Dadurch bleibt Kapitel 10 methodisch allgemein, während Kapitel 11 zeigt, wie die Leitfragen auf konkrete Nutzungskontexte angewendet werden.

10.4 Einstufung, Scoreanwendung und Schwellenprüfung

10.4.1 Einstufung der vier Bewertungsdimensionen

Nach der Beschreibung des Bewertungsgegenstands und des Systemkontexts werden die vier Bewertungsdimensionen Autonomie, Severität, Exposition und Irreversibilität einzeln eingestuft. Diese Einstufung bildet den qualitativen Kern der ASEI-10-Bewertung. Der spätere Score darf daher nicht als Ersatz für die fachliche Begründung verstanden werden, sondern nur als verdichtete Darstellung der zuvor vorgenommenen Einzeleinstufungen.

Die Autonomie wird auf der neunstufigen Skala A1 bis A9 bewertet. Maßgeblich ist dabei, wie selbstständig das System im konkreten Anwendungskontext operieren kann. Zu prüfen ist insbesondere, ob das System lediglich passiv oder reaktiv unterstützt, ob es Werkzeuge nutzt, ob es Handlungsschritte planen oder ausführen kann, ob es dauerhaft ereignisgesteuert aktiv bleibt oder ob es strategisch-systemische Eigenschaften wie Verteilung, Selbstoptimierung oder eingeschränkte Kontrollierbarkeit aufweist.

Die Severität wird auf der fünfstufigen Skala S1 bis S5 bewertet. Maßgeblich ist die mögliche Schwere negativer Folgen im konkreten Einsatzbereich. Dabei geht es nicht um die Wahrscheinlichkeit eines Fehlers, sondern um die Bedeutung der berührten Schutzgüter und die potenzielle Tragweite negativer Folgen. Private, kreative oder vorbereitende Nutzungen sind daher anders zu bewerten als Anwendungen in Bildung, Beschäftigung, Gesundheit, Recht, öffentlicher Sicherheit, kritischer Infrastruktur oder Cybersecurity.

Die Exposition wird auf der fünfstufigen Skala E1 bis E5 bewertet. Maßgeblich ist die Breite des möglichen Wirkungsraums. Zu prüfen ist, ob mögliche Wirkungen auf eine einzelne Person begrenzt bleiben, eine Gruppe betreffen, organisationale Prozesse erreichen, öffentlich sichtbar werden oder systemische Wirkungen über mehrere Organisationen, Plattformen oder Infrastrukturen hinweg entfalten können.

Die Irreversibilität wird auf der fünfstufigen Skala I1 bis I5 bewertet. Maßgeblich ist, wie schwer mögliche negative Folgen nachträglich erkannt, gestoppt, korrigiert, zurückgeholt oder kompensiert werden können. Leicht verwerfbare Entwürfe sind anders zu bewerten als fehlerhafte Entscheidungen, Datenschutzverletzungen, Benachteiligungen, Reputationsschäden, körperliche Schäden oder sicherheitskritische Folgewirkungen.

Jede Dimension ist eigenständig zu begründen. Eine hohe Einstufung in einer Dimension rechtfertigt nicht automatisch eine hohe Einstufung in einer anderen Dimension. Gerade in kritischen Anwendungskontexten können Severität, Exposition und Irreversibilität zwar sachlogisch zusammenwirken; sie erfassen aber unterschiedliche Risikoeigenschaften und müssen getrennt geprüft werden.

10.4.2 Anwendung der Scorelogik

Nach der qualitativen Einstufung der vier Bewertungsdimensionen wird die in Kapitel 9.2 dargestellte Scorelogik angewendet. Die qualitative Autonomiestufe wird dafür in den normierten Autonomiewert A_n übertragen. Anschließend werden die vier Bewertungswerte rechnerisch zusammengeführt und auf die ASEI-10-Skala übertragen.

In Kapitel 10 steht dabei nicht die Herleitung der Berechnungslogik im Mittelpunkt, sondern ihre sachgerechte Anwendung im Bewertungsprozess. Entscheidend ist, dass die Berechnung erst nach der fachlichen Einstufung der Dimensionen erfolgt. Der Score soll eine bereits begründete Bewertung zusammenfassen, nicht eine unzureichende Kontextanalyse ersetzen.

Für die Anwendung ist außerdem zu beachten, dass der ASEI-10-Score als heuristischer Klassifikationsindex zu verstehen ist. Er dient der zusammenfassenden Einordnung, Priorisierung und Dokumentation des Risikopotenzials. Einzelne Dezimalstellen dürfen nicht überbewertet werden. Maßgeblich bleibt die Verbindung aus Score, Einzeldimensionen, Schwellenkriterien und Begründung.

Besondere Aufmerksamkeit ist bei Grenzfällen erforderlich. Liegt ein ungerundeter Score nahe an einer Klassengrenze, sollte die Bewertung nicht mechanisch anhand des gerundeten Werts interpretiert werden. In solchen Fällen sind die Einzeldimensionen, die ausgelösten Schwellenkriterien und der konkrete Anwendungskontext besonders sorgfältig zu würdigen.

10.4.3 Prüfung der Schwellenkriterien

Neben der Scoreanwendung sind die Schwellenkriterien der einzelnen Bewertungsdimensionen zu prüfen. Diese Schwellenprüfung ergänzt den rechnerischen Score um eine qualitative Kontrolllogik. Sie verhindert, dass besonders relevante Einzelmerkmale durch den Gesamtscore verdeckt werden.

Eine Prüfschwelle wird erreicht, wenn eine Dimension ein Merkmal aufweist, das unabhängig vom Gesamtscore zusätzliche Aufmerksamkeit, Kontrolle oder Dokumentation erfordert. Dies kann insbesondere bei agentischer Handlungsausführung, sensiblen Einsatzbereichen, breiter öffentlicher oder systemischer Exposition sowie schwer reversiblen Folgen der Fall sein.

Die Schwellenprüfung ist keine zweite Scoreberechnung. Sie verändert den ASEI-10-Score nicht automatisch, sondern ergänzt ihn um Mindestanforderungen. Ein System kann daher auch dann zusätzliche Prüf-, Freigabe- oder Kontrollanforderungen auslösen, wenn der Gesamtscore allein noch keine sehr hohe Risikoklasse erreicht.

Typische Folgen ausgelöster Schwellenkriterien können eine vertiefte fachliche Prüfung, erweiterte Dokumentation, menschliche Freigabepunkte, Protokollierung, Monitoring, Rollback-Möglichkeiten, Begrenzung von Toolzugriffen, Eskalationsverfahren, Datenschutz- oder Sicherheitsprüfung sowie regelmäßige Neubewertung sein.

Die Schwellenkriterien sind besonders wichtig in Grenzfällen. Sie stellen sicher, dass die Bewertung nicht auf eine Rechenzahl reduziert wird, sondern auch qualitativ bedeutsame Risikomerkmale berücksichtigt. Eine belastbare ASEI-10-Einstufung besteht daher immer aus vier Elementen: begründeter Dimensionseinstufung, berechnetem Score, geprüften Schwellenkriterien und dokumentierten Anwendungskonsequenzen.

10.5 Dokumentation und Ableitung von Mindestanforderungen

10.5.1 Dokumentation der Bewertung

Jede ASEI-10-Bewertung sollte so dokumentiert werden, dass sie auch nachträglich nachvollziehbar bleibt. Die Dokumentation dient nicht nur der formalen Ablage, sondern ist ein wesentlicher Bestandteil der Bewertungsqualität. Ohne Dokumentation wäre nicht erkennbar, auf welchen Annahmen die Einstufung beruht, welcher Systemzustand bewertet wurde und warum bestimmte Werte für Autonomie, Severität, Exposition und Irreversibilität vergeben wurden.

Dokumentiert werden sollten mindestens der Bewertungsgegenstand, der Anwendungskontext, die zugrunde gelegten Systemfunktionen, Datenzugriffe, Werkzeugnutzungen, Nutzergruppen, Einsatzgrenzen und Kontrollmechanismen. Ebenso festzuhalten sind die Einzeleinstufungen der vier Bewertungsdimensionen, die jeweilige Begründung, der berechnete ASEI-10-Score, ausgelöste Schwellenkriterien und daraus abgeleitete Mindestanforderungen.

Besonders wichtig ist die Dokumentation von Unsicherheiten. Wenn bestimmte Informationen fehlen, Annahmen getroffen werden müssen oder der Systemkontext noch nicht vollständig feststeht, sollte dies ausdrücklich vermerkt werden. Eine Bewertung mit unsicheren Annahmen ist nicht automatisch unbrauchbar, darf aber nicht denselben Status haben wie eine Bewertung auf Grundlage vollständiger Systeminformationen.

Die Dokumentation erfüllt damit mehrere Funktionen. Sie macht die Bewertung überprüfbar, erleichtert spätere Neubewertungen, unterstützt interne Freigabe- und Governance-Prozesse und schafft eine Grundlage für Audit, Compliance, Risikomanagement und Verantwortungszuordnung. Zugleich verhindert sie, dass der ASEI-10-Score losgelöst von seinem Kontext verwendet wird.

10.5.2 ASEI-10-Bewertungsbogen

Für die praktische Anwendung kann die ASEI-10-Bewertung in einem standardisierten Bewertungsbogen festgehalten werden. Der Bogen strukturiert die Bewertung und sorgt dafür, dass die wesentlichen Angaben zum Bewertungsgegenstand, zum Bewertungsmodus, zum Systemprofil, zur Einstufung der vier ASEI-Dimensionen, zur Prüfung korrelativer Beziehungen, zur Scorebildung, zu Prüfschwellen, Mindestanforderungen und Neubewertung vollständig erfasst werden.

Der Bewertungsbogen ist nicht als bloßes Rechenformular zu verstehen. Er dient zugleich der fachlichen Begründung, der Plausibilitätsprüfung und der späteren Nachvollziehbarkeit der Bewertung. Deshalb sollen die Felder nicht nur formal ausgefüllt, sondern so beschrieben werden, dass Dritte erkennen können, auf welchen Annahmen die Einstufung beruht.

Der vollständige ASEI-10-Bewertungsbogen ist in Anhang B abgedruckt. Die Anwendung des Bewertungsbogens wird in Kapitel 11 anhand von sechs Fallbeispielen demonstriert; die vollständig ausgefüllten Bewertungsbögen zu diesen Fallbeispielen sind in Anhang C zusammengeführt.

Der Bewertungsbogen ist nicht als bürokratisches Zusatzinstrument zu verstehen. Er soll vorhandene Informationen bündeln, Bewertungsannahmen sichtbar machen und die Nachvollziehbarkeit der Einstufung erhöhen. Die Bearbeitung kann dabei auch KI-gestützt vorbereitet werden, etwa durch die strukturierte Zusammenfassung vorhandener System-, Prozess- oder Governance-Dokumente. Die fachliche Prüfung, endgültige Bewertung und Verantwortung für die Einstufung verbleiben jedoch bei den zuständigen Personen oder Stellen.

10.5.3 Ableitung von Mindestanforderungen

Aus der ASEI-10-Bewertung können Mindestanforderungen an den weiteren Umgang mit dem KI-System abgeleitet werden. Diese Anforderungen ergeben sich nicht allein aus dem numerischen Score, sondern aus der Verbindung von Risikoklasse, Einzeldimensionen, ausgelösten Schwellenkriterien und fachlicher Begründung.

Bei niedrigem Risikopotenzial können einfache organisatorische Vorgaben ausreichen, etwa eine klare Zweckbeschreibung, Nutzerhinweise und eine grundlegende Plausibilitätsprüfung. Bei erhöhtem Risikopotenzial sind regelmäßig stärkere Anforderungen erforderlich, etwa fachliche Endkontrolle, Dokumentation der Nutzung, klare Einsatzgrenzen und Verantwortungszuordnung. Bei hohem oder sehr hohem Risikopotenzial können zusätzliche Freigabeprozesse, Monitoring, Protokollierung, Begrenzung von Toolzugriffen, Rückholmechanismen, Eskalationsverfahren, Datenschutz-, Sicherheits- oder Rechtsprüfungen sowie regelmäßige Neubewertungen erforderlich sein.

Besonders sorgfältig sind Systeme zu behandeln, bei denen Schwellenkriterien ausgelöst werden. Dies gilt insbesondere bei agentischer Handlungsausführung, sensiblen Einsatzbereichen, breiter öffentlicher oder systemischer Exposition und schwer reversiblen Folgen. In solchen Fällen können Mindestanforderungen auch dann erforderlich sein, wenn der Gesamtscore allein noch keine sehr hohe Risikoklasse ausweist.

Die Mindestanforderungen sollten risikoadäquat sein. Das bedeutet, dass sie weder rein formal noch überzogen sein dürfen. Ein einfaches assistives System benötigt keine Kontrollarchitektur für kritische Infrastruktur. Umgekehrt darf ein agentisches System mit Toolzugriff, personenbezogenen Daten oder schwer reversiblen Folgen nicht wie ein bloßes Textwerkzeug behandelt werden.

Die Ableitung von Mindestanforderungen verbindet das ASEI-10-Modell mit praktischer KI-Governance. Die Bewertung endet nicht mit dem Score, sondern führt zu konkreten Anwendungsvorgaben. Dadurch wird ASEI-10 von einem reinen Klassifikationsmodell zu einem nutzbaren Instrument für Freigabeentscheidungen, Priorisierung, Kontrolle, Dokumentation und spätere Neubewertung.

10.6 Aktualisierung und Neubewertung

10.6.1 Funktion der Neubewertung

Eine ASEI-10-Bewertung bezieht sich immer auf einen bestimmten Systemzustand und einen bestimmten Anwendungskontext. Sie ist daher keine dauerhaft gültige Eigenschaft eines KI-Systems, sondern eine kontextbezogene Einstufung zu einem bestimmten Bewertungszeitpunkt. Ändern sich die Funktionen, Berechtigungen, Datenzugriffe, Nutzergruppen, Einsatzgrenzen oder möglichen Folgen des Systems, kann sich auch das Risikopotenzial verändern.

Die Neubewertung dient dazu, solche Veränderungen systematisch zu erfassen. Sie verhindert, dass eine frühere Einstufung auf einen veränderten Systemzustand übertragen wird, obwohl die ursprünglichen Bewertungsannahmen nicht mehr gelten. Besonders bei KI-Systemen ist dies wichtig, weil sich Leistungsfähigkeit, Einbettung, Toolzugriffe, Nutzerkreis und organisatorische Nutzungspraxis häufig schrittweise verändern können.

Eine Aktualisierung der Bewertung ist nicht nur bei großen technischen Änderungen erforderlich. Auch organisatorische oder prozessuale Veränderungen können das Risikopotenzial erhöhen. Ein System, das zunächst nur intern getestet wird, kann bei produktivem Einsatz eine höhere Exposition erreichen. Ein Assistent ohne Toolzugriff kann durch spätere Anbindung an E-Mail, Dateien, Datenbanken oder Prozesssysteme eine höhere Autonomie erhalten. Ein System, das zunächst nur vorbereitend eingesetzt wird, kann risikorelevanter werden, wenn seine Ergebnisse faktisch Entscheidungsprozesse prägen.

10.6.2 Anlässe für eine Neubewertung

Eine Neubewertung sollte insbesondere dann durchgeführt werden, wenn sich risikorelevante Merkmale des KI-Systems oder seines Anwendungskontextes verändern.

Typische Anlässe sind:

Anlass	Mögliche Auswirkung auf die ASEI-10-Bewertung
Erweiterung der Systemfunktionen	Mögliche Erhöhung der Autonomie oder Severität
Zusätzlicher Toolzugriff	Mögliche Erhöhung der Autonomie und Irreversibilität
Neue Schnittstellen zu Datenbanken, Kommunikationssystemen oder Prozesssystemen	Mögliche Erhöhung von Autonomie, Exposition und Irreversibilität
Wechsel vom Testbetrieb in den produktiven Betrieb	Mögliche Erhöhung von Exposition und Severität
Ausweitung auf weitere Nutzergruppen oder Organisationseinheiten	Mögliche Erhöhung der Exposition
Verarbeitung sensiblerer Datenarten	Mögliche Erhöhung von Severität und Irreversibilität
Einsatz in einem sensibleren Fach- oder Entscheidungsbereich	Mögliche Erhöhung der Severität
Wegfall oder Abschwächung menschlicher Kontrolle	Mögliche Erhöhung von Autonomie und Irreversibilität
Neue gesetzliche, organisatorische oder sicherheitsbezogene Anforderungen	Mögliche Anpassung von Mindestanforderungen
Bekannte Fehlfunktionen, Sicherheitsvorfälle oder Missbrauchsfälle	Anlass für vertiefte Prüfung und mögliche Neueinstufung

Tabelle 19: Typische Anlässe für eine Neubewertung der ASEI-10-Einstufung

Diese Anlässe sind nicht abschließend. Entscheidend ist, ob sich eine Bewertungsannahme ändert, die für Autonomie, Severität, Exposition oder Irreversibilität bedeutsam sein kann. Sobald dies der Fall ist, sollte die Bewertung überprüft und gegebenenfalls angepasst werden.

10.6.3 Umgang mit geänderten Systemzuständen

Bei einer Neubewertung sollte nicht lediglich der bisherige Score fortgeschrieben werden. Vielmehr ist erneut zu prüfen, ob der Bewertungsgegenstand, der Systemkontext und die vier Dimensionseinstufungen noch zutreffen. Dabei kann die frühere Bewertung als Vergleichsgrundlage dienen, sie ersetzt jedoch nicht die erneute fachliche Prüfung.

Besonders wichtig ist die Unterscheidung zwischen unveränderten, veränderten und neu hinzugekommenen Risikomerkmalen. Unveränderte Merkmale können aus der bisherigen Bewertung übernommen werden, sofern ihre Annahmen weiterhin gültig sind. Veränderte Merkmale müssen neu begründet werden. Neu hinzugekommene Funktionen, Datenzugriffe, Nutzergruppen, Toolrechte oder Einsatzbereiche sind vollständig in die Bewertung einzubeziehen.

Die Neubewertung sollte dokumentieren, warum sie durchgeführt wurde, welche Systemänderungen vorliegen, welche Dimensionen betroffen sind und ob sich Score, Risikoklasse, Schwellenkriterien oder Mindestanforderungen verändern. Dadurch bleibt nachvollziehbar, wie sich das Risikopotenzial eines KI-Systems über die Zeit entwickelt.

Eine Neubewertung kann zu drei Ergebnissen führen. Erstens kann die bisherige Einstufung bestätigt werden, wenn die Änderungen nicht risikorelevant sind. Zweitens kann sich die Einstufung erhöhen, wenn Autonomie, Severität, Exposition oder Irreversibilität zunehmen. Drittens kann sich die Einstufung verringern, wenn nachweisbare und dauerhafte Begrenzungen oder Schutzmaßnahmen eine niedrigere Bewertung einzelner Dimensionen rechtfertigen. Eine bloß organisatorisch behauptete Schutzmaßnahme reicht dafür jedoch nicht aus; sie muss tatsächlich wirksam, überprüfbar und dauerhaft im System- oder Nutzungskontext verankert sein.

Die Aktualisierung der ASEI-10-Bewertung ist damit ein wesentlicher Bestandteil verantwortbarer KI-Governance. Sie stellt sicher, dass die Risikoeinstufung nicht nur bei Einführung eines Systems vorgenommen wird, sondern den tatsächlichen Entwicklungs- und Nutzungspfad des Systems begleitet.

10.7 Zusammenfassung und Überleitung zu den Fallbeispielen

Das Anwendungskapitel zeigt, wie das ASEI-10-Modell praktisch eingesetzt werden kann. Die Bewertung beginnt nicht mit der Scoreberechnung, sondern mit der eindeutigen Festlegung des Bewertungsgegenstands und der Beschreibung des Systemkontexts. Erst wenn klar ist, welches konkrete KI-System in welchem Nutzungskontext bewertet wird, können Autonomie, Severität, Exposition und Irreversibilität sachgerecht eingestuft werden.

Der Bewertungsprozess verbindet qualitative Analyse und rechnerische Einordnung. Die vier Dimensionen werden zunächst einzeln begründet. Anschließend wird die in Kapitel 9 dargestellte Scorelogik angewendet. Der resultierende ASEI-10-Score fasst das Risikopotenzial zusammen, ersetzt aber nicht die fachliche Begründung. Maßgeblich bleibt stets die Verbindung aus Systembeschreibung, Einzeldimensionen, Score, Schwellenkriterien und Dokumentation.

Die Schwellenprüfung ergänzt die Scorebildung um eine qualitative Kontrolllogik. Sie verhindert, dass besonders relevante Einzelmerkmale durch den Gesamtscore verdeckt werden. Agentische Handlungsausführung, sensible Einsatzbereiche, breite Exposition oder schwer reversible Folgen können zusätzliche Mindestanforderungen auslösen, auch wenn der Score allein noch keine sehr hohe Risikoklasse nahelegt.

Die Dokumentation bildet die Grundlage für Nachvollziehbarkeit, Vergleichbarkeit und spätere Neubewertung. Der ASEI-10-Bewertungsbogen dient dabei als standardisiertes Arbeitsinstrument. Er hält fest, welcher Systemzustand bewertet wurde, welche Annahmen zugrunde lagen, wie die vier Dimensionen begründet wurden, welche Schwellenkriterien ausgelöst wurden und welche Mindestanforderungen sich daraus ergeben.

Da ASEI-10 kontextbezogen arbeitet, ist jede Bewertung an den jeweils beschriebenen Systemzustand gebunden. Verändern sich Funktionen, Berechtigungen, Datenzugriffe, Nutzergruppen, Einsatzbereich, Kontrollmechanismen oder mögliche Folgen, muss geprüft werden, ob eine Aktualisierung oder vollständige Neubewertung erforderlich ist.

Das folgende Kapitel demonstriert die Anwendung des ASEI-10-Modells anhand von sechs Fallbeispielen. Jedes Fallbeispiel folgt derselben Grundlogik: Ausgangssituation, Dimensionseinstufung, Scoreanwendung, Schwellenprüfung und Ergebnisinterpretation. Der vollständige Musterbogen für eine ASEI-10-Bewertung befindet sich in Anhang B. Die ausgefüllten Bewertungsbögen zu den sechs Fallbeispielen sind in Anhang C zusammengeführt.

11 Fallbeispiele zur Anwendung des ASEI-10-Modells

Die folgenden Fallbeispiele demonstrieren die Anwendung des ASEI-10-Modells auf unterschiedliche KI-Systeme und Nutzungskontexte. Sie zeigen, dass das Risikopotenzial eines KI-Systems nicht allein aus der technischen Leistungsfähigkeit eines Modells entsteht, sondern aus dem Zusammenspiel von Systemfunktion, Autonomie, Einsatzbereich, Exposition, Irreversibilität und vorhandenen Kontrollmechanismen.

Die Beispiele sind bewusst unterschiedlich gewählt. Sie reichen von einem privaten, rein assistiven Chatbot über Bildungs- und Kundenserviceanwendungen bis hin zu agentischen Unternehmenssystemen, Personalvorauswahl und KI-Systemen in kritischer Infrastruktur. Dadurch wird sichtbar, dass hohe Risikopotenziale auf unterschiedlichen Wegen entstehen können: durch operative Handlungsausführung, sensible Entscheidungsbereiche, breite Wirkung, schwer reversible Folgen oder deren Kombination.

Jedes Fallbeispiel folgt derselben Grundstruktur. Zunächst wird die Ausgangssituation beschrieben. Anschließend werden die vier ASEI-Dimensionen Autonomie, Severität, Exposition und Irreversibilität begründet eingestuft. Danach wird die in Kapitel 9 dargestellte Scorelogik auf den konkreten Fall angewendet. Ergänzend werden mögliche Prüfschwellen und Mindestanforderungen betrachtet. Abschließend erfolgt eine qualitative Ergebnisinterpretation.

Zusätzlich enthält jedes Fallbeispiel einen ausgefüllten ASEI-10-Bewertungsbogen. Dadurch wird gezeigt, wie die in Kapitel 10 dargestellte Dokumentationsstruktur auf konkrete Nutzungskontexte übertragen werden kann. Die Bewertungsbögen ersetzen nicht die fachliche Interpretation, sondern machen die jeweiligen Annahmen, Einstufungen, Schwellenkriterien und Mindestanforderungen nachvollziehbar.

Die folgenden Fallbeispiele zeigen, wie das ASEI-10-Modell auf unterschiedliche KI-Systeme und Anwendungskontexte angewendet werden kann. Sie dienen nicht nur der rechnerischen Veranschaulichung, sondern auch der methodischen Abgrenzung der vier Bewertungsdimensionen. Dadurch wird sichtbar, dass derselbe technische Systemtyp je nach Autonomie, Einsatzbereich, Exposition und Irreversibilität zu unterschiedlichen Risikoeinstufungen führen kann.

Jedes Fallbeispiel folgt demselben Aufbau: Zunächst wird der konkrete Nutzungskontext beschrieben. Anschließend werden die vier Dimensionen Autonomie, Severität, Exposition und Irreversibilität einzeln bewertet und begründet. Danach werden der multiplikative Rohwert und der ASEI-10-Score berechnet. Abschließend werden die einschlägigen Prüfschwellen betrachtet und das Ergebnis zusammenfassend interpretiert.

Die Fallbeispiele bewerten jeweils den beschriebenen Systemzustand im angegebenen Anwendungskontext. Systemimmanente Begrenzungen, etwa fest eingebaute Freigabepunkte, eingeschränkte Toolzugriffe oder definierte Domänengrenzen, werden als Bestandteil des Bewertungsgegenstands berücksichtigt. Soweit zusätzliche Schutzmaßnahmen beschrieben werden, sind sie nur dann scorewirksam, wenn sie eine ASEI-Dimension tatsächlich verändern. Die Fallbeispiele sind daher als Bewertungen des jeweils konkret beschriebenen Systemdesigns zu verstehen; bei abweichender technischer oder organisatorischer Ausgestaltung wäre eine Neubewertung erforderlich.

Die zugehörigen ASEI-10-Bewertungsbögen werden nicht innerhalb der einzelnen Fallbeispiele abgedruckt, sondern in Anhang C zusammengeführt. Dadurch bleibt das Fallbeispielkapitel auf Ausgangssituation, Dimensionseinstufung, Scoreberechnung, Schwellenprüfung und Ergebnisinterpretation konzentriert, während die formale Dokumentation der Bewertungen im Anhang nachvollzogen werden kann.

11.1 Fallbeispiel 1: Privater Chatbot für kreative Texte

11.1.1 Ausgangssituation

Eine Privatperson nutzt einen textbasierten KI-Chatbot, um kreative Schreibideen zu entwickeln. Das System wird verwendet, um Figurenideen, Dialogansätze, alternative Formulierungen, Gedichtzeilen oder kurze Szenenentwürfe zu erzeugen. Die Nutzung erfolgt ausschließlich privat. Die erzeugten Texte werden nicht automatisch veröffentlicht, nicht an Dritte versendet und nicht in betriebliche, rechtliche, medizinische, finanzielle oder sonstige kritische Entscheidungen übernommen.

Der Chatbot reagiert auf Eingaben des Nutzers. Er führt keine eigenständigen Handlungen außerhalb des Dialogs aus, besitzt keinen Zugriff auf externe Systeme, verändert keine Dateien, versendet keine Nachrichten und löst keine Prozesse aus. Die erzeugten Inhalte können vom Nutzer jederzeit verworfen, überarbeitet oder neu erzeugt werden.

11.1.2 Einstufung der vier Bewertungsdimensionen

Dimension	Einstufung	Begründung
Autonomie	**A2**	Das System ist ein reaktiver Dialogassistent. Es antwortet auf Eingaben des Nutzers, verfolgt aber keine eigenständigen Ziele, nutzt keine Werkzeuge und führt keine Handlungen außerhalb des Dialogs aus.
Severität	**S1**	Der Einsatzbereich ist privat, kreativ und folgenarm. Es werden keine Rechte, Vermögenswerte, personenbezogenen Entscheidungen, Sicherheitsfragen oder kritischen Schutzgüter berührt.
Exposition	**E1**	Die Wirkung bleibt auf eine einzelne Person und eine private Nutzungssituation begrenzt. Eine automatische Weitergabe, Veröffentlichung oder Einbindung in größere Prozesse findet nicht statt.
Irreversibilität	**I1**	Fehlerhafte oder unpassende Ergebnisse können sofort verworfen, korrigiert oder neu erzeugt werden. Relevante Restschäden sind nicht zu erwarten.

Tabelle 20: ASEI-10-Einstufung für Fallbeispiel 1

11.1.3 Berechnung des fallbezogenen ASEI-10-Scores

Aus der qualitativen Einstufung ergibt sich zunächst: A = 2; S = 1; E = 1; I = 1.

Die Autonomiestufe A2 wird für die Scorebildung in den normierten Autonomiewert A_n überführt: $\mathbf{A_n = 1{,}5}$

Der Rohwert lautet: $\boldsymbol{R = 1{,}5 \cdot 1 \cdot 1 \cdot 1 = 1{,}5}$ und wird anschließend logarithmisch normiert:

$$\text{ASEI-10} = 1 + 9 \cdot \frac{\ln(1{,}5)}{\ln(625)} \approx 1{,}6$$

Nach logarithmischer Normierung ergibt sich ein ASEI-10-Score von rund 1,6. Der Score liegt damit in der Risikoklasse RK1.

11.1.4 Prüfung der Schwellenkriterien

In diesem Fallbeispiel werden keine höheren Schwellenkriterien ausgelöst. Das System bleibt innerhalb der assistiven Autonomieklasse A1–A3, der Einsatzbereich ist folgenarm, die Exposition bleibt individuell und mögliche Fehler sind leicht reversibel.

Es gelten daher lediglich die Basisanforderungen:

- Plausibilitätsprüfung durch den Nutzer
- keine ungeprüfte Übernahme sachlicher Aussagen
- bewusster Umgang mit möglichen Halluzinationen oder stilistischen Unstimmigkeiten
- keine Veröffentlichung oder Weiterverwendung in sensiblen Kontexten ohne erneute Prüfung

Eine besondere Hochrisiko-, Expositions- oder Irreversibilitätsprüfung ist in diesem Nutzungskontext nicht erforderlich.

11.1.5 Ergebnisinterpretation

Das Fallbeispiel erreicht mit einem ASEI-10-Score von rund 1,6 ein sehr niedriges Risikopotenzial. Dieses Ergebnis ist plausibel, weil alle Risikodimensionen niedrig ausgeprägt sind. Der Chatbot antwortet lediglich dialogisch-reaktiv, nutzt keine Werkzeuge, führt keine Handlungen aus und bleibt auf eine private, kreative Einzelnutzung beschränkt. Fehlerhafte Ergebnisse können ohne relevante Folgeschäden korrigiert, verworfen oder ignoriert werden.

Gleichzeitig zeigt das Beispiel, dass auch eine sehr niedrige Einstufung stets an den konkreten Nutzungskontext gebunden ist. Wird derselbe Chatbot für veröffentlichte Inhalte, professionelle Kommunikation, Prüfungsleistungen oder automatisierte Verbreitung eingesetzt, verändert sich der Anwendungskontext. In solchen Fällen müssten insbesondere Severität und Exposition, gegebenenfalls auch Irreversibilität, neu bewertet werden.

Das Beispiel verdeutlicht damit einen Grundgedanken des ASEI-10-Modells: Bewertet wird nicht der Chatbot abstrakt, sondern die konkrete Nutzungssituation mit ihren Funktionen, Grenzen, Wirkungen und möglichen Folgen.

11.2 Fallbeispiel 2: KI-Assistent für Unterrichts- und Prüfungsvorbereitung

11.2.1 Ausgangssituation

Ein Dozent nutzt einen KI-Assistenten zur Vorbereitung von Unterrichtsmaterialien und prüfungsnahen Übungsaufgaben. Das System unterstützt bei der Entwicklung von Fallstudien, Arbeitsaufträgen, Musterlösungen, Rechenschemata, Glossaren, Zusammenfassungen und didaktischen Erklärungen. Zusätzlich können vorhandene Unterlagen, Skripte oder Aufgabenentwürfe hochgeladen und vom System analysiert, strukturiert oder sprachlich überarbeitet werden.

Die Nutzung erfolgt durch den Dozenten auf konkrete Anweisung. Der KI-Assistent erstellt Vorschläge, entscheidet aber nicht selbstständig über Lernziele, Prüfungsinhalte, Bewertungen oder Bestehen und Nichtbestehen von Prüfungsteilnehmern. Alle Materialien werden vor einer Verwendung im Unterricht oder in der Prüfungsvorbereitung fachlich geprüft und gegebenenfalls überarbeitet. Eine automatische Veröffentlichung oder direkte Ausspielung an Teilnehmende findet nicht statt.

Das System kann Werkzeuge nutzen, etwa Datei-Analyse, Rechenfunktionen, Tabellenbearbeitung oder Recherchefunktionen. Diese Werkzeuge werden jedoch nur auf ausdrückliche Anweisung des Nutzers eingesetzt. Der KI-Assistent besitzt keine eigenständige Prozesssteuerung, keine dauerhafte Aktivität und keine Berechtigung, ohne Freigabe Inhalte an Lernende oder externe Stellen zu versenden.

11.2.2 Einstufung der vier Bewertungsdimensionen

Dimension	Einstufung	Begründung
Autonomie	**A3**	Das System arbeitet als Werkzeugassistenz auf Nutzeranweisung. Es kann Dateien analysieren, Aufgaben entwerfen, Berechnungen unterstützen oder Materialien strukturieren, handelt aber nicht eigenständig und löst keine Prozesse ohne Anweisung aus.
Severität	**S3**	Der Einsatz ist bildungs- und prüfungsvorbereitungsbezogen. Fehlerhafte Materialien können Lernende fehlleiten, Prüfungsstoff verzerren oder fachliche Fehlvorstellungen erzeugen. Da das System jedoch keine Prüfungsentscheidungen trifft und keine Leistungen bewertet, liegt noch kein hochkritischer Bildungsentscheidungsbereich im Sinne von S4 vor.
Exposition	**E2**	Die Wirkung betrifft eine abgegrenzte Kurs- oder Lerngruppe. Die Materialien werden nicht öffentlich verbreitet und nicht organisationsweit automatisiert eingesetzt.
Irreversibilität	**I3**	Fehler können grundsätzlich korrigiert werden, verursachen aber möglicherweise Nacharbeit, Vertrauensverlust, Fehlvorbereitung oder didaktische Folgeschäden. Eine Korrektur ist möglich, aber nicht immer folgenlos.

Tabelle 21: ASEI-10-Einstufung für Fallbeispiel 2

11.2.3 Berechnung des fallbezogenen ASEI-10-Scores

Aus der qualitativen Einstufung ergibt sich zunächst: A = 3; S = 3; E = 2; I = 3.

Die Autonomiestufe A3 wird für die Scorebildung in den normierten Autonomiewert A_n überführt: $\mathbf{A_n = 2{,}0}$

Der Rohwert lautet $\boldsymbol{R = 2{,}0 \cdot 3 \cdot 2 \cdot 3 = 36}$ und wird anschließend logarithmisch normiert:

$$\text{ASEI-10} = 1 + 9 \cdot \frac{\ln(36)}{\ln(625)} \approx 6{,}0$$

Nach logarithmischer Normierung ergibt sich ein ASEI-10-Score von rund 6,0. Der Score liegt damit in der Risikoklasse RK4.

11.2.4 Prüfung der Schwellenkriterien

In diesem Fallbeispiel werden mehrere mittlere Schwellenkriterien ausgelöst. Die Autonomie bleibt zwar noch innerhalb der assistiven Klasse A1–A3, weil das System nur auf Nutzeranweisung arbeitet. Durch die Werkzeugnutzung ist jedoch besondere Aufmerksamkeit bei Quellen, Berechnungen, Dateiauswertungen und fachlicher Plausibilität erforderlich.

Die Severität erreicht S3, weil die Materialien personenbezogene Lernprozesse und prüfungsnahe Vorbereitung beeinflussen können. Daraus ergeben sich Anforderungen an fachliche Kontrolle, didaktische Prüfung, sachliche Richtigkeit und klare Abgrenzung zwischen Übungsmaterial und offizieller Prüfung. Wenn das System dagegen zur Bewertung von Prüfungsleistungen, zur Auswahl von Prüflingen, zur Notengebung oder zur verbindlichen Entscheidung über Bestehen und Nichtbestehen eingesetzt würde, wäre regelmäßig S4 zu prüfen.

Die Exposition bleibt bei E2, solange die Materialien nur in einer abgegrenzten Kursgruppe verwendet werden. Bei organisationsweiter Nutzung, Veröffentlichung in einem Lernmanagementsystem, Verkauf als Lernprodukt oder automatisierter Verteilung an große Teilnehmergruppen wäre eine höhere Expositionsstufe zu prüfen.

Die Irreversibilität erreicht I3, weil fehlerhafte Lernmaterialien zwar korrigiert werden können, aber bereits entstandene Fehlvorstellungen, Zeitverluste oder Vertrauensschäden nicht immer vollständig folgenlos rückholbar sind.

Als Mindestanforderungen ergeben sich daher insbesondere:

- fachliche Prüfung aller KI-generierten Inhalte vor Verwendung
- Nachrechnen und Kontrollieren von Musterlösungen, Formeln und Rechenschritten
- Kennzeichnung oder interne Dokumentation der KI-Unterstützung
- keine ungeprüfte Übernahme von Rechts-, Prüfungs- oder Normenaussagen
- klare Trennung zwischen Übungsmaterial und verbindlicher Prüfungsbewertung
- Korrekturhinweis an Lernende, falls fehlerhafte Materialien bereits verwendet wurden

11.2.5 Ergebnisinterpretation

Das Fallbeispiel erreicht mit einem ASEI-10-Score von rund 6,0 ein erhöhtes Risikopotenzial. Dieses Ergebnis ist plausibel, weil das System zwar nicht autonom handelt, aber in einem prüfungsnahen Bildungszusammenhang eingesetzt wird und fehlerhafte Ergebnisse reale Lern- und Vorbereitungseffekte haben können.

Das Risikopotenzial entsteht hier nicht primär aus hoher Autonomie, sondern aus dem Zusammenspiel von Werkzeugnutzung, bildungsbezogener Folgenrelevanz, gruppenbezogener Exposition und begrenzter Reversibilität. Eine fehlerhafte private Formulierung wäre leicht folgenlos korrigierbar. Eine fehlerhafte Musterlösung, die in einer Kursgruppe verwendet wird, kann dagegen Lernende systematisch auf eine falsche Spur führen und sich innerhalb der Gruppe vervielfältigen.

Deshalb reicht eine rein private Plausibilitätsprüfung in diesem Anwendungskontext nicht aus. Erforderlich sind fachliche Endkontrolle, dokumentierte Verantwortung und klare Einsatzgrenzen. Der KI-Assistent darf nicht als ungeprüfte Quelle für verbindliche Musterlösungen, Prüfungsinhalte oder Leistungsbewertungen verwendet werden.

Gleichzeitig bleibt der Fall deutlich unterhalb eines hochkritischen Prüfungssystems. Solange der KI-Assistent nur der Vorbereitung dient, keine Prüfungsleistungen bewertet und keine verbindlichen Bildungsentscheidungen vorbereitet oder trifft, ist S3 ausreichend. Sobald das System jedoch in offizielle Prüfungsbewertung, Leistungsdiagnostik oder Zulassungsentscheidungen eingebunden wird, müsste die Einstufung neu vorgenommen werden. In diesem Fall wären insbesondere S4 und gegebenenfalls eine höhere Irreversibilitätsstufe zu prüfen.

11.3 Fallbeispiel 3: KI-System im Kundenservice

11.3.1 Ausgangssituation

Ein Unternehmen setzt auf seiner Website einen KI-gestützten Kundenservice-Assistenten ein. Das System beantwortet häufige Kundenfragen, gibt Auskunft zu Produkten, Lieferzeiten, Retouren, Reklamationen und Vertragsbedingungen und kann auf ausgewählte Kundendaten zugreifen, etwa Bestellnummern, Lieferstatus, Kontaktinformationen oder den Verlauf früherer Serviceanfragen. Zusätzlich kann das System Support-Tickets erstellen und Anfragen an menschliche Servicemitarbeiter weiterleiten.

Der Kundenservice-Assistent ist öffentlich erreichbar und wird von einer großen Zahl von Kundinnen und Kunden genutzt. Die Antworten erscheinen unmittelbar im Dialog mit dem Kunden. Das System kann Informationen aus internen Wissensdatenbanken abrufen, einfache Fallklassifikationen vornehmen und Textbausteine für Antworten erzeugen. Es darf jedoch keine rechtsverbindlichen Vertragsänderungen durchführen, keine Zahlungen auslösen, keine Erstattungen eigenständig genehmigen und keine Kundenkonten ohne menschliche Freigabe verändern.

Das System arbeitet überwiegend dialogisch und werkzeuggestützt. Es reagiert auf Kundenanfragen, ruft Informationen aus angebundenen Systemen ab und erstellt Vorschläge oder Antworten innerhalb des Serviceprozesses. Kritische Fälle, etwa rechtliche Beschwerden, Zahlungsfragen, komplexe Reklamationen oder eskalierte Konflikte, werden an menschliche Mitarbeitende übergeben.

11.3.2 Einstufung der vier Bewertungsdimensionen

Dimension	Einstufung	Begründung
Autonomie	**A3**	Das System nutzt Werkzeuge und interne Datenquellen auf Grundlage konkreter Kundenanfragen. Es handelt jedoch nicht eigenständig außerhalb des Dialogs, führt keine verbindlichen Vertrags- oder Zahlungsaktionen aus und bleibt auf assistive Werkzeugnutzung begrenzt.
Severität	**S3**	Der Einsatz betrifft personenbezogene Daten, Kundenbeziehungen, Bestellungen, Reklamationen und wirtschaftlich relevante Vorgänge. Fehler können zu falschen Kundeninformationen, Vertrauensverlust, Datenschutzproblemen oder wirtschaftlichen Nachteilen führen.
Exposition	**E4**	Das System ist öffentlich erreichbar und kann große Kundengruppen betreffen. Fehlerhafte Antworten können sich massenhaft in der Kundenkommunikation auswirken und reputationsrelevant werden.
Irreversibilität	**I3**	Fehlerhafte Kundeninformationen oder unzutreffende Auskünfte können grundsätzlich korrigiert werden, verursachen aber Nacharbeit, Beschwerden, Vertrauensverlust und organisatorische Folgewirkungen.

Tabelle 22: ASEI-10-Einstufung für Fallbeispiel 3

11.3.3 Berechnung des fallbezogenen ASEI-10-Scores

Aus der qualitativen Einstufung ergibt sich zunächst: A = 3; S = 3; E = 4; I = 3.

Die Autonomiestufe A3 wird für die Scorebildung in den normierten Autonomiewert A_n überführt: $\mathbf{A_n = 2{,}0}$

Der Rohwert lautet $\boldsymbol{R = 2{,}0 \cdot 3 \cdot 4 \cdot 3 = 72}$ und wird anschließend logarithmisch normiert:

$$\textbf{ASEI-10} = 1 + 9 \cdot \frac{\ln(72)}{\ln(625)} \approx 7{,}0 \textbf{ (ungerundet 6,98).}$$

Nach logarithmischer Normierung ergibt sich ein ungerundeter ASEI-10-Score von etwa 6,98. Gerundet entspricht dies einem Score von 7,0. Da die Einordnung nach dem ungerundeten Wert erfolgt, liegt das System formal noch in der Risikoklasse RK4.

11.3.4 Prüfung der Schwellenkriterien

In diesem Fallbeispiel werden mehrere relevante Schwellenkriterien ausgelöst. Die Autonomie bleibt mit A3 noch innerhalb der assistiven Klasse, weil das System zwar Werkzeuge nutzt, aber keine eigenständige Prozesssteuerung oder autonome Handlungsausführung besitzt. Dennoch erfordert die Werkzeugnutzung klare Begrenzungen, insbesondere bei Zugriffen auf Kundendaten, interne Wissensdatenbanken und Ticketsysteme.

Die Severität erreicht S3, weil personenbezogene Daten, Kundenbeziehungen, Bestellungen, Reklamationen und wirtschaftlich relevante Vorgänge betroffen sind. Daraus ergeben sich Anforderungen an Datenschutz, Vertraulichkeit, fachliche Kontrolle, sachliche Richtigkeit und klare Eskalationsregeln. Wenn das System rechtsverbindliche Auskünfte erteilen, Vertragsänderungen auslösen, Erstattungen genehmigen oder Kundenansprüche verbindlich ablehnen dürfte, wäre S4 zu prüfen.

Die Exposition erreicht E4, weil das System öffentlich erreichbar ist und große Kundengruppen betreffen kann. Fehlerhafte Antworten können nicht nur einzelne Kunden irritieren, sondern sich über viele Servicekontakte hinweg wiederholen. Dadurch entstehen besondere Anforderungen an Monitoring, Qualitätssicherung, Freigabe der Wissensbasis, Korrekturprozesse und Kommunikationsfähigkeit im Fehlerfall.

Die Irreversibilität erreicht I3, weil falsche Auskünfte zwar grundsätzlich berichtigt werden können, aber nicht immer folgenlos bleiben. Kunden können aufgrund fehlerhafter Informationen falsche Entscheidungen treffen, Fristen versäumen, Beschwerden einreichen oder das Vertrauen in das Unternehmen verlieren. Die Folgen sind korrigierbar, aber mit Aufwand und möglichen Restwirkungen verbunden.

Als Mindestanforderungen ergeben sich insbesondere:

- klare Begrenzung der Systemrechte und Datenzugriffe
- keine eigenständigen Vertrags-, Zahlungs- oder Kontoänderungen ohne Freigabe
- Datenschutz- und Vertraulichkeitsprüfung
- geprüfte und versionierte Wissensbasis
- Eskalationsregeln für rechtliche, finanzielle und konfliktträchtige Fälle
- Monitoring typischer Fehlantworten und Beschwerdemuster
- Korrekturverfahren für fehlerhafte Kundeninformationen
- Protokollierung relevanter Kundeninteraktionen

11.3.5 Ergebnisinterpretation

Das Fallbeispiel erreicht mit einem ungerundeten ASEI-10-Score von etwa 6,98 ein erhöhtes Risikopotenzial. Zugleich liegt es unmittelbar an der Schwelle zur hohen Risikoklasse. Diese Grenzlage ist fachlich plausibel, weil das Kundenservice-System zwar keine verbindlichen Handlungen eigenständig ausführt, aber gegenüber vielen Kundinnen und Kunden wirkt, personenbezogene oder wirtschaftlich relevante Informationen verarbeitet und fehlerhafte Auskünfte breit wiederholen kann.

Das Risikopotenzial wird nicht durch besonders hohe Autonomie getragen, sondern vor allem durch Severität, Exposition und Irreversibilität. Ein einzelner fehlerhafter Dialog wäre für sich genommen möglicherweise beherrschbar. Wenn sich dieselbe fehlerhafte Antwort jedoch massenhaft gegenüber vielen Kunden wiederholt, kann daraus ein erhebliches Reputations-, Datenschutz- oder Serviceproblem entstehen.

Die gerundete Darstellung von 7,0 darf deshalb nicht mechanisch interpretiert werden. Maßgeblich bleibt der ungerundete Wert in Verbindung mit der fachlichen Begründung. In der praktischen Governance sollte das System als grenznaher Fall behandelt werden, bei dem erhöhte Kontroll-, Dokumentations- und Monitoringanforderungen sachgerecht sind.

Gleichzeitig bleibt das System unterhalb einer höheren Autonomiestufe, solange es keine verbindlichen Handlungen eigenständig ausführt. Würde der Kundenservice-Assistent eigenständig Erstattungen genehmigen, Verträge ändern, Kundenkonten sperren, Forderungen ablehnen oder rechtlich relevante Entscheidungen treffen, müsste die Einstufung neu vorgenommen werden. In diesem Fall wären insbesondere A5, S4 und gegebenenfalls eine höhere Irreversibilitätsstufe zu prüfen.

11.4 Fallbeispiel 4: KI-Agent mit Toolzugriff in einem Unternehmen

11.4.1 Ausgangssituation

Ein mittelständisches Unternehmen setzt einen internen KI-Agenten ein, um operative Büro-, Projekt- und Verwaltungsprozesse zu unterstützen. Der Agent kann auf freigegebene Dokumentenablagen, Kalender, E-Mail-Postfächer, Aufgabenlisten, interne Wissensdatenbanken und ein Projektmanagementsystem zugreifen. Er wird genutzt, um Besprechungen vorzubereiten, Aufgaben nachzuverfolgen, Protokolle zu erstellen, Fristen zu überwachen, interne Informationen zusammenzuführen und einfache Prozessschritte auszulösen.

Der Agent arbeitet nicht nur dialogisch auf einzelne Fragen, sondern kann aus einem Zielauftrag mehrere Teilschritte ableiten. Er kann etwa eine Projektbesprechung vorbereiten, relevante Unterlagen suchen, offene Aufgaben identifizieren, eine Tagesordnung erstellen, Terminvorschläge prüfen, interne Erinnerungen formulieren und Projektstatusinformationen zusammenstellen. Innerhalb eines definierten Berechtigungsrahmens darf er Aufgabenlisten aktualisieren, Kalendertermine vorschlagen, interne Entwürfe vorbereiten und standardisierte Benachrichtigungen versenden.

Das System darf keine Zahlungen auslösen, keine Verträge abschließen, keine Personalentscheidungen treffen, keine Kundenzusagen verbindlich abgeben und keine rechtlich erheblichen Erklärungen ohne menschliche Freigabe versenden. Dennoch besitzt es operative Handlungsmöglichkeiten innerhalb interner Unternehmensprozesse. Fehlerhafte Aktionen können daher nicht nur falsche Texte erzeugen, sondern Arbeitsabläufe beeinflussen, Informationen falsch priorisieren, Fristen verschieben oder interne Abstimmungen stören.

11.4.2 Einstufung der vier Bewertungsdimensionen

Dimension	Einstufung	Begründung
Autonomie	**A5**	Das System handelt als begrenzter autonomer Agent. Es kann Zielvorgaben in mehrere Handlungsschritte übersetzen, Werkzeuge selbst auswählen und innerhalb einer definierten Domäne operative Aktionen ohne Einzelfreigabe ausführen.
Severität	**S3**	Der Einsatz betrifft interne Unternehmensprozesse, personenbezogene Daten, Projektinformationen, Fristen, Aufgaben, Kommunikation und wirtschaftlich relevante Abläufe. Hochkritische Personal-, Rechts-, Finanz- oder Sicherheitsentscheidungen sind im beschriebenen Rahmen ausgeschlossen.
Exposition	**E3**	Die Wirkung betrifft eine Organisationseinheit, Projektteams, interne Prozesse und freigegebene Datenbestände. Eine organisationsweite oder öffentliche Exposition liegt noch nicht zwingend vor.
Irreversibilität	**I3**	Fehlerhafte Aktionen sind grundsätzlich korrigierbar, können aber Nacharbeit, Verzögerungen, Informationsverluste, Vertrauensschäden oder organisatorische Folgewirkungen verursachen.

Tabelle 23: ASEI-10-Einstufung für Fallbeispiel 4

11.4.3 Berechnung des fallbezogenen ASEI-10-Scores

Aus der qualitativen Einstufung ergibt sich zunächst: A = 5; S = 3; E = 3; I = 3.

Die Autonomiestufe A5 wird für die Scorebildung in den normierten Autonomiewert A_n überführt: $\mathbf{A_n = 3{,}0}$

Der Rohwert lautet $R = 3{,}0 \cdot 3 \cdot 3 \cdot 3 = 81$ und wird anschließend logarithmisch normiert:

$$\textbf{ASEI-10} = \mathbf{1} + \mathbf{9} \cdot \frac{\ln(81)}{\ln(625)} \approx \mathbf{7{,}1}$$

Nach logarithmischer Normierung ergibt sich ein ASEI-10-Score von rund 7,1. Der Score liegt damit in der Risikoklasse RK5.

11.4.4 Prüfung der Schwellenkriterien

In diesem Fallbeispiel werden mehrere relevante Schwellenkriterien ausgelöst. Die Autonomie erreicht A5, weil der Agent nicht nur auf einzelne Nutzeranweisungen mit Werkzeugnutzung reagiert, sondern innerhalb einer definierten Domäne selbstständig Handlungsschritte ausführen kann. Damit liegt autonome Handlungsausführung innerhalb eines begrenzten Rahmens vor. Entscheidend ist, dass nicht mehr jede einzelne Aktion individuell freigegeben werden muss, solange sie innerhalb der vorab definierten Berechtigungen liegt.

Die Severität erreicht S3, weil interne Unternehmensprozesse, personenbezogene Daten, vertrauliche Informationen, Fristen, Projektstände und wirtschaftlich relevante Abläufe betroffen sind. Fehler können zu falscher Priorisierung, Informationsverlust, Vertraulichkeitsproblemen, Fehlkommunikation oder operativen Störungen führen. Da der Agent keine rechtsverbindlichen Erklärungen, Personalentscheidungen, Zahlungen oder sicherheitskritischen Aktionen ausführen darf, ist S4 in der beschriebenen Ausgangssituation noch nicht erreicht.

Die Exposition erreicht E3, weil die Wirkungen über eine einzelne Person oder kleine Gruppe hinausgehen und in interne Prozesse, Projektteams, Datenbestände und Organisationseinheiten hineinwirken. Wird der Agent dagegen organisationsweit für alle Mitarbeitenden eingeführt, an zentrale Unternehmenssysteme angebunden oder in großflächige Kunden-, Finanz- oder Personalprozesse integriert, wäre E4 zu prüfen.

Die Irreversibilität erreicht I3, weil fehlerhafte Aktionen grundsätzlich korrigiert werden können, aber nicht zwingend folgenlos bleiben. Eine falsch gesetzte Frist, eine irreführende interne Nachricht, eine unvollständige Projektzusammenfassung oder eine fehlerhafte Aufgabenpriorisierung kann Nacharbeit, Verzögerung oder Vertrauensverlust verursachen. Die Folgen bleiben meist korrigierbar, können aber operative Restwirkungen haben.

Als Mindestanforderungen ergeben sich insbesondere:

- klar definiertes Rollen- und Rechtekonzept
- Begrenzung der erlaubten Werkzeuge und Systemzugriffe
- Freigabepflicht für rechtlich, finanziell, personell oder extern relevante Handlungen
- Protokollierung aller agentischen Aktionen
- nachvollziehbare Aufgaben- und Entscheidungsketten
- Möglichkeit zur menschlichen Übersteuerung
- Abschalt- oder Deaktivierungsmechanismus
- Eskalationsregeln bei Unsicherheit, Konflikten oder sensiblen Vorgängen
- regelmäßige Überprüfung der Berechtigungen und Fehlhandlungen

11.4.5 Ergebnisinterpretation

Das Fallbeispiel erreicht mit einem ASEI-10-Score von rund 7,1 ein hohes Risikopotenzial. Dieses Ergebnis zeigt, dass bereits ein begrenzter Unternehmensagent deutlich risikorelevant sein kann, auch wenn er nicht in einem hochkritischen Einsatzbereich verwendet wird. Ausschlaggebend ist vor allem der Übergang von bloßer Werkzeugassistenz zu begrenzter agentischer Handlungsausführung.

Das System kann nicht nur unterstützen, sondern innerhalb definierter Grenzen operative Vorgänge beeinflussen, Informationen weiterleiten, Aufgaben anstoßen oder Prozesse verändern. Dadurch entsteht ein anderes Risikoprofil als bei einem rein reaktiven Assistenzsystem. Das Risiko liegt weniger in einem einzelnen falschen Text, sondern in der Fähigkeit des Systems, Arbeitsabläufe zu beeinflussen und Fehler in Prozessketten fortzusetzen.

Ein fehlerhafter Agent kann Aufgaben falsch priorisieren, Informationen unvollständig weitergeben, Fristen verändern, interne Nachrichten auslösen oder Dokumente fehlerhaft einordnen. Auch wenn solche Folgen grundsätzlich korrigierbar bleiben, können sie organisatorische Folgewirkungen erzeugen und zusätzlichen Kontroll-, Klärungs- oder Korrekturaufwand verursachen.

Gleichzeitig bleibt das System unterhalb der höchsten Risikobereiche, solange seine Handlungsmöglichkeiten klar begrenzt sind. Würde der Agent dauerhaft ereignisgesteuert aktiv bleiben, eigenständig auf neue Systemzustände reagieren oder über längere Zeit ohne erneute Nutzerinteraktion Prozesse überwachen, wäre A6 zu prüfen. Würde er organisationsweit ausgerollt, wäre E4 naheliegend. Dürfte er Zahlungen, Verträge, Personalmaßnahmen, Kundenentscheidungen oder sicherheitsrelevante Aktionen ausführen, wären S4, I4 und gegebenenfalls weitere Schwellenkriterien zu prüfen.

Das Beispiel verdeutlicht damit den zentralen Unterschied zwischen Werkzeugassistenz und agentischer Unternehmensautomation. Mit zunehmender Autonomie verschiebt sich die Kontrolle vom einzelnen Befehl zur vorgelagerten Systemgestaltung. Entscheidend werden daher Berechtigungsarchitektur, Freigabepunkte, Protokollierung, Monitoring, Eskalation und Rückholbarkeit.

11.5 Fallbeispiel 5: KI-System zur Personalvorauswahl

11.5.1 Ausgangssituation

Ein Unternehmen setzt ein KI-System ein, um eingehende Bewerbungen für offene Stellen vorzustrukturieren. Das System analysiert Lebensläufe, Anschreiben, Qualifikationsprofile, Zeugnisse und weitere Bewerbungsunterlagen. Es extrahiert relevante Merkmale, gleicht diese mit dem Anforderungsprofil der ausgeschriebenen Stelle ab und erstellt eine Rangfolge oder Vorauswahlliste geeigneter Bewerberinnen und Bewerber.

Die endgültige Entscheidung über Einladung, Ablehnung oder Einstellung liegt formal bei der Personalabteilung. Das KI-System versendet keine Absagen, führt keine Bewerbungsgespräche und trifft keine rechtlich verbindliche Einstellungsentscheidung. Seine Vorschläge haben jedoch erheblichen Einfluss darauf, welche Bewerbungen von den Personalverantwortlichen vorrangig geprüft werden und welche Bewerberinnen und Bewerber realistische Chancen auf ein Vorstellungsgespräch erhalten.

Das System verarbeitet personenbezogene und beruflich sensible Daten. Fehlerhafte Gewichtungen, unvollständige Datenauswertung, stereotype Muster, Verzerrungen in Trainingsdaten oder unangemessene Kriterien können dazu führen, dass geeignete Personen benachteiligt oder ungeeignete Personen bevorzugt werden. Auch wenn die menschliche Letztentscheidung erhalten bleibt, kann die KI-gestützte Vorstrukturierung den Entscheidungsprozess faktisch stark prägen.

11.5.2 Einstufung der vier Bewertungsdimensionen

Dimension	Einstufung	Begründung
Autonomie	**A4**	Das System strukturiert Bewerbungen mehrstufig vor, analysiert Unterlagen, vergleicht Profile mit Anforderungen und erzeugt Rangfolgen oder Vorschlagslisten. Die endgültige Entscheidung verbleibt jedoch beim Menschen; verbindliche Handlungen erfolgen nicht ohne menschliche Freigabe.
Severität	**S4**	Der Einsatz betrifft berufliche Chancen, Zugang zu Beschäftigung, personenbezogene Daten und potenzielle Benachteiligungen. Fehler können erhebliche Auswirkungen auf Lebensläufe, Erwerbschancen und Gleichbehandlung haben.
Exposition	**E3**	Die Wirkung betrifft einen abgegrenzten Recruitingprozess, eine Personalabteilung, Bewerbergruppen oder bestimmte Stellenbesetzungen. Eine organisationsweite oder öffentliche Exposition liegt in der beschriebenen Ausgangssituation noch nicht vor.
Irreversibilität	**I4**	Fehlerhafte Vorauswahlentscheidungen können auch nachträglich nur schwer vollständig korrigiert werden. Verpasste Chancen, Diskriminierung, Vertrauensverlust oder rechtliche Folgen können bestehen bleiben, selbst wenn Fehler später erkannt werden.

Tabelle 24: ASEI-10-Einstufung für Fallbeispiel 5

11.5.3 Berechnung des fallbezogenen ASEI-10-Scores

Aus der qualitativen Einstufung ergibt sich zunächst: A = 4; S = 4; E = 3; I = 4.

Die Autonomiestufe A4 wird für die Scorebildung in den normierten Autonomiewert A_n überführt: $\mathbf{A_n = 2{,}5}$

Der Rohwert lautet $\boldsymbol{R = 2{,}5 \cdot 4 \cdot 3 \cdot 4 = 120}$ und wird anschließend logarithmisch normiert:

$$\text{ASEI-10} = 1 + 9 \cdot \frac{\ln(120)}{\ln(625)} \approx 7{,}7$$

Nach logarithmischer Normierung ergibt sich ein ASEI-10-Score von rund 7,7. Der Score liegt damit in der Risikoklasse RK5.

11.5.4 Prüfung der Schwellenkriterien

In diesem Fallbeispiel werden mehrere gewichtige Schwellenkriterien ausgelöst. Die Autonomie erreicht A4, weil das System Bewerbungen nicht nur passiv analysiert, sondern einen mehrstufigen Prozess der Vorstrukturierung durchführt. Es ordnet Informationen, bewertet Profile im Verhältnis zu Anforderungen und erzeugt eine Vorauswahl oder Rangfolge. Damit liegt agentische Prozesssteuerung vor, auch wenn die Letztentscheidung beim Menschen verbleibt.

Die Severität erreicht S4, weil der Einsatzbereich berufliche Chancen, Bewerbungsverfahren und potenzielle Benachteiligungen betrifft. Auch eine nur vorbereitende KI-Bewertung kann faktisch erheblichen Einfluss auf die Entscheidung haben. Wenn Personalverantwortliche die Rangfolge des Systems unkritisch übernehmen, entsteht ein Automatisierungsbias: Die menschliche Letztentscheidung besteht formal weiter, wird aber praktisch durch die KI-Vorentscheidung geprägt.

Die Exposition erreicht E3, weil die Wirkung über einzelne Personen hinausgeht und einen organisatorischen Teilprozess betrifft. Die Einstufung bleibt bei E3, solange das System nur für bestimmte Stellen, Bewerbergruppen oder eine abgegrenzte Personalabteilung eingesetzt wird. Wird das System dagegen organisationsweit für alle Bewerbungsprozesse, über mehrere Standorte hinweg oder bei sehr großen Bewerberzahlen verwendet, wäre E4 zu prüfen.

Die Irreversibilität erreicht I4, weil fehlerhafte Vorauswahlentscheidungen nicht vollständig folgenlos korrigiert werden können. Wird eine geeignete Person nicht eingeladen, kann diese Chance später häufig nicht realistisch wiederhergestellt werden. Auch Diskriminierung, Vertrauensverlust, Reputationsschäden und rechtliche Auseinandersetzungen können bestehen bleiben, selbst wenn der Fehler nachträglich erkannt wird.

Als Mindestanforderungen ergeben sich insbesondere:

- klare menschliche Letztentscheidung mit echter Prüfung, nicht nur formaler Bestätigung
- dokumentierte Kriterien für die Bewerbungsanalyse
- regelmäßige Prüfung auf Verzerrungen und Diskriminierungsrisiken
- Nachvollziehbarkeit der Rangfolge oder Empfehlung
- Möglichkeit zur menschlichen Korrektur und Übersteuerung
- keine automatische Ablehnung ohne qualifizierte menschliche Prüfung
- Datenschutz- und Vertraulichkeitsprüfung
- Beschwerde- und Überprüfungsverfahren für betroffene Personen
- regelmäßige Evaluation der Trefferqualität und Fehlklassifikationen

11.5.5 Ergebnisinterpretation

Das Fallbeispiel erreicht mit einem ASEI-10-Score von rund 7,7 ein hohes Risikopotenzial. Dieses Ergebnis ist plausibel, weil das System in einem besonders schutzwürdigen Entscheidungsumfeld eingesetzt wird. Die Einstufung wird vor allem durch den hochsensiblen Beschäftigungskontext, die Vorstrukturierung von Auswahlentscheidungen und die schwer reversiblen Folgen möglicher Benachteiligungen getragen.

Das Risikopotenzial entsteht nicht aus vollständig autonomem Handeln, sondern aus der Kombination von agentischer Vorstrukturierung, hoher Severität, organisationaler Reichweite und begrenzter Rückholbarkeit negativer Folgen. Eine fehlerhafte oder verzerrte Vorauswahl kann berufliche Chancen beeinflussen, Bewerberinnen und Bewerber benachteiligen und Entscheidungsprozesse prägen, bevor eine menschliche Letztentscheidung getroffen wird.

Besonders bedeutsam ist, dass die menschliche Letztentscheidung den Risikowert nicht automatisch stark reduziert. Wenn die KI-Vorauswahl den Entscheidungsprozess faktisch dominiert, muss das System dennoch als hoch risikorelevant behandelt werden. Entscheidend ist nicht nur, wer formal entscheidet, sondern wie stark die KI die Wahrnehmung, Priorisierung und Auswahl der Personalverantwortlichen beeinflusst.

Das System muss daher nicht erst vollständig autonom entscheiden, um ein hohes Risikopotenzial zu erreichen. Schon die strukturierende Wirkung auf Bewerbungsprozesse kann ausreichen, wenn sie in einem sensiblen Beschäftigungskontext erfolgt und mögliche Benachteiligungen nur schwer vollständig korrigiert werden können.

Würde das System zusätzlich automatische Absagen versenden, Bewerberinnen und Bewerber ohne menschliche Prüfung aussortieren oder organisationsweit in allen Recruitingprozessen eingesetzt werden, müsste die Bewertung verschärft werden. In diesem Fall wären insbesondere A5, E4 sowie weiterhin S4 und I4 zu prüfen. Bei besonders sensiblen Beschäftigungskontexten oder systematischer Benachteiligungsgefahr könnte zudem eine weitergehende externe Prüfung erforderlich sein.

11.6 Fallbeispiel 6: KI-System in kritischer Infrastruktur und Cybersecurity

11.6.1 Ausgangssituation

Ein Betreiber kritischer Infrastruktur setzt ein KI-gestütztes System zur Überwachung und Absicherung seiner digitalen Betriebsumgebung ein. Das System analysiert fortlaufend technische Ereignisdaten, Netzwerkaktivitäten, Systemmeldungen, Zugriffsmuster und sicherheitsrelevante Auffälligkeiten. Es soll ungewöhnliche Vorgänge erkennen, Risiken priorisieren, interne Warnungen erzeugen und bei bestimmten Ereignissen vordefinierte Schutzmaßnahmen vorbereiten oder auslösen.

Das System arbeitet dauerhaft und ereignisgesteuert. Es wartet nicht nur auf einzelne Nutzeranfragen, sondern überwacht kontinuierlich technische Zustände und reagiert auf neue Ereignisse. Innerhalb definierter Grenzen kann es Sicherheitsmeldungen priorisieren, Vorfälle klassifizieren, betroffene Systeme markieren, interne Eskalationen auslösen oder vorbereitete Schutzmaßnahmen anstoßen. Besonders weitreichende Maßnahmen, etwa vollständige Abschaltungen, Eingriffe in zentrale Betriebsprozesse oder externe Meldungen mit erheblicher Tragweite, erfordern eine menschliche Freigabe.

Der Einsatzbereich ist hochsensibel. Fehlerhafte Klassifikationen können dazu führen, dass reale Angriffe zu spät erkannt, harmlose Vorgänge fälschlich als kritisch eingestuft, wichtige Betriebsprozesse unterbrochen oder Ressourcen falsch priorisiert werden. Da das System in eine sicherheitskritische Umgebung eingebunden ist, können Fehlentscheidungen nicht nur interne Abläufe betreffen, sondern Auswirkungen auf Versorgungssicherheit, technische Stabilität, Datenschutz, öffentliche Sicherheit und Vertrauen in den Betreiber haben.

11.6.2 Einstufung der vier Bewertungsdimensionen

Dimension	Einstufung	Begründung
Autonomie	**A6**	Das System ist dauerhaft und ereignisgesteuert aktiv. Es überwacht kontinuierlich technische Zustände, priorisiert Vorfälle und kann innerhalb definierter Grenzen Schutzmaßnahmen vorbereiten oder auslösen. Damit liegt persistente autonome Agentik vor.
Severität	**S5**	Der Einsatz betrifft kritische Infrastruktur, Cybersecurity, Versorgungssicherheit und sicherheitsrelevante technische Prozesse. Fehler können systemkritische Folgen haben.
Exposition	**E5**	Die Wirkung kann über einzelne Organisationseinheiten hinausgehen und technische Infrastrukturen, angeschlossene Systeme, externe Nutzer, Lieferketten oder öffentliche Versorgungsräume betreffen.
Irreversibilität	**I5**	Fehlhandlungen oder übersehene Angriffe können schwer oder kaum rückholbare Folgen haben, etwa dauerhafte Datenabflüsse, Betriebsunterbrechungen, infrastrukturelle Schäden oder sicherheitskritische Folgewirkungen.

Tabelle 25: ASEI-10-Einstufung für Fallbeispiel 6

11.6.3 Berechnung des fallbezogenen ASEI-10-Scores

Aus der qualitativen Einstufung ergibt sich zunächst: A = 6; S = 5; E = 5; I = 5.

Die Autonomiestufe A6 wird für die Scorebildung in den normierten Autonomiewert A_n überführt: $\mathbf{A_n = 3{,}5}$

Der Rohwert lautet $\boldsymbol{R = 3{,}5 \cdot 5 \cdot 5 \cdot 5 = 437{,}5}$ und wird anschließend logarithmisch normiert:

$$\mathbf{ASEI\text{-}10 = 1 + 9 \cdot \frac{\ln(437{,}5)}{\ln(625)} \approx 9{,}5}$$

Nach logarithmischer Normierung ergibt sich ein ASEI-10-Score von rund 9,5. Der Score liegt damit in der Risikoklasse RK6.

11.6.4 Prüfung der Schwellenkriterien

In diesem Fallbeispiel werden die höchsten Schwellenbereiche mehrerer Dimensionen erreicht. Die Autonomie liegt mit A6 im Bereich persistenter autonomer Agentik. Das System ist nicht nur auf einzelne Nutzeranfragen angewiesen, sondern bleibt dauerhaft aktiv, beobachtet technische Zustände und reagiert ereignisgesteuert. Daraus ergeben sich besondere Anforderungen an Monitoring, Zustandskontrolle, Protokollierung, Eingriffsmöglichkeiten und Abschaltmechanismen.

Die Severität erreicht S5, weil der Einsatzbereich kritische Infrastruktur und Cybersecurity betrifft. Fehlerhafte Entscheidungen können nicht nur betriebliche Nachteile verursachen, sondern sicherheitskritische oder infrastrukturelle Folgen auslösen. Deshalb reicht eine normale fachliche Prüfung nicht aus. Erforderlich sind eine systemkritische Folgenabschätzung, strenge Zugriffskontrollen, klare Verantwortlichkeiten und eine ausdrückliche Vertretbarkeitsprüfung.

Die Exposition erreicht E5, weil mögliche Wirkungen nicht auf eine einzelne Person, Gruppe oder Organisationseinheit begrenzt bleiben. Störungen, Fehlklassifikationen oder unterlassene Reaktionen können über technische Systeme, angeschlossene Prozesse, externe Nutzergruppen oder infrastrukturelle Zusammenhänge weiterwirken. Dadurch entsteht eine systemische Exposition.

Die Irreversibilität erreicht I5, weil bestimmte Folgen nicht vollständig rückholbar sind. Ein übersehener Angriff, ein nicht erkannter Datenabfluss, eine falsche Priorisierung sicherheitskritischer Meldungen oder eine ungeeignete Schutzmaßnahme kann Schäden verursachen, die später nur teilweise kompensiert werden können. Selbst wenn technische Systeme wiederhergestellt werden, können Datenabflüsse, Vertrauensverluste, Folgeschäden und Sicherheitslücken bestehen bleiben.

Als Mindestanforderungen ergeben sich insbesondere:

- strenge Rollen-, Rechte- und Zugriffskontrolle
- dauerhafte Protokollierung sicherheitsrelevanter Systemhandlungen
- menschliche Freigabe für weitreichende Eingriffe in Betriebsprozesse
- klare Eskalations- und Verantwortlichkeitsstruktur
- unabhängige Prüfung der Systemarchitektur
- regelmäßige Simulation und Überprüfung von Fehlalarmen und Nichterkennungen
- Notfall- und Rückfallkonzept für Fehlfunktionen
- Abschalt- oder Isolationsmöglichkeit des KI-Systems
- kontinuierliches Monitoring der Modell- und Systemleistung
- regelmäßige Neubewertung nach Systemänderungen, neuen Bedrohungslagen oder erweiterten Berechtigungen

11.6.5 Ergebnisinterpretation

Das Fallbeispiel erreicht mit einem ASEI-10-Score von rund 9,5 ein sehr hohes Risikopotenzial. Dieses Ergebnis ist plausibel, weil mehrere höchste Risikodimensionen zusammenwirken: Der Einsatzbereich besitzt systemkritische Severität, die Exposition reicht über einzelne interne Prozesse hinaus, mögliche Folgen können kaum rückholbar sein und das System arbeitet dauerhaft ereignisgesteuert.

Der Score wird nicht allein durch Autonomie getrieben. Auch wenn das System noch nicht die Stufen A7 bis A9 erreicht, führt die Kombination aus A6, S5, E5 und I5 zu einer sehr hohen Gesamteinstufung. Das Beispiel zeigt damit eine wichtige Stärke des ASEI-10-Modells: Ein KI-System muss nicht unkontrollierbar, verteilt oder selbstoptimierend sein, um ein sehr hohes Risikopotenzial zu besitzen. In kritischer Infrastruktur kann bereits eine persistente, begrenzte Agentik ausreichen, wenn Severität, Exposition und Irreversibilität maximal ausgeprägt sind.

Gleichzeitig macht das Beispiel die Bedeutung der Schwellenkriterien sichtbar. Ein solches System darf nicht allein nach dem rechnerischen Score beurteilt werden. S5, E5 und I5 lösen jeweils eigenständige Prüf- und Mindestanforderungen aus. Entscheidend sind daher nicht nur technische Leistungsfähigkeit und Erkennungsqualität, sondern auch Begrenzbarkeit, Protokollierung, menschliche Eingriffsmöglichkeit, Notfallfähigkeit und unabhängige Überprüfbarkeit.

Würde das System zusätzlich mehrere Agenten koordinieren, eigene Abwehrstrategien optimieren, seine Schutzmaßnahmen selbstständig verändern oder sich über mehrere Infrastrukturbereiche hinweg verteilen, wären A7 oder A8 zu prüfen. Sollte es nicht mehr zuverlässig begrenzbar, überwachbar, abschaltbar oder rückholbar sein, wäre A9 zu prüfen. In diesem Fall wäre eine grundlegende Vertretbarkeitsprüfung erforderlich, bevor ein produktiver Einsatz verantwortbar wäre.

Das Beispiel markiert damit den oberen Bereich der ASEI-10-Fallbeispiele. Es verdeutlicht, dass sehr hohes Risikopotenzial nicht erst bei Kontrollverlustszenarien entsteht, sondern bereits dann, wenn persistente agentische Aktivität mit systemkritischem Einsatzbereich, systemischer Exposition und kaum reversiblen Folgen zusammentrifft.

11.7 Vergleichende Zusammenfassung der Fallbeispiele und Überleitung zu den Grenzen des Modells

Die sechs Fallbeispiele zeigen die praktische Spannweite des ASEI-10-Modells. Sie reichen vom privaten, reaktiven Chatbot mit sehr niedrigem Risikopotenzial bis zu einem persistenten KI-System in kritischer Infrastruktur und Cybersecurity mit sehr hohem Risikopotenzial. Dadurch wird deutlich, dass ASEI-10 nicht auf eine abstrakte Bewertung von KI-Technologie zielt, sondern auf die konkrete Einordnung von KI-Systemen in ihrem jeweiligen Anwendungskontext.

Derselbe technische Grundtyp kann je nach Systemeinbettung, Datenzugriff, Nutzerkreis, Entscheidungsnähe, Reichweite und Rückholbarkeit möglicher Folgen zu sehr unterschiedlichen Risikowerten führen. Entscheidend ist daher nicht allein die Leistungsfähigkeit eines Modells, sondern die konkrete Verbindung von Autonomie, Severität, Exposition und Irreversibilität.

Die ausgefüllten ASEI-10-Bewertungsbögen zu den einzelnen Fallbeispielen sind in Anhang C zusammengeführt. Sie dokumentieren jeweils, wie Bewertungsgegenstand, Systemkontext, Dimensionseinstufung, Prüfung korrelativer Beziehungen, Scorebildung, Schwellenprüfung und Mindestanforderungen zusammengeführt werden.

Die folgende Tabelle verdichtet diese Einzelbewertungen zu einer vergleichenden Übersicht.

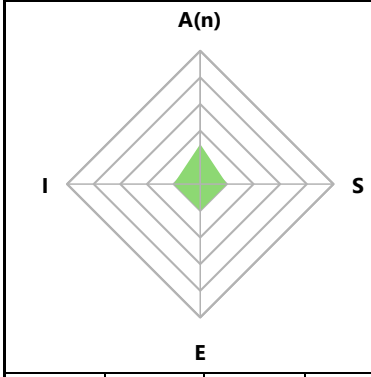
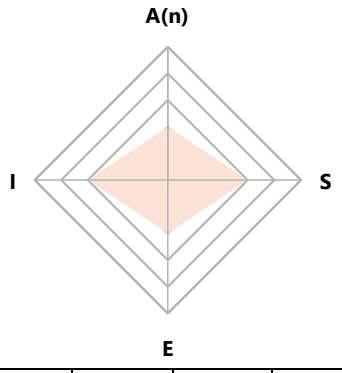
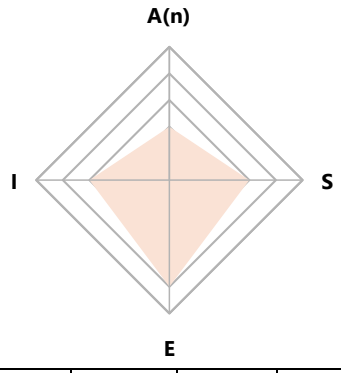
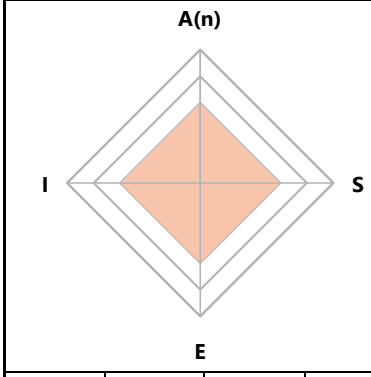
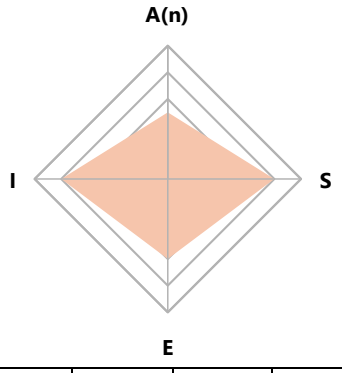
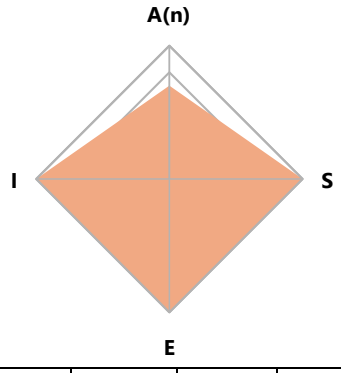
Nr.	Fallbeispiel	Qualitative Einstufung	Rechenwerte	Rohwert *R*	ASEI-10-Score	Risikoklasse
1	Privater Chatbot für kreative Texte	A2 / S1 / E1 / I1	1,5 / 1 / 1 / 1	1,5	1,6	RK1
2	KI-Assistent für Unterrichts- und Prüfungsvorbereitung	A3 / S3 / E2 / I3	2,0 / 3 / 2 / 3	36,0	6,0	RK4
3	KI-System im Kundenservice	A3 / S3 / E4 / I3	2,0 / 3 / 4 / 3	72,0	6,98 (ungerundet) 7,0 (gerundet)	RK4, grenznah RK5
4	KI-Agent mit Toolzugriff in einem Unternehmen	A5 / S3 / E3 / I3	3,0 / 3 / 3 / 3	81,0	7,1	RK5
5	KI-System zur Personalvorauswahl	A4 / S4 / E3 / I4	2,5 / 4 / 3 / 4	120,0	7,7	RK5
6	KI-System in kritischer Infrastruktur und Cybersecurity	A6 / S5 / E5 / I5	3,5 / 5 / 5 / 5	437,5	9,5	RK6

Tabelle 26: Vergleichende Übersicht der sechs Fallbeispiele nach normierter Autonomie

Zur Ergänzung der tabellarischen Übersicht werden die sechs Fallbeispiele zusätzlich als Netzdiagramme dargestellt. Die Diagramme visualisieren nicht den ASEI-10-Score selbst, sondern die zugrunde liegenden Dimensionsprofile. Dargestellt werden der normierte Autonomiewert A(n), die Severität S, die Exposition E und die Irreversibilität I auf einer einheitlichen Skala von 0 bis 5. A(n) bezeichnet dabei den normierten Autonomiewert A_n, der für die Scorebildung aus der qualitativen Autonomiestufe A1 bis A9 abgeleitet wird.

Die Netzdiagramme machen sichtbar, aus welchen Risikotreibern sich die jeweilige Einstufung zusammensetzt. Dadurch wird erkennbar, ob ein Fallbeispiel eher durch Autonomie, durch den Einsatzbereich, durch breite Exposition, durch schwer reversible Folgen oder durch ein gleichmäßiges Zusammenwirken mehrerer Dimensionen geprägt ist. Sie ergänzen damit die Score- und Tabellenbetrachtung um eine profilbezogene Darstellung des Risikopotenzials.

Abbildung 2: Dimensionsprofile der sechs Fallbeispiele auf Basis von A(n), S, E und I
Anmerkung: A(n) bezeichnet den normierten Autonomiewert A_n.

(1) Privater Chatbot für kreative Texte	(2) KI-Assistent für Unterrichts- und Prüfungsvorbereitung	(3) KI-System im Kundenservice
		
A_n = 1,5 | S = 1 | E = 1 | I = 1	A_n = 2,0 | S = 3 | E = 2 | I = 3	A_n = 2,0 | S = 3 | E = 4 | I = 3
R = 1,5 | ASEI-10-Score = 1,6 | Risikoklasse: RK1	*R* = 36 | ASEI-10-Score = 6 | Risikoklasse: RK4	*R* = 72 | ASEI-10-Score = 6,98 | Risikoklasse: RK4
(4) KI-Agent mit Toolzugriff in einem Unternehmen	**(5) KI-System zur Personalvorauswahl**	**(6) KI-System in kritischer Infrastruktur und Cybersecurity**
		
A_n = 3,0 | S = 3 | E = 3 | I = 3	A_n = 2,5 | S = 4 | E = 3 | I = 4	A_n = 3,5 | S = 5 | E = 5 | I = 5
R = 81 | ASEI-10-Score = 7,1 | Risikoklasse: RK5	*R* = 120 | ASEI-10-Score = 7,7 | Risikoklasse: RK5	*R* = 437,5 | ASEI-10-Score = 9,5 | Risikoklasse: RK6

Die tabellarische Übersicht und die Netzdiagramme machen deutlich, dass das Risikopotenzial nicht linear mit der Autonomie steigt. Zwar nimmt die Autonomie von Fallbeispiel 1 bis Fallbeispiel 6 tendenziell zu, doch die Höhe des ASEI-10-Scores wird stets durch das Zusammenspiel aller vier Dimensionen bestimmt. Ein System mit begrenzter Autonomie kann ein erhebliches Risikopotenzial besitzen, wenn es in einem sensiblen Einsatzbereich wirkt, viele Personen betrifft oder schwer reversible Folgen erzeugen kann.

Fallbeispiel 1 zeigt den unteren Rand der Skala. Der private Chatbot bleibt reaktiv, wirkt nur im privaten Einzelkontext und erzeugt leicht korrigierbare Folgen. Das Risikopotenzial ist deshalb sehr niedrig.

Fallbeispiel 2 verdeutlicht, dass bereits ein assistives System im Bildungs- und Prüfungskontext ein erhöhtes Risikopotenzial erreichen kann, weil fachliche Fehler Lernprozesse und Prüfungsvorbereitung beeinflussen können. Die Einstufung wird nicht durch hohe Autonomie getragen, sondern durch die Folgenrelevanz prüfungsnaher Materialien, die Wirkung auf eine Kursgruppe und die begrenzte Rückholbarkeit fehlerhafter Vorbereitung.

Fallbeispiel 3 zeigt einen wichtigen Grenzfall. Das Kundenservice-System besitzt nur eine begrenzte Autonomie, erreicht aber durch breite Kundenexposition, personenbezogene oder wirtschaftlich relevante Informationen und begrenzt reversible Folgen einen Score unmittelbar an der Schwelle zur hohen Risikoklasse. Die gerundete Darstellung von 7,0 darf daher nicht mechanisch interpretiert werden; maßgeblich bleiben der ungerundete Wert und die fachliche Begründung.

Fallbeispiel 4 zeigt den Übergang zu agentischen Unternehmenssystemen. Hier entsteht das hohe Risikopotenzial vor allem aus der Verbindung von Toolzugriff, begrenzter autonomer Handlungsausführung und organisationaler Einbettung. Das System muss nicht in einem hochkritischen Bereich eingesetzt werden, um durch operative Prozesswirkung prüf- und kontrollbedürftig zu werden.

Fallbeispiel 5 verdeutlicht, dass hohe Risikowerte nicht zwingend aus hoher technischer Autonomie folgen. Das System zur Personalvorauswahl bleibt formal in eine menschliche Letztentscheidung eingebettet, berührt aber einen hochsensiblen Beschäftigungskontext und kann schwer reversible Benachteiligungen auslösen. Die Severität und Irreversibilität tragen hier das Risikopotenzial besonders stark.

Fallbeispiel 6 markiert den oberen Bereich der betrachteten Beispiele. Das System in kritischer Infrastruktur und Cybersecurity verbindet persistente agentische Aktivität mit systemkritischem Einsatzbereich, systemischer Exposition und kaum reversiblen Folgen. Dadurch entsteht ein sehr hohes Risikopotenzial, auch ohne dass das System bereits als nicht zuverlässig kontrollierbarer KI-Akteur eingestuft werden müsste.

Die vergleichende Betrachtung zeigt damit, dass unterschiedliche Risikopotenziale auf unterschiedlichen Wegen entstehen können. Ein System kann vor allem durch hohe Autonomie risikorelevant werden, ein anderes durch hohe Severität, breite Exposition oder schwer reversible Folgen. Für die praktische Steuerung ist daher nicht nur der Score selbst entscheidend, sondern auch die Frage, welche Dimensionen den Score tragen.

Die Fallbeispiele verdeutlichen außerdem, dass der ASEI-10-Score nicht isoliert interpretiert werden darf. Der Score verdichtet die Einzeldimensionen zu einem gemeinsamen Klassifikationswert, erklärt aber nicht allein, wodurch das Risikopotenzial entsteht. Maßgeblich bleibt die Verbindung von Score, Einzeldimensionen, Prüfschwellen, Bewertungsbogen und fachlicher Begründung. Erst diese Verbindung zeigt, ob ein System vor allem wegen seiner Handlungstiefe, seines Einsatzbereichs, seiner Reichweite oder seiner mangelnden Rückholbarkeit besondere Aufmerksamkeit erfordert.

Gerade die Grenzfälle zeigen den praktischen Nutzen des Modells. Ein Score nahe an einer Bereichsgrenze darf nicht mechanisch behandelt werden. Ebenso dürfen ausgelöste Prüfschwellen nicht durch einen moderaten Gesamtscore relativiert werden. ASEI-10 soll daher keine schematische Ampellogik erzeugen, sondern eine nachvollziehbare Risikopotenzialbewertung ermöglichen, die fachliche Begründung, Dokumentation und organisatorische Verantwortung miteinander verbindet.

Die Fallbeispiele zeigen damit sowohl die Leistungsfähigkeit als auch die Grenzen des Modells. ASEI-10 hilft, konkrete KI-Systeme strukturiert zu bewerten, Risikotreiber sichtbar zu machen, Prioritäten zu setzen und Mindestanforderungen aus Prüfschwellen abzuleiten. Sein Nutzen liegt gerade darin, abstrakte Diskussionen über KI-Risiken in eine vergleichbare Bewertung einzelner Anwendungskontexte zu überführen.

Gleichzeitig darf dieser Nutzen nicht überdehnt werden. ASEI-10 ist kein rechtlicher Compliance-Nachweis, keine Zertifizierung, keine technische Sicherheitsgarantie und kein vollständiges Managementsystem. Die Aussagekraft des Modells hängt von der Qualität der Systembeschreibung, der fachlichen Begründung der Einstufungen, der Prüfung von Schwellenkriterien, der Dokumentation und der regelmäßigen Neubewertung ab. Deshalb muss im folgenden Kapitel ausdrücklich geklärt werden, wo die Grenzen des ASEI-10-Modells liegen und welche Fehlanwendungen vermieden werden müssen.

12 Grenzen und Fehlanwendungen des ASEI-10-Modells

12.1 Funktion und Reichweite des Grenzenkapitels

Ein Klassifikationsmodell für KI-Risiken muss nicht nur seinen Anwendungsbereich, sondern auch seine Grenzen offenlegen. Das gilt besonders für ein Modell wie ASEI-10, das unterschiedliche Risikodimensionen in einem gemeinsamen Score zusammenführt. Die rechnerische Verdichtung schafft Übersichtlichkeit und Vergleichbarkeit, darf aber nicht den Eindruck erzeugen, ein komplexes Risikoprofil lasse sich vollständig durch eine einzige Zahl erfassen.

Das ASEI-10-Modell ist daher als Bewertungs- und Strukturierungsinstrument zu verstehen. Es hilft, Risikopotenziale konkreter KI-Systeme systematisch zu erfassen, zu vergleichen und zu dokumentieren. Es ersetzt jedoch weder rechtliche Prüfung noch technische Sicherheitsanalyse, fachliche Qualitätssicherung, ethische Abwägung, organisatorische Verantwortung oder kontinuierliches Monitoring.

Die Aussagekraft des Modells hängt wesentlich von der Qualität der zugrunde liegenden Systembeschreibung ab. Werden Funktionen, Datenzugriffe, Werkzeugrechte, Nutzergruppen, Einsatzgrenzen oder mögliche Fehlwirkungen unvollständig beschrieben, kann auch der resultierende ASEI-10-Score nur eingeschränkt belastbar sein. Das Modell liefert keine bessere Bewertung als die Informationen, auf denen es beruht.

12.2 Grundsätzliche Grenzen des Modells

12.2.1 Kein vollständiger Risikowert im klassischen Sinn

ASEI-10 bewertet nicht Risiko im klassischen Sinn einer Verbindung von Eintrittswahrscheinlichkeit und Schadensausmaß.[30] Das Modell enthält bewusst keine eigene Wahrscheinlichkeitsdimension. Es bestimmt also nicht, wie wahrscheinlich ein konkretes Schadensereignis tatsächlich eintritt. Stattdessen klassifiziert es das strukturelle Risikopotenzial eines KI-Systems im konkreten Anwendungskontext.

Diese Abgrenzung ist zentral. Zwei KI-Systeme mit identischer Einbettung, identischem Severität, identischer Exposition und identischer Irreversibilität können denselben ASEI-10-Score erhalten, obwohl eines technisch robuster, zuverlässiger oder weniger fehleranfällig ist als das andere. Solche Unterschiede müssen durch ergänzende technische Evaluation, Fehlerratenanalysen, Robustheitstests, Red Teaming, Monitoring oder Betriebserfahrungen erfasst werden.

Ein niedriger ASEI-10-Score bedeutet daher nicht, dass ein System zuverlässig oder fehlerfrei ist. Ein hoher ASEI-10-Score bedeutet nicht, dass ein Schaden wahrscheinlich eintreten wird. Der Score beschreibt vielmehr, wie gravierend das Risikopotenzial wäre, wenn Fehler, Fehlverwendung, Missbrauch oder Kontrollprobleme im beschriebenen Anwendungskontext auftreten.

[30] ISO 31000:2018; IEC 31010:2019

12.2.2 Kein Ersatz für rechtliche, technische oder fachliche Prüfung

ASEI-10 ist kein Rechtsgutachten und kein Compliance-Nachweis. Das Modell kann Hinweise darauf geben, welche KI-Systeme einer vertieften Prüfung bedürfen, entscheidet aber nicht verbindlich, ob ein System unter bestimmte gesetzliche Pflichten fällt. Insbesondere ersetzt ein niedriger ASEI-10-Score keine Prüfung nach dem EU AI Act, Datenschutzrecht, Arbeitsrecht, Produktsicherheitsrecht, Urheberrecht, Antidiskriminierungsrecht oder sonstigen einschlägigen Rechtsgebieten.

Ebenso ist ASEI-10 keine technische Sicherheitsprüfung. Das Modell bewertet das Risikopotenzial eines KI-Systems anhand von Autonomie, Severität, Exposition und Irreversibilität. Es prüft jedoch nicht unmittelbar, ob ein System robust, manipulationsresistent, adversarial getestet, cybersicher, erklärbar, bias-arm oder technisch zuverlässig ist. Solche Eigenschaften müssen durch geeignete technische Verfahren, Tests, Audits, Red Teaming, Monitoring und Qualitätssicherungsmaßnahmen gesondert untersucht werden.

Auch die fachliche Prüfung bleibt unverzichtbar. Ein KI-System kann rechnerisch niedrig oder moderat eingestuft werden und dennoch fachlich falsche, irreführende oder ungeeignete Ergebnisse liefern. Das Modell bewertet nicht automatisch die inhaltliche Richtigkeit einzelner Ausgaben. Fachliche Verantwortung bleibt daher bei den Personen oder Organisationen, die das System einsetzen, freigeben oder seine Ergebnisse weiterverwenden.

12.2.3 Bewertungsunsicherheiten und Ermessensspielräume

Die Einstufung eines KI-Systems in ASEI-10 enthält fachliche Ermessensspielräume. Nicht jeder Anwendungskontext lässt sich eindeutig einer Stufe zuordnen. Besonders Grenzfälle zwischen A3 und A4, S3 und S4, E3 und E4 oder I3 und I4 können eine sorgfältige Begründung erfordern.

Diese Ermessensspielräume sind kein methodischer Fehler, sondern Ausdruck der Komplexität realer KI-Systeme. Ein Bewertungssystem kann fachliches Urteil nicht vollständig ersetzen. Es kann jedoch die Entscheidung strukturieren, die maßgeblichen Kriterien offenlegen und die Begründung nachvollziehbar machen.

Für Grenzfälle gilt im ASEI-10-Modell ein Vorsichtsprinzip. Wenn zwei Einstufungen fachlich vertretbar erscheinen, sollte die höhere Stufe gewählt werden, sofern plausible Schadensszenarien, unklare Kontrollmöglichkeiten oder schwer abschätzbare Folgewirkungen bestehen. Die niedrigere Stufe ist nur dann angemessen, wenn ihre Voraussetzungen belastbar begründet und dokumentiert werden können.

12.3 Grenzen der Score- und Skaleninterpretation

12.3.1 Grenzen der Scorebildung

Der ASEI-10-Score verdichtet vier Bewertungsdimensionen zu einem gemeinsamen Indexwert. Diese Verdichtung ist für Vergleichbarkeit und Priorisierung hilfreich, erzeugt aber zwangsläufig Informationsverlust. Zwei KI-Systeme können denselben Score erreichen, obwohl ihre Risikostruktur unterschiedlich ist. Ein System kann vor allem durch hohe Autonomie auffällig sein, ein anderes durch hohen Severität, hohe Exposition oder hohe Irreversibilität.

Deshalb darf der ASEI-10-Score nie isoliert interpretiert werden. Er muss immer gemeinsam mit der qualitativen Autonomiestufe A, dem normierten Autonomiewert A_n, den übrigen Einzeldimensionen S, E und I sowie den ausgelösten Prüfschwellen betrachtet werden. Ein Score von 7,0 kann je nach Konstellation sehr unterschiedliche Maßnahmen erfordern. Ein Kundenservice-System mit hoher Exposition benötigt andere Kontrollen als ein Recruiting-System mit hohem Severität und schwer reversiblen Folgen.

Auch die mathematische Form der Scorebildung ist eine Modellentscheidung. Die Multiplikation der vier Rechenwerte bildet eine heuristische Verknüpfung der Risikodimensionen. Die logarithmische Normierung macht die Skala praktikabel. Dennoch bleibt jede Scoreformel eine bewusste Vereinfachung. Sie ist keine naturgesetzliche Beschreibung von Risiko, sondern eine methodische Konstruktion zur strukturierten Klassifikation von Risikopotenzialen.

12.3.2 Ordinalskalen und Indexcharakter

Die Skalen des ASEI-10-Modells sind geordnete Bewertungsskalen; die methodischen Grenzen solcher ordinalen Bewertungslogiken sind in der Risikomatrix-Literatur ausführlich diskutiert worden.[31] Eine höhere Stufe bedeutet eine stärkere Ausprägung der jeweiligen Dimension. Daraus folgt jedoch nicht, dass die Abstände zwischen den Stufen exakt gleich groß sind oder dass eine Stufe im mathematischen Sinn doppelt so stark ist wie eine andere.

Der ASEI-10-Score ist deshalb als heuristischer Klassifikationsindex zu verstehen, nicht als exakter Messwert. Die Zahlenwerte dienen der Vergleichbarkeit, Priorisierung und Dokumentation. Sie dürfen nicht so interpretiert werden, als handele es sich um objektiv gemessene Risikoeinheiten.

Diese Einschränkung betrifft insbesondere die Interpretation knapper Unterschiede. Ein System mit einem Score von 7,3 ist nicht automatisch in einem messgenauen Sinn riskanter als ein System mit 7,1. Entscheidend sind der Risikobereich, die ausgelösten Prüfschwellen, die Einzeldimensionen und die fachliche Begründung der Einstufung.

12.3.3 Gefahr rechnerischer Scheingenauigkeit

Ein numerischer Score kann den Eindruck hoher Präzision erzeugen. Diese Präzision darf nicht überschätzt werden. Wenn ein System mit 7,3 und ein anderes mit 7,5 bewertet wird, bedeutet dies nicht zwangsläufig, dass das zweite System objektiv und messgenau gefährlicher ist. Die Scorewerte dienen der Orientierung, nicht der exakten Messung im naturwissenschaftlichen Sinne.

Besonders bei knappen Unterschieden sollten daher nicht einzelne Dezimalstellen über Maßnahmen entscheiden. Wichtiger sind Risikobereich, Einzeldimensionen, Schwellenkriterien und fachliche Begründung. Ein System mit 6,9 kann praktisch ähnlich streng zu behandeln sein wie ein System mit 7,0, wenn relevante Schwellenwerte ausgelöst werden.

Die Rundung auf eine Dezimalstelle ist deshalb bewusst pragmatisch. Sie soll Vergleichbarkeit ermöglichen, ohne eine übertriebene Genauigkeit vorzutäuschen. Für Entscheidungen mit hoher Tragweite sollte der Score immer durch qualitative Analyse, Szenariobetrachtung und zusätzliche Prüfverfahren ergänzt werden.

[31] Cox 2008

12.3.4 Kommunikationsregel bei ausgelösten Höchstschwellen

Der ASEI-10-Score darf nicht isoliert berichtet werden, wenn eine der höchsten qualitativen Schwellen ausgelöst wird. Dies gilt insbesondere für A9, S5, E5 oder I5. In solchen Fällen ist neben dem Score stets der Schwellenvermerk anzugeben, weil der zusammenfassende Score allein die qualitative Bedeutung einer Höchstschwelle verdecken könnte.

Ein bewusst zugespitzter Eckfall verdeutlicht diese Grenze: Ein nicht zuverlässig kontrollierbarer KI-Akteur der Stufe A9 in einem ansonsten folgenarmen, individuell begrenzten und leicht reversiblen Kontext würde rechnerisch mit $A_n = 5$, $S = 1$, $E = 1$ und $I = 1$ bewertet. Daraus ergäbe sich $R = 5$ und ein ASEI-10-Score von rund 3,3. Eine isolierte Kommunikation als „niedriges Risikopotenzial“ wäre jedoch irreführend. Die Autonomiestufe A9 löst unabhängig vom Score eine Vertretbarkeitsprüfung aus und muss ausdrücklich berichtet werden.

Zulässig wäre daher nicht die Aussage:

> *„Das System erreicht nur einen ASEI-10-Score von 3,3.“*

Sachgerecht wäre vielmehr:

> *„Das System erreicht rechnerisch einen ASEI-10-Score von 3,3; zugleich liegt mit A9 eine Höchstschwelle der Autonomiedimension vor, die eine gesonderte Vertretbarkeitsprüfung auslöst.“*

Diese Kommunikationsregel gilt allgemein: Der ASEI-10-Score ist immer gemeinsam mit den Einzelwerten A, A_n, S, E, I und den ausgelösten Prüfschwellen zu dokumentieren. Der Score verdichtet die Bewertung, ersetzt aber nicht die qualitative Schwellenprüfung.

12.3.5 Vorläufigkeit der Interpretationsbereiche

Die Interpretationsbereiche des ASEI-10-Scores sind als vorläufige Orientierung zu verstehen. Sie wurden nicht empirisch kalibriert und stellen keine validierten Risikogrenzen dar. Ihre Funktion besteht darin, Organisationen eine praktikable Einordnung zu ermöglichen und erste Priorisierungen zu unterstützen.

Daraus folgt, dass die Bereichsgrenzen nicht mechanisch angewendet werden sollten. Ein System knapp unterhalb einer Grenze kann ähnliche Anforderungen auslösen wie ein System knapp oberhalb einer Grenze. Umgekehrt kann ein System innerhalb desselben Scorebereichs aufgrund anderer Risikotreiber andere Maßnahmen erfordern.

Eine zukünftige Weiterentwicklung des Modells sollte die Interpretationsbereiche anhand realer Fallbewertungen, Expertenurteilen, Sensitivitätsanalysen und Inter-Rater-Reliabilitätsprüfungen überprüfen. Bis dahin sind die Scorebereiche als begründete, aber nicht validierte Klassifikationshilfe zu verstehen.

12.4 Grenzen der Dimensionsabgrenzung und Systembeschreibung

12.4.1 Überschneidungen zwischen den Dimensionen

Die vier Dimensionen des ASEI-10-Modells sind analytisch getrennt, aber in der Praxis nicht immer vollständig unabhängig. Hohe Severität, hohe Exposition und hohe Irreversibilität können gemeinsam auftreten. Besonders deutlich wird dies bei kritischer Infrastruktur, sicherheitskritischen Anwendungen oder großflächig eingesetzten Systemen.

Solche Überschneidungen müssen bei der Anwendung reflektiert werden. Ein System darf nicht automatisch in mehreren Dimensionen hoch eingestuft werden, nur weil eine Dimension hoch ist. Vielmehr ist jede Dimension eigenständig zu begründen. Severität fragt nach der Schwere möglicher negativer Folgen. Exposition fragt nach der Breite des Wirkungsraums. Irreversibilität fragt nach der Rückholbarkeit möglicher Folgen.

Wenn alle drei Dimensionen hoch ausgeprägt sind, kann dies sachlich gerechtfertigt sein. Es muss dann aber erkennbar sein, dass unterschiedliche Risikomerkmale bewertet wurden und nicht derselbe Sachverhalt mehrfach ungeprüft gezählt wurde. Die Qualität der Bewertung hängt daher wesentlich von der diskriminierenden Begründung der Einzeldimensionen ab.

12.4.2 Dynamik von KI-Systemen und Neubewertung

KI-Systeme sind nicht statisch. Modelle werden aktualisiert, Werkzeuge ergänzt, Berechtigungen erweitert, Datenbestände verändert, Schnittstellen angebunden und Einsatzbereiche ausgeweitet. Dadurch kann sich das Risikopotenzial erheblich verändern, ohne dass der Systemname oder die Benutzeroberfläche sichtbar anders erscheinen.

Eine ASEI-10-Bewertung ist daher immer zeit- und kontextbezogen. Sie gilt nur für den dokumentierten Systemzustand und den beschriebenen Anwendungskontext. Wird ein KI-System erweitert, produktiv geschaltet, in neue Prozesse integriert, für größere Nutzergruppen freigegeben oder mit zusätzlichen Handlungsmöglichkeiten ausgestattet, ist eine Neubewertung erforderlich.

Besonders kritisch sind schleichende Risikoerhöhungen. Ein System beginnt als assistiver Chatbot, erhält später Dateizugriff, danach Zugriff auf E-Mail, Kalender oder Datenbanken und schließlich die Fähigkeit, Prozesse eigenständig auszulösen. Jede einzelne Erweiterung kann harmlos erscheinen, in der Summe aber zu einer deutlich höheren Autonomie, Exposition oder Irreversibilität führen.

12.4.3 Grenzen bei unbekannten oder verdeckten Systemfähigkeiten

ASEI-10 setzt voraus, dass die relevanten Systemeigenschaften bekannt oder zumindest plausibel rekonstruierbar sind. In der Praxis ist dies nicht immer der Fall. Bei proprietären Modellen, geschlossenen Plattformen oder komplexen integrierten KI-Systemen können Trainingsdaten, Sicherheitsmechanismen, Toolentscheidungen, Modellgrenzen, Updatezyklen oder interne Bewertungslogiken nur eingeschränkt transparent sein.

Solche Unsicherheiten müssen in der Bewertung ausdrücklich dokumentiert werden. Unbekannte Fähigkeiten oder unklare Kontrollmechanismen dürfen nicht automatisch zugunsten einer niedrigeren Einstufung ausgelegt werden. Wenn zentrale Systemmerkmale nicht belastbar bekannt sind, sollte die Bewertung konservativ erfolgen oder mit einem Unsicherheitsvermerk versehen werden.

Dies gilt besonders bei Systemen mit hoher Autonomie, sensiblen Datenzugriffen, öffentlicher Exposition oder schwer reversiblen Folgen. In solchen Fällen reicht es nicht aus, sich allein auf Herstellerangaben, Marketingbeschreibungen oder allgemeine Nutzungsversprechen zu verlassen. Erforderlich sind technische Nachweise, Prüfprotokolle, kontrollierte Tests oder vertraglich abgesicherte Informationen.

12.5 Fehlanwendungen und Umgang mit Unsicherheit

12.5.1 Typische Fehlanwendungen des ASEI-10-Modells

Typische Fehlanwendungen des ASEI-10-Modells entstehen meist nicht durch die Scorebildung selbst, sondern durch eine verkürzte oder missverständliche Nutzung des Modells. Besonders problematisch ist es, den Score isoliert zu betrachten, den Anwendungskontext zu vernachlässigen oder ASEI-10 mit einer vollständigen Risiko-, Rechts- oder Sicherheitsprüfung gleichzusetzen.

Die folgende Tabelle fasst zentrale Fehlanwendungen zusammen, benennt das jeweilige Problem und zeigt geeignete Gegenmaßnahmen. Sie dient zugleich als praktische Kontrollliste für die Anwendung des Modells in Organisationen.

Fehlanwendung	Problem	Gegenmaßnahme
Bewertung eines Modells ohne Anwendungskontext	Das Risiko wird vom konkreten Einsatz getrennt und dadurch verzerrt	Immer das konkrete KI-System im konkreten Nutzungskontext bewerten
Verwechslung mit klassischer Risikoanalyse	Eintrittswahrscheinlichkeit wird fälschlich als im Score enthalten verstanden	ASEI-10 als Risikopotenzial-Index verwenden und Wahrscheinlichkeiten gesondert prüfen
Isolierte Interpretation des Scores	Einzeldimensionen und Schwellenkriterien werden übersehen	Score stets gemeinsam mit A, A_n, S, E, I und Prüfschwellen ausweisen
Einstufung nach beabsichtigter Idealnutzung	Realistische Fehlverwendung und Eskalation bleiben unberücksichtigt	Normalbetrieb, Fehlverwendung und plausible Schadensszenarien prüfen
Zu niedrige Einstufung bei Unsicherheit	Unklare Systemfähigkeiten werden verharmlost	Bei relevanter Unsicherheit konservativ einstufen oder Unsicherheitsvermerk dokumentieren
Verwechslung mit Rechtskonformität	Ein Score wird fälschlich als Compliance-Nachweis verstanden	Rechtliche Prüfung gesondert durchführen
Verwechslung mit technischer Sicherheit	Der Score wird als Nachweis von Robustheit oder Zuverlässigkeit verwendet	Technische Tests, Audits, Red Teaming und Monitoring ergänzen
Keine Neubewertung nach Systemänderungen	Erweiterte Fähigkeiten verändern das Risikoprofil unbemerkt	Neubewertung bei neuen Funktionen, Datenzugriffen, Nutzergruppen oder Einsatzbereichen
Überbewertung von Dezimalstellen	Scheinpräzision führt zu unangemessen feinen Unterscheidungen	Risikobereiche, Schwellen und qualitative Begründung priorisieren

Tabelle 27: Typische Fehlanwendungen und Gegenmaßnahmen

12.5.2 Umgang mit Unsicherheit

Unsicherheit ist bei der Bewertung von KI-Systemen nicht vollständig vermeidbar. Sie kann aus unvollständigen Informationen, technischen Intransparenzen, dynamischen Modelländerungen, unbekannten Fehlverwendungen, neuen Angriffsmethoden oder unklaren organisatorischen Verantwortlichkeiten entstehen. Ein seriöses Bewertungsmodell muss diese Unsicherheit nicht ausblenden, sondern sichtbar machen.

Im ASEI-10-Modell sollte Unsicherheit auf drei Ebenen behandelt werden. Erstens ist zu dokumentieren, welche Annahmen der Bewertung zugrunde liegen. Zweitens ist festzuhalten, welche Informationen fehlen oder unsicher sind. Drittens ist zu prüfen, ob diese Unsicherheit eine höhere Einstufung oder zusätzliche Mindestanforderungen rechtfertigt.

Besonders in hochkritischen, stark exponierten oder schwer reversiblen Kontexten sollte Unsicherheit nicht als Entlastung verstanden werden. Wenn unklar ist, ob ein System bestimmte Handlungen ausführen, Daten weitergeben, externe Systeme beeinflussen oder Schäden rückholen kann, spricht dies eher für erhöhte Vorsicht als für eine niedrigere Bewertung.

12.6 Zusammenfassung und Überleitung zum Fazit

Die Grenzen des ASEI-10-Modells liegen nicht in seiner Unbrauchbarkeit, sondern in seinem bewusst begrenzten Anspruch. ASEI-10 ist ein Klassifikationsmodell zur strukturierten Bewertung des Risikopotenzials konkreter KI-Systeme. Es schafft Transparenz, Vergleichbarkeit und Priorisierung, ersetzt aber keine rechtliche, technische, fachliche oder organisatorische Verantwortung.

Die wichtigste Schutzmaßnahme gegen Fehlanwendung besteht darin, Score, Einzeldimensionen, Prüfschwellen und Begründung immer gemeinsam zu betrachten. Der Score verdichtet, aber er erklärt nicht allein. Die Einzeldimensionen zeigen die Risikostruktur. Die Prüfschwellen definieren Mindestanforderungen. Die Begründung macht die Einstufung überprüfbar.

Damit ist der methodische Rahmen des ASEI-10-Modells vollständig beschrieben: Die Bewertungsdimensionen sind definiert, die methodische Einordnung ist erfolgt, die Scorebildung ist erläutert, die praktische Anwendung wurde anhand von Fallbeispielen gezeigt, die Einordnung in bestehende Rahmenwerke wurde vorgenommen und die Grenzen des Modells sind benannt. Das abschließende Kapitel fasst den Beitrag des Modells zusammen und gibt einen Ausblick auf mögliche Weiterentwicklung, Validierung und Anwendung.

13 Fazit und Ausblick

13.1 Zusammenfassung des Modellansatzes

Das ASEI-10-Modell wurde entwickelt, um das Risikopotenzial konkreter KI-Systeme in konkreten Anwendungskontexten strukturiert, nachvollziehbar und vergleichbar zu klassifizieren. Ausgangspunkt ist die Annahme, dass KI-Risiken nicht allein aus der Leistungsfähigkeit eines Modells entstehen, sondern aus dessen Einbettung in reale Nutzungssituationen. Entscheidend ist daher nicht nur, was ein KI-Modell grundsätzlich kann, sondern was ein KI-System im jeweiligen Kontext tatsächlich tun darf, wen es betrifft, welche Schutzgüter berührt werden und wie rückholbar mögliche Folgen sind.

ASEI-10 verdichtet diese Grundannahme in vier Bewertungsdimensionen: Autonomie, Severität, Exposition und Irreversibilität. Autonomie beschreibt die operative Selbstständigkeit eines KI-Systems. Severität erfasst die mögliche Schwere negativer Folgen im Einsatzbereich. Exposition beschreibt die Breite des möglichen Wirkungsraums. Irreversibilität bewertet, wie schwer negative Folgen korrigiert, zurückgeholt oder kompensiert werden können.

Die vier Dimensionen werden zunächst getrennt bewertet und anschließend rechnerisch zusammengeführt. Die qualitative Autonomiestufe wird hierfür auf einen normierten Autonomiewert A_n übertragen. Danach werden A_n, Severität, Exposition und Irreversibilität zu einem multiplikativen Rohwert verbunden, der logarithmisch auf eine Skala von 1 bis 10 normiert wird. Dadurch entsteht ein kompakter ASEI-10-Score, der unterschiedliche KI-Systeme vergleichbar macht. Gleichzeitig verhindern ergänzende Prüfschwellen, dass qualitativ bedeutsame Einzelmerkmale durch den Gesamtscore verdeckt werden.

13.2 Zentrale Erkenntnis

Die zentrale Erkenntnis des ASEI-10-Modells lautet: Das Risikopotenzial eines KI-Systems entsteht aus dem Zusammenspiel von Autonomie, Severität, Exposition und Irreversibilität. Keine dieser Dimensionen genügt allein, um das Risikopotenzial eines KI-Systems angemessen zu erfassen. Erst ihre gemeinsame Betrachtung macht sichtbar, warum ein System folgenarm, moderat, erhöht, hoch oder sehr hoch risikobehaftet sein kann.

Ein hochautonomes System ist nicht automatisch hochriskant, wenn es in einem eng begrenzten, folgenarmen und gut rückholbaren Kontext eingesetzt wird. Umgekehrt kann ein nur begrenzt autonomes System ein hohes Risikopotenzial aufweisen, wenn es in einem folgenreichen Bereich eingesetzt wird, große Personengruppen betrifft oder schwer reversible Folgen auslösen kann. Genau diese Differenzierung soll ASEI-10 leisten.

Damit wendet sich das Modell gegen zwei Vereinfachungen. Einerseits reicht es nicht aus, KI-Risiken allein aus der Modellgröße, Leistungsfähigkeit oder technischen Fortschrittlichkeit abzuleiten. Andererseits reicht es ebenso wenig, nur auf den Einsatzbereich zu schauen und Autonomie, Exposition oder Irreversibilität auszublenden. ASEI-10 verbindet technische, organisatorische und wirkungsbezogene Perspektiven in einem gemeinsamen Bewertungsrahmen.

13.3 Praktischer Nutzen des Modells

Der praktische Nutzen des ASEI-10-Modells liegt in seiner Anwendbarkeit auf unterschiedliche Organisationen, Branchen und Nutzungskontexte. Das Modell kann für KI-Inventare, interne Freigabeprozesse, Risikoanalysen, Projektbewertungen, Schulungen, Governance-Entscheidungen, Auditvorbereitungen oder Compliance-Vorprüfungen eingesetzt werden. Es eignet sich sowohl für einfache assistive Systeme als auch für agentische oder systemkritische KI-Anwendungen.

Besonders hilfreich ist die Trennung von Score und Risikotreibern. Der ASEI-10-Score liefert eine kompakte Gesamteinstufung. Die Einzeldimensionen zeigen, wodurch diese Einstufung zustande kommt. Die Prüfschwellen benennen zusätzliche Mindestanforderungen. Dadurch kann eine Organisation nicht nur feststellen, dass ein KI-System risikorelevant ist, sondern auch, warum es risikorelevant ist und welche Schutzmaßnahmen naheliegen.

Das Modell unterstützt außerdem eine wiederholbare Bewertungslogik. Unterschiedliche KI-Systeme können nach denselben Kriterien bewertet werden. Das verbessert Vergleichbarkeit, Priorisierung und Dokumentation. Gleichzeitig bleibt Raum für fachliche Begründung, weil jede Einstufung an den konkreten Anwendungskontext gebunden ist.

13.4 Anschlussfähigkeit an Regulierung und Governance

ASEI-10 ist kein Ersatz für bestehende rechtliche, normative oder organisatorische Rahmenwerke. Es ersetzt weder den EU AI Act noch das NIST AI Risk Management Framework, ISO/IEC 23894, ISO/IEC 42001 oder andere Governance- und Compliance-Anforderungen. Sein Beitrag liegt vielmehr darin, diese Rahmenwerke durch eine operative Klassifikationslogik zu ergänzen.

Im Verhältnis zum EU AI Act kann ASEI-10 helfen, Systeme vorzustrukturieren, Risikotreiber sichtbar zu machen und eine vertiefte rechtliche Prüfung zu priorisieren. Im Verhältnis zu NIST und ISO kann das Modell als Bewertungsinstrument in bestehende Risikomanagement- und Managementsystemprozesse eingebunden werden. Es liefert damit keine vollständige Governance-Architektur, aber ein methodisches Werkzeug innerhalb einer solchen Architektur.

Diese Anschlussfähigkeit ist wichtig, weil Organisationen zunehmend vor der Aufgabe stehen, sehr unterschiedliche KI-Systeme zu erfassen, zu bewerten und zu steuern. ASEI-10 kann dabei helfen, aus einer abstrakten Risikodiskussion eine konkrete, dokumentierbare und überprüfbare Bewertungspraxis zu entwickeln.

13.5 Grenzen und Verantwortung

Die Stärke des ASEI-10-Modells liegt in seiner Strukturierungskraft, nicht in einer automatischen Sicherheitsgarantie. Ein niedriger Score beweist nicht, dass ein System fehlerfrei, rechtlich unproblematisch oder technisch sicher ist. Ein hoher Score bedeutet nicht automatisch, dass ein System unzulässig ist. Der Score ist eine Klassifikation des Risikopotenzials, kein abschließendes Urteil.

ASEI-10 bewertet außerdem nicht die Eintrittswahrscheinlichkeit konkreter Schadensereignisse. Zuverlässigkeit, Fehlerraten, Robustheit, Halluzinationsneigung, Missbrauchswahrscheinlichkeit oder technische Sicherheitsleistung müssen bei Bedarf gesondert untersucht werden. Das Modell kann anzeigen, welche Systeme aufgrund ihrer Struktur besonders prüfbedürftig sind; es ersetzt aber keine technische Evaluation und keine empirische Risikoanalyse.

Die Verantwortung für den Einsatz eines KI-Systems bleibt bei den handelnden Personen und Organisationen. Sie müssen prüfen, ob die Systembeschreibung vollständig ist, ob die Einstufungen plausibel begründet sind, ob rechtliche Vorgaben eingehalten werden, ob technische Sicherheitsmaßnahmen ausreichen und ob fachliche Kontrolle gewährleistet ist. ASEI-10 kann diese Verantwortung unterstützen, aber nicht ersetzen.

Besonders wichtig ist die regelmäßige Neubewertung. KI-Systeme verändern sich durch neue Modelle, neue Werkzeuge, zusätzliche Berechtigungen, andere Datenbestände, weitere Nutzergruppen oder veränderte Einsatzbereiche. Eine einmalige Einstufung bleibt nur so lange belastbar, wie der bewertete Systemzustand und Anwendungskontext unverändert bleiben.

13.6 Weiterentwicklung und Validierung

Das ASEI-10-Modell ist als methodischer Vorschlag zu verstehen, der weiterentwickelt und validiert werden kann. Eine sinnvolle Weiterentwicklung bestünde zunächst darin, das Modell auf eine größere Zahl realer oder realistischer Fallbeispiele anzuwenden. Dadurch ließe sich prüfen, ob die Skalen trennscharf genug sind, ob die Prüfschwellen praxistauglich greifen und ob die Interpretationsbereiche des Scores angemessen gewählt sind.

Darüber hinaus könnte das Modell branchenspezifisch ergänzt werden. Für Bildung, Gesundheitswesen, Finanzdienstleistungen, öffentliche Verwaltung, industrielle Produktion, kritische Infrastruktur oder kreative Anwendungen könnten zusätzliche Leitfragen, Beispiele und Mindestanforderungen entwickelt werden. Die Grundstruktur aus Autonomie, Severität, Exposition und Irreversibilität bliebe dabei erhalten, würde aber für bestimmte Anwendungsfelder präzisiert.

Ein weiterer Entwicklungsschritt bestünde in einer empirischen oder expertengestützten Validierung. Fachleute aus Technik, Recht, Risikomanagement, Ethik, Datenschutz, Bildung und Organisation könnten unabhängige Bewertungen vornehmen und die Verständlichkeit, Praktikabilität und Nachvollziehbarkeit des Modells prüfen. Auf dieser Grundlage ließen sich Grenzfälle schärfen, Beispiele erweitern, Gewichtungen überprüfen und Interpretationsbereiche kalibrieren.

Besonders naheliegend wäre als erster Validierungsschritt eine kleine Inter-Rater-Studie. Dabei würden mehrere fachkundige Personen dieselben Fallbeispiele unabhängig voneinander anhand des ASEI-10-Modells bewerten. Verglichen würden insbesondere die Einstufungen der vier Dimensionen, die Begründungen der Grenzfälle und die daraus resultierenden Scores. Eine solche Studie könnte zeigen, ob unterschiedliche Bewerterinnen und Bewerter bei gleichem Bewertungsgegenstand zu vergleichbaren Ergebnissen gelangen oder ob systematische Abweichungen auftreten. Damit ließe sich zugleich erkennen, an welchen Stellen die Skalenbeschreibungen, Prüfschwellen oder Anwendungshinweise weiter geschärft werden müssen.

13.7 Ausblick

Mit ASEI-10 liegt ein Klassifikationsmodell vor, das KI-Risiken nicht auf einzelne technische Eigenschaften reduziert, sondern in ihrem konkreten Anwendungskontext bewertet. Der Ansatz verbindet qualitative Analyse, dimensionsbezogene Strukturierung, rechnerische Indexbildung und schwellenbasierte Mindestanforderungen. Dadurch kann das Modell eine Brücke zwischen abstrakter KI-Risikodiskussion und praktischer Governance schlagen.

Die weitere Entwicklung von KI-Systemen wird diesen Bedarf voraussichtlich verstärken. Mit zunehmender Agentik, Toolintegration, Persistenz, Multimodalität, Automatisierung und Einbettung in reale Prozesse wird es immer wichtiger, nicht nur die Leistungsfähigkeit von Modellen zu betrachten, sondern ihre operative Wirkung im jeweiligen Systemkontext. Genau hier setzt ASEI-10 an.

Das Modell erhebt keinen Anspruch auf abschließende Vollständigkeit oder empirische Validierung in seiner jetzigen Form. Sein Wert liegt in einer klaren, nachvollziehbaren und erweiterbaren Bewertungslogik. Es kann Organisationen dabei unterstützen, KI-Systeme differenzierter zu beurteilen, Risikotreiber sichtbar zu machen, Prioritäten zu setzen und Verantwortungsstrukturen zu stärken.

ASEI-10 fasst diese Logik in einem einfachen Grundsatz zusammen:

Das Risikopotenzial eines KI-Systems entsteht nicht allein aus seiner Modellleistung, sondern aus dem Zusammenspiel von Autonomie, Severität, Exposition und Irreversibilität.

Literatur- und Quellenverzeichnis

Rechtsquellen und regulatorische Grundlagen

Europäische Kommission (o. J.): AI Act. Shaping Europe's Digital Future. Directorate-General for Communications Networks, Content and Technology.
https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
Abrufdatum: 22.06.2026.

Europäisches Parlament; Rat der Europäischen Union (2024): Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Artificial Intelligence Act). Amtsblatt der Europäischen Union, OJ L, 2024/1689, 12.07.2024.
https://data.europa.eu/eli/reg/2024/1689/oj
Abrufdatum: 22.06.2026.

Allgemeines Risikomanagement, Bewertungsmethodik und Risikomatrizen

Cox, Louis Anthony Jr. (2008): What's Wrong with Risk Matrices? In: Risk Analysis, Vol. 28, No. 2, S. 497–512. https://doi.org/10.1111/j.1539-6924.2008.01030.x

IEC (2018): IEC 60812:2018 – Failure modes and effects analysis (FMEA and FMECA).
Geneva: International Electrotechnical Commission.

IEC (2019): IEC 31010:2019 – Risk management — Risk assessment techniques.
Geneva: International Electrotechnical Commission.

ISO (2018): ISO 31000:2018 – Risk management — Guidelines.
Geneva: International Organization for Standardization.

KI-Risikomanagement, KI-Governance und Normen

DIN (2019): DIN SPEC 92001-1:2019-04 – Artificial Intelligence — Life Cycle Processes and Quality Requirements – Part 1: Quality Meta Model. Berlin: Deutsches Institut für Normung.
https://doi.org/10.31030/3050203

DIN (2020): DIN SPEC 92001-2:2020-12 – Artificial Intelligence — Life Cycle Processes and Quality Requirements – Part 2: Robustness. Berlin: Deutsches Institut für Normung.
https://doi.org/10.31030/3205018

ISO/IEC (2022): ISO/IEC 22989:2022 – Information technology — Artificial intelligence — Artificial intelligence concepts and terminology.
Geneva: International Organization for Standardization / International Electrotechnical Commission.

ISO/IEC (2022): ISO/IEC 23053:2022 – Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML).
Geneva: International Organization for Standardization / International Electrotechnical Commission.

ISO/IEC (2023): ISO/IEC 23894:2023 – Information technology — Artificial intelligence — Guidance on risk management.
Geneva: International Organization for Standardization / International Electrotechnical Commission.

ISO/IEC (2023): ISO/IEC 42001:2023 – Information technology — Artificial intelligence — Management system.
Geneva: International Organization for Standardization / International Electrotechnical Commission.

National Institute of Standards and Technology (NIST) (2023): Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. Gaithersburg, MD: U.S. Department of Commerce, National Institute of Standards and Technology.
https://doi.org/10.6028/NIST.AI.100-1

National Institute of Standards and Technology (NIST) (2024): Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. NIST AI 600-1. Gaithersburg, MD: U.S. Department of Commerce, National Institute of Standards and Technology.
https://doi.org/10.6028/NIST.AI.600-1

AI-Risk-Taxonomien und Klassifikationsansätze

MIT AI Risk Initiative (o. J.): MIT AI Risk Repository. Living database and taxonomy of AI risks.
https://airisk.mit.edu/
Abrufdatum: 22.06.2026.

Slattery, Peter; Saeri, Alexander K.; Grundy, Emily A. C.; Graham, Jess; Noetel, Michael; Uuk, Risto; Dao, James; Pour, Soroush; Casper, Stephen; Thompson, Neil (2024): The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence.
https://doi.org/10.48550/arXiv.2408.12622

Uuk, Risto; Gutierrez, Carlos Ignacio; Guppy, Daniel; Lauwaert, Lode; Kasirzadeh, Atoosa; Velasco, Lucia; Slattery, Peter; Prunkl, Carina (2024): A Taxonomy of Systemic Risks from General-Purpose AI.
https://doi.org/10.48550/arXiv.2412.07780

Zeng, Yi; Klyman, Kevin; Zhou, Andy; Yang, Yu; Pan, Minzhou; Jia, Ruoxi; Song, Dawn; Liang, Percy; Li, Bo (2024): AI Risk Categorization Decoded (AIR 2024): From Government Regulations to Corporate Policies.
https://doi.org/10.48550/arXiv.2406.17864

Internationale AI-Safety-Berichte

Bengio, Yoshua et al. (2024): International Scientific Report on the Safety of Advanced AI. Interim Report. International Scientific Report on Advanced AI Safety.
https://doi.org/10.48550/arXiv.2412.05282

Bengio, Yoshua et al. (2025): International AI Safety Report 2025. DSIT 2025/001.
https://internationalaisafetyreport.org/publication/international-ai-safety-report-2025
Abrufdatum: 22.06.2026

Bengio, Yoshua et al. (2026): International AI Safety Report 2026. DSIT 2026/001.
https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026
Abrufdatum: 22.06.2026

KI-Prüfung, Qualitätsstandards und institutionelle KI-Sicherheit

AIQI Consortium (o. J.): AIQI Member Spotlight – TÜV AI.Lab.
https://www.aiqi.org/newsroom/member-spotlight-tuv-ai-lab
Abrufdatum: 22.06.2026.

AI Quality & Testing Hub GmbH (o. J.): Offizielle Website.
https://aiqualityhub.com/en/home/
Abrufdatum: 22.06.2026.

Bundesregierung (o. J.): Der Nationale Sicherheitsrat.
https://www.bundesregierung.de/breg-de/bundesregierung/nationaler-sicherheitsrat

Abrufdatum: 22.06.2026.

CertifAI GmbH (o. J.): Offizielle Website / News.
https://www.getcertif.ai/news
Abrufdatum: 22.06.2026.

DLR – Deutsches Zentrum für Luft- und Raumfahrt (o. J.): Institut für KI-Sicherheit – Abteilungen.
https://www.dlr.de/de/ki/ueber-uns/abteilungen
Abrufdatum: 22.06.2026.

DLR – Deutsches Zentrum für Luft- und Raumfahrt (o. J.): Institut für KI-Sicherheit – Sicherheitskritische KI-Anwendungen.
https://www.dlr.de/de/ki/forschung-transfer/themen/sicherheitskritische-ki-anwendungen
Abrufdatum: 22.06.2026.

Fraunhofer IAIS (o. J.): Vertrauenswürdige KI. Abschnitt „MISSION KI“.
https://www.iais.fraunhofer.de/de/kuenstliche-intelligenz/vertrauenswuerdige-ki.html
Abrufdatum: 22.06.2026.

Fraunhofer IAIS (2023): KI-Prüfkatalog. Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz.
https://www.iais.fraunhofer.de/en/publications/studies/2023/ai-assessment-catalog.html
Direkter PDF-Verweis: https://www.iais.fraunhofer.de/s/ki-pruefkatalog/epaper/ausgabe.pdf
Abrufdatum: 22.06.2026.

Hessisches Ministerium für Digitalisierung und Innovation (o. J.): AI Quality & Testing Hub.
https://digitales.hessen.de/kuenstliche-intelligenz/ki-handlungsfelder/ki-innovation-und-ki-anwendungen-foerdern/ai-quality-testing-hub
Abrufdatum: 22.06.2026.

MISSION KI (o. J.): National Initiative for AI & Data Economy.
https://mission-ki.de/en/mission-ki-national-initiative-developed
Abrufdatum: 22.06.2026.

MISSION KI (2025): MISSION KI Qualitätsstandard für Niedrigrisiko-KI. KI-Qualität messbar machen.
https://mission-ki.de/de/pruefstandards
Abrufdatum: 22.06.2026.

Presse- und Informationsamt der Bundesregierung (2026): Sitzung des Nationalen Sicherheitsrates im Juni 2026. Pressemitteilung 103 vom 08.06.2026.
https://www.bundesregierung.de/breg-de/aktuelles/sitzung-des-nationalen-sicherheitsrates-im-juni-2026-2438050
Abrufdatum: 22.06.2026.

TÜV AI.Lab (o. J.): Offizielle Website.
https://www.tuev-lab.ai/
Abrufdatum: 22.06.2026.

VDE (2025): Prüfbare KI-Qualität: MISSION KI präsentiert Qualitätsstandard und digitales Prüfportal für Niedrigrisiko-KI. Pressemitteilung vom 26.11.2025.
https://www.vde.com/de/presse/pressemitteilungen/mission-ki-praesentiert-qualitaetsstandard
Abrufdatum: 22.06.2026.

VDE (2026): Quality Evaluation of Low-Risk AI Systems. Bewertung von Niedrigrisiko-KI-Systemen auf Grundlage des MISSION-KI-Qualitätsstandards.
https://www.vde.com/topics-de/kuenstliche-intelligenz/beratung/bewertung-niedrig-risiko-ki
Abrufdatum: 22.06.2026.

Wissenschaftstheorie und Forschungsmethodik

Hevner, Alan R.; March, Salvatore T.; Park, Jinsoo; Ram, Sudha (2004): Design Science in Information Systems Research. In: MIS Quarterly, Vol. 28, No. 1, S. 75–105.
https://doi.org/10.2307/25148625

Krippendorff, Klaus (2019): Content Analysis. An Introduction to Its Methodology. 4. Auflage. Los Angeles/London: SAGE Publications. [Erstveröffentlichung 1980.]

Popper, Karl R. (2005): Logik der Forschung. Hrsg. von Herbert Keuth. 11., durchgesehene und ergänzte Auflage. Tübingen: Mohr Siebeck (Gesammelte Werke in deutscher Sprache, Bd. 3). [Erstveröffentlichung 1935, Wien: Julius Springer.]

Weiterführende Grundlagenliteratur zu künstlicher Intelligenz

Russell, Stuart (2019): Human Compatible: Artificial Intelligence and the Problem of Control. New York: Viking.

Russell, Stuart; Norvig, Peter (2021): Artificial Intelligence: A Modern Approach. 4th edition. Hoboken: Pearson.

Anhang A: Begriffsverzeichnis

ASEI-10
Bezeichnung des Klassifikationsmodells zur Bewertung des Risikopotenzials konkreter KI-Systeme in konkreten Anwendungskontexten. Der Name verweist auf die vier Bewertungsdimensionen Autonomie, Severität, Exposition und Irreversibilität sowie auf die normierte Gesamtskala von 1 bis 10.

ASEI-10-Score
Normierter Klassifikationswert des Modells. Er wird aus dem multiplikativen Rohwert der vier Bewertungsdimensionen gebildet und logarithmisch auf eine Skala von 1 bis 10 übertragen. Der Score dient der zusammenfassenden Einordnung des Risikopotenzials, ist aber keine exakte Messgröße und kein vollständiger Risikowert im klassischen Sinn.

Agentisches KI-System
KI-System, das über eine rein reaktive Assistenz hinausgeht und Zielvorgaben in Handlungsschritte übersetzen, Werkzeuge auswählen, Zwischenergebnisse verarbeiten oder innerhalb definierter Grenzen eigenständig handeln kann.

Anwendungskontext
Konkreter Einsatzrahmen eines KI-Systems. Dazu gehören Einsatzzweck, Nutzergruppen, Datenarten, Berechtigungen, technische Einbettung, organisatorische Umgebung, mögliche Außenwirkung und realistische Folgen einer fehlerhaften, missbräuchlichen oder unzureichend kontrollierten Nutzung.

Autonomie
Bewertungsdimension des ASEI-10-Modells. Sie beschreibt den Grad der operativen Selbstständigkeit eines KI-Systems. Die qualitative Skala reicht von passiven oder reaktiven Assistenzsystemen bis zu persistenten, verteilten, selbstoptimierenden oder nicht zuverlässig kontrollierbaren agentischen Systemen.

Autonomiestufe A
Qualitative Einstufung der Autonomie auf der neunstufigen Skala A1 bis A9. Sie beschreibt die operative Selbstständigkeit des KI-Systems. Für die Scorebildung wird die qualitative Autonomiestufe nicht direkt verwendet, sondern in den normierten Autonomiewert A_n übertragen.

Bewertungsbogen
Standardisiertes Dokumentationsinstrument zur Durchführung und Nachvollziehbarkeit einer ASEI-10-Bewertung. Der Bewertungsbogen hält Bewertungsgegenstand, Bewertungsmodus, Systemprofil, Wirkungskontext, Einzeleinstufungen der vier ASEI-Dimensionen, Scorebildung, Prüfschwellen, Mindestanforderungen und Anlässe für eine Neubewertung fest. Er ersetzt nicht die fachliche Bewertung, sondern macht sie prüfbar, vergleichbar und dokumentierbar.

Bewertungsdimension
Eine der vier Grunddimensionen des ASEI-10-Modells. Bewertet werden Autonomie, Severität, Exposition und Irreversibilität. Jede Dimension wird zunächst eigenständig eingestuft und anschließend in die Scorebildung einbezogen.

Bewertungsgegenstand

Das konkret zu bewertende KI-System im konkret beschriebenen Nutzungskontext. Bewertet wird nicht „KI“ im Allgemeinen und nicht ein Modell in abstrakter Form, sondern ein bestimmtes System mit bestimmten Funktionen, Rechten, Datenzugriffen, Nutzern und Einsatzgrenzen.

Bewertungsmodus

Festlegung, unter welchen Annahmen eine ASEI-10-Bewertung durchgeführt wird. Im Modell werden insbesondere Brutto-Bewertung, Netto-Bewertung und die Bewertung eines systemimmanent begrenzten Systemdesigns unterschieden. Der Bewertungsmodus ist wichtig, weil er bestimmt, ob Schutz- und Begrenzungsmaßnahmen bereits in die Dimensionseinstufung einbezogen werden oder lediglich als ergänzende Governance-Maßnahmen dokumentiert werden.

Brutto-Bewertung

Bewertung des Risikopotenzials eines KI-Systems ohne Berücksichtigung zusätzlicher Schutz-, Kontroll- oder Begrenzungsmaßnahmen. Die Brutto-Bewertung beschreibt das strukturelle Ausgangspotenzial des Systems, wie es sich aus Funktionen, Rechten, Datenzugriffen, Nutzergruppen, Einsatzbereich und möglichen Folgen ergibt. Sie eignet sich insbesondere, um sichtbar zu machen, welches Risikopotenzial ein System ohne wirksame Mitigation besitzt.

Design-Science-Artefakt

Ein im Rahmen der Design-Science-Forschung entwickeltes methodisches, technisches oder organisatorisches Ergebnis, das ein praktisches Problem adressiert und zugleich wissenschaftlich reflektiert wird. Ein Design-Science-Artefakt kann beispielsweise ein Modell, ein Verfahren, ein Bewertungsinstrument, eine Methode, ein Prototyp oder ein konzeptioneller Rahmen sein.

Im Kontext dieses Papiers ist das ASEI-10-Modell selbst als Design-Science-Artefakt zu verstehen. Es wurde als Klassifikationsmodell entwickelt, um das strukturelle Risikopotenzial konkreter KI-Systeme in konkreten Anwendungskontexten nachvollziehbar zu bewerten. Dies bedeutet nicht, dass ASEI-10 bereits empirisch validiert ist. Die wissenschaftliche Qualität des Artefakts ergibt sich vielmehr aus der nachvollziehbaren Problembeschreibung, der begründeten Konstruktion des Modells, der Demonstration seiner Anwendung anhand von Fallbeispielen und der Möglichkeit einer späteren systematischen Evaluation.

Dimensionsprofil

Darstellung der vier ASEI-Dimensionen Autonomie, Severität, Exposition und Irreversibilität als gemeinsames Risikoprofil eines KI-Systems. Das Dimensionsprofil zeigt, welche Dimensionen eine Einstufung besonders tragen. Dadurch wird sichtbar, ob ein System vor allem wegen hoher Autonomie, eines sensiblen Einsatzbereichs, breiter Exposition, schwer reversibler Folgen oder wegen des Zusammenwirkens mehrerer Dimensionen risikorelevant ist.

Eintrittswahrscheinlichkeit

Wahrscheinlichkeit, mit der ein bestimmtes Schadensereignis tatsächlich eintritt. Die Eintrittswahrscheinlichkeit ist kein Bestandteil des ASEI-10-Scores. ASEI-10 bewertet das strukturelle Risikopotenzial eines KI-Systems; Wahrscheinlichkeiten müssen bei Bedarf ergänzend durch technische Evaluation, Fehlerratenanalyse, Monitoring, Red Teaming oder Betriebserfahrung bestimmt werden.

Exposition
Bewertungsdimension des ASEI-10-Modells. Sie beschreibt die Breite des möglichen Wirkungsraums eines KI-Systems. Exposition erfasst, ob Wirkungen auf eine Einzelperson begrenzt bleiben, Gruppen betreffen, Organisationseinheiten erreichen, öffentlich sichtbar werden oder systemische Reichweite entfalten können.

FMEA
Abkürzung für *Failure Mode and Effects Analysis*, deutsch meist Fehlermöglichkeits- und Einflussanalyse. FMEA ist ein Verfahren zur systematischen Untersuchung möglicher Fehlerarten, ihrer Ursachen und ihrer Auswirkungen in technischen Systemen, Produkten oder Prozessen. Im Kontext des ASEI-10-Modells kann FMEA als ergänzende Detailanalyse eingesetzt werden, wenn ein KI-System aufgrund seines Risikopotenzials einer vertieften technischen oder prozessualen Prüfung bedarf.

FMECA
Abkürzung für *Failure Mode, Effects and Criticality Analysis*. FMECA erweitert die FMEA um eine zusätzliche Bewertung der Bedeutung und Priorität möglicher Fehlerfolgen im jeweiligen System- oder Prozesskontext. Im ASEI-10-Modell ersetzt FMECA nicht die Klassifikation des strukturellen Risikopotenzials, kann aber als nachgelagerte Analyse genutzt werden, um konkrete Fehlermöglichkeiten, ihre Ursachen, ihre Folgen und ihre Dringlichkeit detaillierter zu bewerten. Der Begriff *Criticality* ist Bestandteil der etablierten englischen Methodenbezeichnung und nicht mit der Bewertungsdimension Severität im ASEI-10-Modell gleichzusetzen.

General-Purpose-AI-Modell
KI-Modell mit breitem Einsatzspektrum, das für unterschiedliche Aufgaben und Anwendungskontexte verwendet oder angepasst werden kann. Im ASEI-10-Modell wird ein solches Modell nicht isoliert bewertet, sondern erst in seiner konkreten Systemeinbettung und Nutzung.

Governance
Gesamtheit der Regeln, Prozesse, Verantwortlichkeiten und Kontrollmechanismen, mit denen eine Organisation den Einsatz von KI-Systemen steuert. ASEI-10 kann als Klassifikations- und Priorisierungsinstrument innerhalb einer solchen Governance-Struktur verwendet werden.

Grenznahe Einstufung
Bewertungssituation, in der der ungerundete ASEI-10-Score unmittelbar an einer Klassengrenze liegt. In solchen Fällen darf die Einstufung nicht mechanisch anhand eines gerundeten Werts interpretiert werden. Maßgeblich sind der ungerundete Score, die Einzeldimensionen, ausgelöste Prüfschwellen und die fachliche Begründung. Grenznahe Einstufungen erfordern regelmäßig besondere Plausibilitätsprüfung.

Heuristischer Klassifikationsindex
Ein Index, der komplexe Bewertungsmerkmale strukturiert zusammenführt, um Orientierung, Priorisierung und Vergleichbarkeit zu ermöglichen. Der ASEI-10-Score ist ein heuristischer Klassifikationsindex und keine exakte quantitative Risikomessung.

Inter-Rater-Reliabilität
Maß dafür, inwieweit unterschiedliche Bewerterinnen und Bewerter bei Anwendung desselben Bewertungsmodells auf denselben Bewertungsgegenstand zu übereinstimmenden oder hinreichend ähnlichen Ergebnissen gelangen. Im ASEI-10-Modell wäre die Inter-Rater-Reliabilität besonders wichtig, um zu prüfen, ob die Einstufungen der vier Dimensionen Autonomie, Severität, Exposition und Irreversibilität auch durch verschiedene fachkundige Personen nachvollziehbar und konsistent vorgenommen werden können.

Irreversibilität
Bewertungsdimension des ASEI-10-Modells. Sie beschreibt, wie schwer negative Folgen eines KI-Systems nachträglich erkannt, gestoppt, korrigiert, zurückgeholt oder kompensiert werden können.

KI-Modell
Technischer Modellkern, der Eingaben verarbeitet und Ausgaben erzeugt. Ein KI-Modell kann Bestandteil eines KI-Systems sein, ist aber nicht mit dem gesamten System gleichzusetzen.

KI-System
Konkrete technische und organisatorische Einheit, in der ein KI-Modell mit Funktionen, Schnittstellen, Werkzeugen, Datenzugriffen, Nutzerrollen, Berechtigungen und Einsatzgrenzen verbunden ist. ASEI-10 bewertet KI-Systeme, nicht isolierte Modelle.

Klassische Risikoanalyse
Risikobetrachtung, bei der Risiko häufig als Verbindung von Eintrittswahrscheinlichkeit und Schadensausmaß verstanden wird. ASEI-10 grenzt sich davon ab, weil das Modell keine Eintrittswahrscheinlichkeit berechnet, sondern das strukturelle Risikopotenzial eines KI-Systems klassifiziert.

Logarithmische Normierung
Rechenschritt, mit dem der multiplikative Rohwert des ASEI-10-Modells auf eine Skala von 1 bis 10 übertragen wird. Die Normierung erhält die Rangordnung des Rohwerts, verdichtet aber hohe Rohwerte und macht die Ergebnisse besser interpretierbar.

Mindestanforderung
Aus einer Prüfschwelle abgeleitete grundlegende Anforderung an Kontrolle, Dokumentation, Freigabe, Monitoring, Rückholbarkeit oder Governance. Mindestanforderungen gelten unabhängig davon, ob der Gesamtscore allein bereits eine hohe Risikostufe nahelegt.

Mitigation
Allgemein bezeichnet Mitigation Maßnahmen zur Verringerung, Begrenzung oder Beherrschung eines Risikos. Der Begriff wird im Risiko-, Sicherheits- und Krisenmanagement verwendet und ist nicht auf KI-Systeme beschränkt.

Im Kontext des ASEI-10-Modells bezeichnet Mitigation technische, organisatorische oder prozessuale Maßnahmen zur Verringerung des Risikopotenzials eines KI-Systems. Dazu können etwa Begrenzungen von Toolzugriffen, Freigabepunkte, Rollen- und Rechtekonzepte, Monitoring, Protokollierung, Rücksetzmechanismen, fachliche Prüfungen oder Beschwerdeverfahren gehören.

Im ASEI-10-Modell ist eine Mitigation nur dann scorewirksam, wenn sie mindestens eine Bewertungsdimension tatsächlich, nachweisbar und dauerhaft so reduziert, dass eine niedrigere Einstufung begründet ist. Andernfalls bleibt sie eine ergänzende Schutz- oder Governance-Maßnahme, ohne den ASEI-10-Score selbst zu senken.

Modellleistung

Technische Leistungsfähigkeit eines KI-Modells, etwa in Bezug auf Sprachverarbeitung, Analysefähigkeit, Problemlösung, Planung oder Generierung. Im ASEI-10-Modell ist Modellleistung nicht allein ausschlaggebend für das Risikopotenzial.

Netto-Bewertung

Bewertung des Risikopotenzials eines KI-Systems unter Berücksichtigung tatsächlich wirksamer, überprüfbarer und dauerhaft verankerter Schutz- oder Begrenzungsmaßnahmen. Eine Netto-Bewertung darf nur vorgenommen werden, wenn die Maßnahmen eine oder mehrere ASEI-Dimensionen tatsächlich reduzieren. Bloße organisatorische Absichtserklärungen, unverbindliche Hinweise oder nicht überprüfbare Kontrollen rechtfertigen keine niedrigere Einstufung.

Normierter Autonomiewert A_n

Für die Scorebildung verwendeter Rechenwert der Autonomie. Die qualitative Autonomiestufe A1 bis A9 wird nach der Formel $\boldsymbol{A_n = 1 + 4 \cdot \frac{A-1}{8}}$ auf den Wertebereich von 1 bis 5 übertragen. Dadurch werden alle vier Bewertungsdimensionen rechnerisch auf einem gemeinsamen Wertebereich geführt.

Mathematisch äquivalent lässt sich die Normierung auch vereinfacht als $\boldsymbol{A_n = \frac{A+1}{2}}$ darstellen. Die ausführlichere Formel macht jedoch den Normierungsschritt transparenter, weil sie die Übertragung von der neunstufigen Autonomieskala auf den fünfstufigen Wertebereich der übrigen Dimensionen sichtbar macht.

Ordinalskala

Geordnete Bewertungsskala, bei der höhere Stufen eine stärkere Ausprägung anzeigen, ohne dass die Abstände zwischen den Stufen exakt gleich groß sein müssen. Die Skalen des ASEI-10-Modells sind ordinal zu verstehen. Ihre rechnerische Verknüpfung dient der heuristischen Indexbildung, nicht der exakten Messung.

Prüfschwelle

Qualitativer Schwellenpunkt innerhalb einer Bewertungsdimension. Wird eine Prüfschwelle erreicht, sind zusätzliche Mindestanforderungen zu prüfen. Prüfschwellen verhindern, dass besonders relevante Risikomerkmale durch den Gesamtscore verdeckt werden.

Prüf- und Qualitätsbewertungsansatz

Ansatz zur strukturierten Bewertung von Qualität, Vertrauenswürdigkeit, Sicherheit oder Prüfbarkeit eines KI-Systems. Solche Ansätze können Kriterien, Prüffragen, Testmethoden, Schutzbedarfsanalysen, Dokumentationsanforderungen oder Prüfberichte umfassen. ASEI-10 ersetzt solche Ansätze nicht, kann aber vorgelagert helfen, risikorelevante Systeme zu identifizieren und zu priorisieren.

Red Teaming
Prüf- und Testverfahren, bei dem ein System gezielt aus der Perspektive möglicher Angreifer, Fehlanwender oder kritischer Nutzungsszenarien untersucht wird. Ziel ist es, Schwachstellen, Umgehungsmöglichkeiten, Missbrauchspotenziale oder unerwartetes Fehlverhalten sichtbar zu machen. Im Kontext von KI-Systemen kann Red Teaming dazu dienen, Sicherheitslücken, problematische Ausgaben, Manipulationsanfälligkeit, Prompt-Injection-Risiken, Tool-Missbrauch oder Kontrollprobleme zu identifizieren. Im ASEI-10-Modell ist Red Teaming kein Bestandteil der Scorebildung selbst, sondern ein ergänzendes Prüfverfahren, das insbesondere bei erhöhtem oder hohem Risikopotenzial angezeigt sein kann.

Reliabilität
Zuverlässigkeit oder Konsistenz eines Bewertungsverfahrens. Ein Modell ist reliabel, wenn es bei wiederholter Anwendung unter vergleichbaren Bedingungen zu stabilen und nachvollziehbaren Ergebnissen führt. Im Kontext von ASEI-10 betrifft Reliabilität insbesondere die Frage, ob die Skalenbeschreibungen, Prüfschwellen und Anwendungshinweise so eindeutig sind, dass Bewertungen nicht beliebig ausfallen.

Risikoklasse
Modellinterne Einordnungskategorie des ASEI-10-Scores. Die Risikoklasse ordnet den normierten Score einem qualitativen Risikopotenzial zu: sehr niedrig, niedrig, moderat, erhöht, hoch oder sehr hoch. Die Risikoklasse dient der Orientierung und Priorisierung, ersetzt aber nicht die Betrachtung der Einzeldimensionen, Prüfschwellen und fachlichen Begründung.

Risikopotenzial
Zentrale Zielgröße des ASEI-10-Modells. Risikopotenzial beschreibt die mögliche Gefährdungs- und Schadensrelevanz eines KI-Systems im konkreten Anwendungskontext. Es ergibt sich aus dem Zusammenspiel von Autonomie, Severität, Exposition und Irreversibilität. Risikopotenzial ist nicht gleichbedeutend mit Eintrittswahrscheinlichkeit.

Rohwert *R*
Nicht normiertes Risikoprodukt der vier Bewertungsdimensionen. Der Rohwert wird nach der Formel $\boldsymbol{R = A_n \cdot S \cdot E \cdot I}$ berechnet und anschließend logarithmisch auf die ASEI-10-Skala normiert.

Schutzbedarfsanalyse
Analyse, mit der bestimmt wird, welche Schutzgüter, Anforderungen oder möglichen Schäden in einem konkreten Anwendungskontext besonders relevant sind. Im Kontext von KI-Systemen kann eine Schutzbedarfsanalyse dazu dienen, die mögliche Schwere negativer Folgen zu bestimmen und daraus Prüf-, Qualitäts- oder Sicherheitsanforderungen abzuleiten. Für ASEI-10 ist sie besonders anschlussfähig an die Severitätsdimension.

Schwellenkriterium
Konkretes Merkmal, das das Erreichen einer Prüfschwelle begründet. Ein Schwellenkriterium beschreibt, warum eine bestimmte Mindestanforderung ausgelöst wird, etwa wegen agentischer Prozesssteuerung, hoher Severität, systemischer Exposition oder schwer reversibler Folgen.

Schwellenprüfung
Prüfung, ob innerhalb einer ASEI-Dimension qualitative Prüfschwellen erreicht werden, die zusätzliche Mindestanforderungen auslösen. Die Schwellenprüfung ergänzt den ASEI-10-Score um eine qualitative Kontrolllogik. Sie verhindert, dass besonders relevante Einzelmerkmale durch den Gesamtscore verdeckt werden. Eine ausgelöste Prüfschwelle verändert den Score nicht automatisch, kann aber zusätzliche Prüf-, Freigabe-, Kontroll- oder Dokumentationsanforderungen begründen.

Scorebildung
Rechnerische Zusammenführung der vier Bewertungsdimensionen. Im ASEI-10-Modell erfolgt sie zunächst über einen multiplikativen Rohwert und anschließend über eine logarithmische Normierung auf die Skala von 1 bis 10.

Scorewirksamkeit
Eigenschaft einer Schutz-, Kontroll- oder Begrenzungsmaßnahme, die dazu führt, dass sie bei der ASEI-10-Einstufung eine oder mehrere Bewertungsdimensionen tatsächlich senkt. Scorewirksam ist eine Maßnahme nur dann, wenn sie nachweisbar, dauerhaft und im konkreten System- oder Nutzungskontext wirksam ist. Nicht jede Governance-Maßnahme ist automatisch scorewirksam.

Severität
Bewertungsdimension des ASEI-10-Modells. Die Severität bezeichnet den Schweregrad möglicher negativer Folgen eines KI-Systems im konkreten Anwendungskontext. Bewertet wird nicht die Eintrittswahrscheinlichkeit, nicht die Reichweite der Wirkung und nicht die Rückholbarkeit möglicher Folgen, sondern die Bedeutung der berührten Schutzgüter und das mögliche Ausmaß negativer Konsequenzen.

Systemimmanent begrenztes Systemdesign
Systemzustand, bei dem risikobegrenzende Eigenschaften fest in Architektur, Berechtigungen, Funktionsumfang oder technische Einsatzgrenzen eines KI-Systems eingebaut sind. Anders als nachträgliche organisatorische Schutzmaßnahmen gehören diese Begrenzungen zum Systemdesign selbst. Sie können scorewirksam sein, wenn sie die Handlungsmöglichkeiten, Reichweite, Rückholbarkeit oder mögliche Folgenschwere tatsächlich und dauerhaft begrenzen.

Systemkontext
Gesamtheit der technischen, organisatorischen und funktionalen Bedingungen, unter denen ein KI-System eingesetzt wird. Dazu gehören Werkzeuge, Schnittstellen, Datenzugriffe, Nutzergruppen, Rechte, Kontrollmechanismen und mögliche Wirkungsräume.

Toolzugriff
Möglichkeit eines KI-Systems, externe Werkzeuge, Dateien, Datenbanken, Rechenfunktionen, Webzugänge, Kommunikationssysteme oder andere Anwendungen zu nutzen. Toolzugriff kann die Autonomie und das Risikopotenzial eines Systems erheblich erhöhen.

Validierung
Systematische Prüfung, ob ein Modell in der Anwendung belastbare, nachvollziehbare und hinreichend übereinstimmende Ergebnisse liefert. ASEI-10 ist in der vorliegenden Fassung ein methodischer Vorschlag und noch kein empirisch validiertes Messinstrument.

Validität
Gültigkeit oder sachliche Angemessenheit eines Bewertungsverfahrens. Ein Modell ist valide, wenn es tatsächlich das bewertet, was es zu bewerten beansprucht. Im Kontext von ASEI-10 bedeutet Validität, dass das Modell das strukturelle Risikopotenzial eines KI-Systems im Anwendungskontext angemessen erfasst und nicht lediglich scheinpräzise Zahlenwerte erzeugt.

Vertretbarkeitsprüfung
Prüfung, ob der Einsatz eines KI-Systems angesichts seines Risikopotenzials, seiner Kontrollierbarkeit, seiner Schutzmaßnahmen und seiner möglichen Folgen verantwortbar ist. Sie ist besonders bei hohen Scores oder kritischen Prüfschwellen erforderlich.

Wirkungskontext
Teil des Anwendungskontexts, der beschreibt, auf wen oder was sich die Wirkungen eines KI-Systems tatsächlich erstrecken können. Dazu gehören betroffene Personen, Nutzergruppen, Organisationseinheiten, externe Empfänger, technische Systeme, öffentliche Räume oder nachgelagerte Prozesse. Der Wirkungskontext ist besonders relevant für die Bewertung von Exposition und Irreversibilität, weil er sichtbar macht, wie weit mögliche Folgen reichen und wie schwer sie nachträglich begrenzt, korrigiert oder zurückgeholt werden können.

Anhang B: Bewertungsbogen

Der ASEI-10-Bewertungsbogen soll die fachliche Bewertung nicht ersetzen, sondern nachvollziehbar machen. Er ist nicht als bürokratisches Kontrollformular zu verstehen, sondern als Arbeits- und Dokumentationshilfe für eine strukturierte KI-Bewertung. Entscheidend ist daher nicht das bloße Ausfüllen einzelner Felder, sondern die begründete Dokumentation der Bewertung. Insbesondere die Einstufung der vier ASEI-Dimensionen, die Prüfung korrelativer Beziehungen sowie die Begründung möglicher Scorewirksamkeit von Schutz-, Kontroll- und Begrenzungsmaßnahmen müssen so festgehalten werden, dass sie später überprüft werden können.

Der Bewertungsbogen muss nicht zwingend manuell ausgefüllt werden. Gerade bei komplexeren KI-Systemen kann es sinnvoll sein, die Erstellung des Bewertungsbogens mithilfe geeigneter KI-Werkzeuge vorzubereiten, etwa durch die strukturierte Auswertung technischer Beschreibungen, Nutzungskonzepte, Rollen- und Rechtekonzepte, Datenschutzinformationen, Prozessbeschreibungen oder vorhandener Governance-Dokumente. Eine KI-gestützte Erstellung kann den Bewertungsprozess entlasten, ersetzt jedoch nicht die fachliche Prüfung und Verantwortung der zuständigen Personen oder Stellen. Die abschließende Einstufung, Begründung und Freigabe bleiben eine menschliche bzw. organisatorische Verantwortungsentscheidung.

Der in diesem Anhang abgedruckte Bewertungsbogen stellt die allgemeine Struktur für eine ASEI-10-Bewertung bereit. Er soll Organisationen dabei unterstützen, vorhandene Informationen geordnet zusammenzuführen, Risikotreiber sichtbar zu machen und notwendige Mindestanforderungen abzuleiten, ohne unnötige zusätzliche Bürokratie aufzubauen. Ein ausgefüllter Beispielbogen wird an dieser Stelle bewusst nicht aufgenommen, damit der Musterbogen als neutrales Arbeitsinstrument verwendbar bleibt. Die praktische Anwendung des Bewertungsbogens wird in Kapitel 11 anhand der Fallbeispiele erläutert. Die vollständig ausgefüllten Bewertungsbögen zu den sechs Fallbeispielen sind aus Gründen der Übersichtlichkeit in Anhang C zusammengeführt.

A. Grunddaten der Bewertung

Dieser Abschnitt identifiziert den Bewertungsgegenstand und die für die Bewertung verantwortlichen Personen oder Stellen. Er soll sicherstellen, dass später eindeutig nachvollziehbar ist, welches KI-System in welchem Nutzungskontext bewertet wurde und auf welchen Zeitpunkt sich die Bewertung bezieht.

Bearbeitungshinweis: Die Bezeichnung des KI-Systems sollte so konkret sein, dass keine Verwechslung mit anderen Modellen, Anwendungen, Versionen oder Einsatzvarianten möglich ist. Der Nutzungskontext sollte nicht nur allgemein benannt, sondern so beschrieben werden, dass Zweck, organisatorische Einbettung und praktische Verwendung des Systems erkennbar werden.

Bezeichnung des KI-Systems:

Bewerteter Nutzungskontext:

Bewertungsdatum:

Bewertende Person / Stelle:

Verantwortliche Stelle:

B. Bewertungsmodus und berücksichtigte Begrenzungen

Dieser Abschnitt legt fest, unter welchen Bewertungsannahmen die ASEI-10-Einstufung vorgenommen wird. Zu unterscheiden ist insbesondere, ob das Risikopotenzial ohne zusätzliche Schutzmaßnahmen, nach wirksam umgesetzten Schutzmaßnahmen oder auf Grundlage eines von vornherein technisch begrenzten Systemdesigns bewertet wird.

Bearbeitungshinweis: Bei einer Brutto-Bewertung wird das strukturelle Risikopotenzial ohne zusätzliche Schutz-, Kontroll- oder Begrenzungsmaßnahmen betrachtet. Bei einer Netto-Bewertung werden nur solche Maßnahmen berücksichtigt, die tatsächlich wirksam, überprüfbar und dauerhaft implementiert sind. Bei der Bewertung eines systemimmanent begrenzten Systemdesigns werden Begrenzungen berücksichtigt, die bereits fest in Architektur, Berechtigungen, Funktionsumfang oder technische Einsatzgrenzen eingebaut sind.

Hinweis zur Scorewirksamkeit: Schutz-, Kontroll- und Begrenzungsmaßnahmen dürfen bei der ASEI-10-Bewertung nur dann scorewirksam berücksichtigt werden, wenn sie eine oder mehrere Bewertungsdimensionen tatsächlich, nachweisbar und dauerhaft reduzieren. Bloße organisatorische Absichtserklärungen, unverbindliche Nutzungshinweise oder nicht überprüfbare Kontrollen rechtfertigen keine niedrigere Einstufung. Maßnahmen, die zwar sinnvoll sind, aber keine ASEI-Dimension unmittelbar senken, sind als ergänzende Governance- oder Mindestanforderungen zu dokumentieren.

Bewertungsmodus:

☐ Brutto-Bewertung

☐ Netto-Bewertung

☐ Bewertung des systemimmanent begrenzten Systemdesigns

Berücksichtigte Schutz- und Begrenzungsmaßnahmen:

Begründung der Scorewirksamkeit:

C. Systemprofil und Wirkungskontext

Dieser Abschnitt beschreibt die technische, organisatorische und funktionale Einbettung des KI-Systems. Er bildet die Grundlage für die spätere Einstufung der vier ASEI-Dimensionen. Entscheidend ist nicht nur, was das System technisch leisten kann, sondern auch, wer es nutzt, welche Daten verarbeitet werden, welche Schnittstellen bestehen und auf wen oder was sich seine Wirkungen erstrecken können.

Bearbeitungshinweis: Das Systemprofil sollte so beschrieben werden, dass die späteren Einstufungen von Autonomie, Severität, Exposition und Irreversibilität nachvollziehbar werden. Insbesondere ist festzuhalten, ob das System nur Informationen erzeugt, Entscheidungen vorbereitet, Prozesse beeinflusst oder selbst Handlungen auslösen kann. Der Wirkungskontext beschreibt, welche Personen, Organisationseinheiten, technischen Systeme, externen Empfänger oder nachgelagerten Prozesse von den Systemwirkungen betroffen sein können.

Nutzergruppen:

Verarbeitete Datenarten:

Werkzeugzugriffe oder Schnittstellen:

Mögliche Außenwirkung:

D. Bewertung der ASEI-Dimensionen

In diesem Abschnitt werden die vier Bewertungsdimensionen einzeln eingestuft. Jede Dimension muss eigenständig begründet werden. Die Einstufung soll nicht allein aus einem Gesamteindruck abgeleitet werden, sondern aus den jeweils maßgeblichen Merkmalen des Systems und seines Anwendungskontexts.

Bearbeitungshinweis: Die qualitative Autonomiestufe A wird auf der Skala A1 bis A9 bestimmt. Für die Scorebildung wird sie anschließend in den normierten Autonomiewert A_n im Wertebereich von 1 bis 5 übertragen. Severität, Exposition und Irreversibilität werden jeweils auf der Skala von 1 bis 5 bewertet. Für jede Dimension ist zu begründen, warum die gewählte Stufe angemessen ist und warum niedrigere oder höhere Stufen nicht passender erscheinen.

Autonomie	**A =**	**Begründung der Einstufung:**
Normierter Autonomiewert	**A_n =**	
Severität	**S =**	**Begründung der Einstufung:**
Exposition	**E =**	**Begründung der Einstufung:**
Irreversibilität	**I =**	**Begründung der Einstufung:**

E. Prüfung korrelativer Beziehungen zwischen den Dimensionen

Dieser Abschnitt dient der Plausibilitätsprüfung. Die vier ASEI-Dimensionen sind analytisch getrennt, können im konkreten Anwendungskontext aber sachlogisch zusammenwirken. Die Prüfung korrelativer Beziehungen soll verhindern, dass derselbe Risikotreiber unreflektiert mehrfach gezählt wird oder dass eine hohe Einstufung in einer Dimension automatisch auf andere Dimensionen übertragen wird.

Bearbeitungshinweis: Die Prüfung korrelativer Beziehungen verändert den ASEI-10-Score nicht automatisch. Sie dient der Plausibilitätskontrolle, der Vermeidung unreflektierter Mehrfachzählung und der späteren qualitativen Ergebnisinterpretation. Eine hohe Einstufung in einer Dimension rechtfertigt nicht automatisch eine hohe Einstufung in einer anderen Dimension. Jede mittlere oder hohe Einstufung muss eigenständig begründet werden können.

Eigenständige Begründung der Dimensionen: *Wurde jede Dimension unabhängig begründet, oder besteht die Gefahr, dass derselbe Risikotreiber mehrfach gezählt wurde?*

Relevante Wechselwirkungen: *Bestehen sachlogische Wechselwirkungen zwischen einzelnen Dimensionen, etwa zwischen Autonomie und Irreversibilität, Severität und Exposition, Severität und Irreversibilität oder Exposition und Irreversibilität?*

Plausibilitätsbewertung: *Erscheinen die Dimensionseinstufungen auch dann plausibel, wenn jede Dimension getrennt betrachtet wird? Die Dimensionseinstufungen erscheinen plausibel, wenn jede hohe oder mittlere Einstufung eigenständig begründet werden kann. Eine hohe Einstufung in einer Dimension rechtfertigt nicht automatisch eine hohe Einstufung in einer anderen Dimension.*

Auswirkung auf die Interpretation: *Welche Bedeutung haben die festgestellten Wechselwirkungen für die qualitative Interpretation der Bewertung? Die Prüfung korrelativer Beziehungen verändert den ASEI-10-Score nicht automatisch. Sie dient der Plausibilitätskontrolle, der Vermeidung unreflektierter Mehrfachzählung und der späteren qualitativen Ergebnisinterpretation.*

F. Scorebildung

In diesem Abschnitt werden die zuvor bestimmten Dimensionswerte rechnerisch zusammengeführt. Die Scorebildung erfolgt auf Grundlage des normierten Autonomiewerts A_n sowie der Werte für Severität, Exposition und Irreversibilität. Der resultierende Score ist als heuristischer Klassifikationsindex zu verstehen und nicht als exakte Messgröße.

Bearbeitungshinweis: Für die fachliche Bewertung sollte der ungerundete Score dokumentiert werden, weil grenznahe Einstufungen sonst verdeckt werden können. Der gerundete Score dient nur der übersichtlichen Darstellung. Die Risikoklasse ergibt sich aus der Zuordnung des Scores zu den definierten Risikoklassen des Modells. Sie ersetzt nicht die Betrachtung der Einzeldimensionen, Prüfschwellen und fachlichen Begründung.

Rohwert	$\boldsymbol{R = A_n \cdot S \cdot E \cdot I}$	**Ergebnis:**
ASEI-10-Score	$\mathbf{ASEI\text{-}10 = 1 + 9 \cdot \frac{\ln(R)}{\ln(625)}}$	**Ergebnis:**
Ungerundeter ASEI-10-Score:		
Risikoklasse:		

G. Prüfschwellen und Mindestanforderungen
Dieser Abschnitt prüft, ob qualitative Schwellenkriterien erreicht werden, die zusätzliche Mindestanforderungen auslösen. Prüfschwellen ergänzen die Scorelogik und verhindern, dass besonders relevante Einzelmerkmale durch den Gesamtscore verdeckt werden. **Bearbeitungshinweis:** Eine ausgelöste Prüfschwelle verändert den ASEI-10-Score nicht automatisch. Sie kann aber zusätzliche Anforderungen an Kontrolle, Freigabe, Dokumentation, Monitoring, Rollen- und Rechtebegrenzung, Rückholbarkeit, Datenschutz, Sicherheit oder regelmäßige Neubewertung begründen. Mindestanforderungen sollen so formuliert werden, dass erkennbar ist, welche organisatorische oder technische Konsequenz aus der Bewertung folgt.
Ausgelöste Prüfschwellen:
Begründung der ausgelösten Prüfschwellen:
Abgeleitete Mindestanforderungen:
Verantwortliche Stelle für Umsetzung oder Prüfung:
H. Neubewertung
Dieser Abschnitt dokumentiert, wann und unter welchen Bedingungen eine erneute ASEI-10-Bewertung erforderlich wird. Eine Bewertung gilt nur für den beschriebenen Systemzustand und den beschriebenen Anwendungskontext. Änderungen an Funktionen, Berechtigungen, Datenzugriffen, Nutzergruppen, Schnittstellen, Schutzmaßnahmen oder Einsatzbereichen können eine Neubewertung erforderlich machen. **Bearbeitungshinweis:** Eine Neubewertung ist insbesondere erforderlich, wenn sich Systemfunktionen, Toolzugriffe, Schnittstellen, Einsatzbereiche, Nutzergruppen, Datenarten, Kontrollmechanismen oder Außenwirkungen wesentlich verändern. Auch bekannte Fehlfunktionen, Sicherheitsvorfälle, Missbrauchsmöglichkeiten, geänderte rechtliche Anforderungen oder neue organisatorische Rahmenbedingungen können eine erneute Bewertung auslösen.
Anlass für Neubewertung:
Spätester Zeitpunkt der erneuten Bewertung:
Auslösende Änderungen oder Ereignisse:
Dokumentationshinweise:

Tabelle 28: Bewertungsbogen für eine ASEI-10-Bewertung

Anhang C: Bewertungsbögen zu den sechs Fallbeispielen

Bewertungsbogen zu Fallbeispiel 1: Privater Chatbot für kreative Texte

A. Grunddaten der Bewertung

Bezeichnung des KI-Systems: Privater Chatbot für kreative Texte.

Bewerteter Nutzungskontext: Private Nutzung eines dialogischen KI-Chatbots zur Erstellung, Überarbeitung oder Variation kreativer Texte, Ideen, Formulierungen und Entwürfe.

Bewertungsdatum: Juni 2026.

Bewertende Person / Stelle: Beispielhafte ASEI-10-Bewertung im Rahmen des Fallbeispiels.

Verantwortliche Stelle: Private Nutzerin bzw. privater Nutzer.

B. Bewertungsmodus und berücksichtigte Begrenzungen

Bewertungsmodus: ☒ Bewertung des systemimmanent begrenzten Systemdesigns.

Berücksichtigte Schutz- und Begrenzungsmaßnahmen: Das System wird ausschließlich privat und kreativ genutzt. Es besitzt keinen Zugriff auf externe Werkzeuge, Dateien, E-Mail-Systeme, Datenbanken, Veröffentlichungsplattformen oder produktive Prozesssysteme. Es kann keine Nachrichten versenden, keine Dateien verändern, keine Entscheidungen treffen und keine Handlungen außerhalb des Dialogs ausführen.

Begründung der Scorewirksamkeit: Die fehlenden Werkzeugzugriffe, die rein dialogisch-reaktive Nutzung und die Beschränkung auf eine private Einzelnutzung begrenzen Autonomie, Exposition und Irreversibilität unmittelbar. Eine einfache Plausibilitätsprüfung durch den Nutzer ist sinnvoll, wird aber nicht als zusätzliche scorewirksame Schutzmaßnahme angesetzt.

C. Systemprofil und Wirkungskontext

Nutzergruppen: Eine private Nutzerin bzw. ein privater Nutzer.

Verarbeitete Datenarten: Kreative Textideen, Formulierungen, Entwürfe, Stilvarianten und allgemeine private Eingaben ohne sensiblen Entscheidungs-, Sicherheits- oder Rechtsbezug.

Werkzeugzugriffe oder Schnittstellen: Keine produktiven Werkzeugzugriffe oder externen Schnittstellen.

Mögliche Außenwirkung: Keine automatische Außenwirkung. Eine Außenwirkung entsteht nur dann, wenn der Nutzer erzeugte Inhalte später eigenständig veröffentlicht oder in einen anderen Kontext überträgt.

D. Bewertung der ASEI-Dimensionen

Autonomie A = 2; Normierter Autonomiewert A_n = 1,5. Das System ist ein reaktiver Dialogassistent. Es antwortet auf Eingaben des Nutzers, verfolgt keine eigenständigen Ziele, nutzt keine Werkzeuge und führt keine Handlungen außerhalb des Dialogs aus.

Severität S = 1. Der Einsatzbereich ist privat, kreativ und folgenarm. Es werden keine Rechte, Vermögenswerte, personenbezogenen Entscheidungen, Sicherheitsfragen oder kritischen Schutzgüter berührt.

Exposition E = 1. Die Wirkung bleibt auf eine einzelne Person und eine private Nutzungssituation begrenzt. Eine automatische Weitergabe, Veröffentlichung oder Einbindung in größere Prozesse findet nicht statt.

Irreversibilität I = 1. Fehlerhafte oder unpassende Ergebnisse können sofort verworfen, korrigiert oder neu erzeugt werden. Relevante Restschäden sind im beschriebenen Nutzungskontext nicht zu erwarten.

E. Prüfung korrelativer Beziehungen zwischen den Dimensionen

Es bestehen keine risikoverstärkenden korrelativen Beziehungen zwischen den Dimensionen. Die niedrige Autonomie, geringe Severität, minimale Exposition und leichte Rückholbarkeit stützen sich jeweils auf eigenständige Merkmale des Nutzungskontexts. Eine Mehrfachzählung ist nicht ersichtlich.

F. Scorebildung

$R = 1{,}5 \cdot 1 \cdot 1 \cdot 1 = 1{,}5$ ⇒ ASEI-10-Score: ≈ 1,6

Risikoklasse: RK1 — sehr niedriges Risikopotenzial.

G. Prüfschwellen und Mindestanforderungen

Ausgelöste Prüfschwellen: Höhere Prüfschwellen werden nicht ausgelöst.

Abgeleitete Mindestanforderungen: Es gelten lediglich Basisanforderungen an eine bewusste Nutzung. Dazu gehören eine Plausibilitätsprüfung durch den Nutzer, keine ungeprüfte Übernahme sachlicher Aussagen, bewusster Umgang mit möglichen Halluzinationen oder stilistischen Unstimmigkeiten sowie keine Übertragung in sensible, professionelle, prüfungsbezogene oder öffentlich wirksame Kontexte ohne erneute Bewertung.

H. Neubewertung

Anlass für Neubewertung: Eine Neubewertung ist erforderlich, wenn derselbe Chatbot nicht mehr nur privat-kreativ genutzt wird, sondern für veröffentlichte Inhalte, professionelle Kommunikation, Unterrichts- oder Prüfungsmaterialien, automatisierte Verbreitung, personenbezogene Entscheidungen oder produktive Prozesse eingesetzt wird.

Spätester Zeitpunkt der erneuten Bewertung: Vor jeder wesentlichen Änderung des Nutzungskontexts, insbesondere vor Veröffentlichung, beruflicher Weiterverwendung, Integration in externe Systeme oder Nutzung in sensiblen Entscheidungszusammenhängen.

Dokumentationshinweise: Die Bewertung gilt ausschließlich für den beschriebenen privaten, kreativen und rein dialogischen Nutzungskontext. Sie darf nicht auf andere Einsatzformen desselben Chatbots übertragen werden.

Bewertungsbogen zu Fallbeispiel 2: KI-Assistent für Unterrichts- und Prüfungsvorbereitung

A. Grunddaten der Bewertung

Bezeichnung des KI-Systems: KI-Assistent für Unterrichts- und Prüfungsvorbereitung.

Bewerteter Nutzungskontext: Nutzung eines KI-Assistenten durch einen Dozenten zur Erstellung, Überarbeitung und Strukturierung von Unterrichtsmaterialien, Übungsaufgaben, Musterlösungen, Rechenschemata, Glossaren, Zusammenfassungen und prüfungsnahen Vorbereitungsmaterialien.

Bewertungsdatum: Juni 2026.

Bewertende Person / Stelle: Beispielhafte ASEI-10-Bewertung im Rahmen des Fallbeispiels.

Verantwortliche Stelle: Dozentin bzw. Dozent oder Bildungseinrichtung, die den KI-Assistenten zur Materialvorbereitung einsetzt.

B. Bewertungsmodus und berücksichtigte Begrenzungen

Bewertungsmodus: ☒ Bewertung des systemimmanent begrenzten Systemdesigns

Berücksichtigte Schutz- und Begrenzungsmaßnahmen: Das System wird zur Vorbereitung und Unterstützung eingesetzt, trifft aber keine Prüfungsentscheidungen, bewertet keine Leistungen und entscheidet nicht über Bestehen, Nichtbestehen, Zulassung oder Notengebung. Die Nutzung erfolgt auf konkrete Anweisung des Dozenten. Eine automatische Veröffentlichung oder direkte Ausspielung an Teilnehmende findet nicht statt. Materialien werden vor der Verwendung fachlich geprüft und gegebenenfalls überarbeitet.

Begründung der Scorewirksamkeit: Die Beschränkung auf vorbereitende Assistenz begrenzt die Autonomie auf A3 und verhindert eine Einstufung als eigenständig entscheidendes oder bewertendes Prüfungssystem. Die fachliche Endkontrolle reduziert jedoch nicht automatisch Severität, Exposition oder Irreversibilität auf ein niedriges Niveau, weil fehlerhafte Materialien dennoch Lernprozesse beeinflussen und sich innerhalb einer Kursgruppe vervielfältigen können.

C. Systemprofil und Wirkungskontext

Nutzergruppen: Dozentin bzw. Dozent, gegebenenfalls weitere Lehrende oder Fachverantwortliche.

Verarbeitete Datenarten: Unterrichtsmaterialien, Aufgabenentwürfe, Musterlösungen, Rechenschemata, Glossare, Zusammenfassungen, didaktische Erklärungen und gegebenenfalls hochgeladene Skripte oder vorhandene Aufgabenunterlagen.

Werkzeugzugriffe oder Schnittstellen: Möglichkeit zur Analyse, Strukturierung oder sprachlichen Überarbeitung hochgeladener Unterlagen. Keine eigenständige Veröffentlichung, keine automatische Weitergabe an Teilnehmende und keine unmittelbare Verbindung zu Prüfungsverwaltung, Notensystemen oder Zulassungsentscheidungen.

Mögliche Außenwirkung: Mittelbare Wirkung auf eine abgegrenzte Kurs- oder Lerngruppe, wenn die erzeugten oder überarbeiteten Materialien im Unterricht oder in der Prüfungsvorbereitung verwendet werden.

D. Bewertung der ASEI-Dimensionen

Autonomie A = 3. Normierter Autonomiewert A_n = 2,0. Das System arbeitet auf konkrete Anweisung des Dozenten und erzeugt Vorschläge, Materialien, Strukturierungen oder Überarbeitungen. Es kann vorhandene Unterlagen analysieren und inhaltlich verarbeiten, entscheidet aber nicht selbstständig über Lernziele, Prüfungsinhalte, Bewertungen oder Prüfungsergebnisse.

Severität S = 3. Der Einsatz ist bildungs- und prüfungsvorbereitungsbezogen. Fehlerhafte Materialien können Lernende fehlleiten, Prüfungsstoff verzerren oder fachliche Fehlvorstellungen erzeugen. Da das System jedoch keine Prüfungsleistungen bewertet und keine verbindlichen Bildungsentscheidungen trifft, liegt noch kein hochkritischer Bildungsentscheidungsbereich im Sinne von S4 vor.

Exposition E = 2. Die Wirkung betrifft eine abgegrenzte Kurs- oder Lerngruppe. Die Materialien werden nicht öffentlich verbreitet und nicht organisationsweit automatisiert eingesetzt.

Irreversibilität I = 3. Fehler können grundsätzlich korrigiert werden, sind aber nicht immer folgenlos. Fehlerhafte Musterlösungen, unklare Rechenschemata oder sachlich falsche Erklärungen können Nacharbeit, Vertrauensverlust, Fehlvorbereitung oder didaktische Folgeschäden verursachen.

E. Prüfung korrelativer Beziehungen zwischen den Dimensionen

Es bestehen sachlogische Beziehungen zwischen Severität, Exposition und Irreversibilität. Der prüfungsnahe Bildungskontext erhöht die Folgenrelevanz fehlerhafter Materialien; die Nutzung in einer Kursgruppe erweitert den Wirkungsraum; fehlerhafte Lern- oder Prüfungsvorbereitung ist zwar korrigierbar, kann aber Nachwirkungen entfalten. Die Dimensionen bleiben dennoch unterscheidbar: Severität betrifft den Bildungs- und Prüfungskontext, Exposition die Kursgruppe, Irreversibilität die begrenzte Rückholbarkeit fehlerhafter Lernwirkungen.

F. Scorebildung

$R = 2{,}0 \cdot 3 \cdot 2 \cdot 3 = 36{,}0$ ⇒ ASEI-10-Score: ≈ 6,0

Risikoklasse: RK4 — erhöhtes Risikopotenzial.

G. Prüfschwellen und Mindestanforderungen

Ausgelöste Prüfschwellen: Mehrere mittlere Schwellenkriterien werden berührt. Die Autonomie bleibt zwar im assistiven Bereich, durch Werkzeugnutzung und Materialanalyse ist jedoch besondere Aufmerksamkeit bei Quellen, Berechnungen, Dateiauswertungen und fachlicher Plausibilität erforderlich. Die Severität erreicht S3, weil Lernprozesse und prüfungsnahe Vorbereitung betroffen sein können. Die Irreversibilität erreicht I3, weil Korrekturen möglich sind, aber nicht immer folgenlos bleiben.

Abgeleitete Mindestanforderungen: Erforderlich sind fachliche Endkontrolle, didaktische Prüfung, sachliche Plausibilitätskontrolle, klare Abgrenzung zwischen Übungsmaterial und offiziellen Prüfungsinhalten, dokumentierte Verantwortung des Dozenten sowie transparente Einsatzgrenzen. KI-generierte Musterlösungen, Rechenschemata oder Aufgabenstellungen dürfen nicht ungeprüft in Unterricht oder Prüfungsvorbereitung übernommen werden.

H. Neubewertung

Anlass für Neubewertung: Eine Neubewertung ist erforderlich, wenn das System nicht mehr nur vorbereitend eingesetzt wird, sondern Prüfungsleistungen bewertet, Notenvorschläge erzeugt, Zulassungsentscheidungen vorbereitet, offizielle Prüfungsinhalte erstellt, Materialien automatisch an Teilnehmende ausspielt oder in Lernmanagement- bzw. Prüfungssysteme integriert wird.

Spätester Zeitpunkt der erneuten Bewertung: Vor jeder Einbindung in offizielle Prüfungsbewertung, Leistungsdiagnostik, Zulassungsentscheidung, automatisierte Materialverteilung oder organisationsweite Nutzung.

Dokumentationshinweise: Die Bewertung gilt ausschließlich für die beschriebene vorbereitende Nutzung durch den Dozenten mit fachlicher Endkontrolle. Sie darf nicht auf automatisierte Prüfungs-, Bewertungs- oder Entscheidungssysteme übertragen werden.

Bewertungsbogen zu Fallbeispiel 3: KI-System im Kundenservice

A. Grunddaten der Bewertung

Bezeichnung des KI-Systems: KI-System im Kundenservice.

Bewerteter Nutzungskontext: Einsatz eines KI-gestützten Kundenservice-Systems zur Beantwortung von Kundenanfragen, zur Bereitstellung von Produkt-, Vertrags- oder Serviceinformationen und zur Unterstützung standardisierter Kommunikationsprozesse gegenüber Kundinnen und Kunden.

Bewertungsdatum: Juni 2026.

Bewertende Person / Stelle: Beispielhafte ASEI-10-Bewertung im Rahmen des Fallbeispiels.

Verantwortliche Stelle: Unternehmen bzw. zuständige Kundenservice-, Vertriebs- oder Supporteinheit.

B. Bewertungsmodus und berücksichtigte Begrenzungen

Bewertungsmodus: ☒ Bewertung des systemimmanent begrenzten Systemdesigns

Berücksichtigte Schutz- und Begrenzungsmaßnahmen: Das System beantwortet Kundenanfragen dialogisch und unterstützend, trifft aber keine verbindlichen Rechts-, Vertrags-, Erstattungs-, Sperr- oder Ablehnungsentscheidungen. Es kann keine Kundenkonten sperren, keine Verträge ändern, keine Zahlungen auslösen und keine rechtlich bindenden Zusagen eigenständig erteilen. Für komplexe oder kritische Fälle sind Übergaben an menschliche Servicemitarbeitende vorgesehen.

Begründung der Scorewirksamkeit: Die fehlende eigenständige Handlungsausführung begrenzt die Autonomie auf A3. Die hohe Kundenzahl, der mögliche Umgang mit personenbezogenen oder wirtschaftlich relevanten Informationen und die Wiederholbarkeit fehlerhafter Auskünfte bleiben jedoch scorewirksam für Severität, Exposition und Irreversibilität.

C. Systemprofil und Wirkungskontext

Nutzergruppen: Kundinnen und Kunden, Servicepersonal, gegebenenfalls interne Fachabteilungen.

Verarbeitete Datenarten: Kundenanfragen, Kontaktdaten, Vertrags- oder Bestelldaten, Serviceinformationen, Produktinformationen, Kommunikationshistorien und gegebenenfalls personenbezogene oder wirtschaftlich relevante Kundendaten.

Werkzeugzugriffe oder Schnittstellen: Zugriff auf Wissensdatenbanken, FAQ-Systeme, Produktinformationen, Serviceunterlagen oder Kundendatenansichten. Keine eigenständige Änderung von Vertrags-, Zahlungs-, Sperr- oder Kontodaten im bewerteten Nutzungskontext.

Mögliche Außenwirkung: Direkte Wirkung gegenüber einer großen Zahl von Kundinnen und Kunden. Fehlerhafte oder missverständliche Auskünfte können wiederholt ausgespielt werden und dadurch Servicequalität, Vertrauen, Datenschutz, Kundenbeziehungen und Unternehmensreputation beeinflussen.

D. Bewertung der ASEI-Dimensionen

Autonomie A = 3. Normierter Autonomiewert A_n = 2,0. Das System reagiert auf Kundenanfragen, erzeugt Antworten und kann Informationen aus bereitgestellten Wissens- oder Kundendatenquellen verarbeiten. Es handelt jedoch nicht eigenständig verbindlich und führt keine operativen Maßnahmen wie Vertragsänderungen, Erstattungen, Sperrungen oder Zahlungsaktionen aus.

Severität S = 3. Der Einsatz betrifft personenbezogene und wirtschaftlich relevante Kundenkommunikation. Fehlerhafte Auskünfte können zu Fehlentscheidungen, Vertrauensverlust, Serviceproblemen, Beschwerden oder Datenschutzproblemen führen. Da das System im beschriebenen Kontext keine verbindlichen rechtlichen oder wirtschaftlichen Entscheidungen selbst trifft, ist S3 ausreichend.

Exposition E = 4. Das System wirkt gegenüber einer großen externen Nutzergruppe. Fehlerhafte Antworten können massenhaft wiederholt und vielen Kundinnen und Kunden ausgespielt werden. Die Exposition geht damit deutlich über eine interne Gruppe oder einzelne Organisationseinheit hinaus.

Irreversibilität I = 3. Fehlerhafte Auskünfte können grundsätzlich korrigiert werden, sind aber nicht immer folgenlos. Wiederholte Falschinformationen, Datenschutzprobleme oder missverständliche Kundenkommunikation können Nacharbeit, Beschwerden, Reputationsschäden oder Vertrauensverlust verursachen.

E. Prüfung korrelativer Beziehungen zwischen den Dimensionen

Es bestehen relevante Wechselwirkungen zwischen Exposition und Irreversibilität sowie zwischen Severität und Exposition. Die breite Kundenerreichbarkeit kann dazu führen, dass fehlerhafte Auskünfte wiederholt und gegenüber vielen Personen verbreitet werden. Dadurch werden Korrekturen schwieriger und mögliche Reputations-, Datenschutz- oder Servicefolgen verstärkt. Die Dimensionen sind dennoch eigenständig begründet: Severität betrifft die personenbezogene und wirtschaftliche Relevanz der Auskünfte, Exposition die große externe Nutzerzahl, Irreversibilität die begrenzte Rückholbarkeit wiederholter Fehlkommunikation.

F. Scorebildung

$R = 2{,}0 \cdot 3 \cdot 4 \cdot 3 = 72{,}0$ ⇒ ASEI-10-Score: ungerundet ≈ 6,98; gerundet 7,0

Risikoklasse: Formal RK4 — erhöhtes Risikopotenzial; grenznah zur Risikoklasse RK5.

G. Prüfschwellen und Mindestanforderungen

Ausgelöste Prüfschwellen: Die breite externe Exposition erreicht E4. Zusätzlich sind personenbezogene oder wirtschaftlich relevante Informationen betroffen. Die Irreversibilität erreicht I3, weil fehlerhafte Auskünfte korrigierbar sind, aber wiederholte Fehlkommunikation nicht immer folgenlos bleibt.

Abgeleitete Mindestanforderungen: Erforderlich sind fachliche Pflege der Wissensbasis, klare Eskalationsregeln, menschliche Übernahme bei kritischen oder unklaren Fällen, Protokollierung relevanter Dialoge, Monitoring häufiger Fehlantworten, Beschwerde- und Korrekturprozesse sowie klare Begrenzung rechtlich oder wirtschaftlich verbindlicher Aussagen. Das System sollte nicht ohne zusätzliche Prüfung für Vertragsänderungen, Erstattungen, Sperrungen, Forderungsentscheidungen oder rechtlich relevante Zusagen eingesetzt werden.

H. Neubewertung

Anlass für Neubewertung: Eine Neubewertung ist erforderlich, wenn das System eigenständig Erstattungen genehmigt, Verträge ändert, Kundenkonten sperrt, Forderungen ablehnt, rechtlich relevante Zusagen erteilt, automatisiert Beschwerden entscheidet oder stärker in operative Kundenprozesse integriert wird.

Spätester Zeitpunkt der erneuten Bewertung: Vor jeder Erweiterung um verbindliche Handlungsmöglichkeiten, vor Anbindung an produktive Vertrags-, Zahlungs- oder Kontosysteme sowie vor organisationsweiter oder internationaler Ausweitung des Einsatzes.

Dokumentationshinweise: Die Bewertung gilt ausschließlich für ein Kundenservice-System mit dialogischer Antwortfunktion und ohne eigenständige verbindliche Handlungsausführung. Die gerundete Darstellung von 7,0 darf nicht isoliert interpretiert werden; maßgeblich bleibt der ungerundete Score in Verbindung mit den Einzeldimensionen und Schwellenkriterien.

Bewertungsbogen zu Fallbeispiel 4: KI-Agent mit Toolzugriff in einem Unternehmen

A. Grunddaten der Bewertung

Bezeichnung des KI-Systems: KI-Agent mit Toolzugriff in einem Unternehmen.

Bewerteter Nutzungskontext: Einsatz eines KI-Agenten zur Unterstützung interner Arbeitsprozesse, etwa durch Strukturierung von Aufgaben, Weiterleitung von Informationen, Erstellung interner Nachrichten, Bearbeitung einfacher Vorgänge, Dokumentenablage oder Anstoßen definierter Prozessschritte.

Bewertungsdatum: Juni 2026.

Bewertende Person / Stelle: Beispielhafte ASEI-10-Bewertung im Rahmen des Fallbeispiels.

Verantwortliche Stelle: Unternehmen bzw. zuständige Fachabteilung, IT-, Prozess- oder Governance-Verantwortliche.

B. Bewertungsmodus und berücksichtigte Begrenzungen

Bewertungsmodus: ☒ Bewertung des systemimmanent begrenzten Systemdesigns

Berücksichtigte Schutz- und Begrenzungsmaßnahmen: Der Agent arbeitet innerhalb eines definierten Aufgaben- und Berechtigungsrahmens. Er kann interne Informationen verarbeiten, Aufgaben anstoßen, Nachrichten vorbereiten oder versenden und einfache Prozessschritte unterstützen. Er darf jedoch keine Zahlungen auslösen, keine Verträge abschließen oder ändern, keine Personalmaßnahmen durchführen, keine Kundenentscheidungen treffen und keine sicherheitsrelevanten Aktionen eigenständig ausführen. Kritische Vorgänge erfordern menschliche Freigabe oder werden aus dem Funktionsumfang ausgeschlossen.

Begründung der Scorewirksamkeit: Die Möglichkeit zur eigenständigen operativen Handlungsausführung begründet die Einstufung in A5. Die Begrenzung auf definierte interne Prozesse und der Ausschluss besonders kritischer Handlungen verhindern jedoch eine Einstufung in höhere Autonomiestufen. Da das System im beschriebenen Kontext nicht dauerhaft ereignisgesteuert aktiv bleibt und keine eigenständige Langzeitüberwachung von Systemzuständen vornimmt, ist A6 nicht erreicht.

C. Systemprofil und Wirkungskontext

Nutzergruppen: Mitarbeitende einer Fachabteilung, Prozessverantwortliche, Teamleitungen, gegebenenfalls interne Support- oder Verwaltungsstellen.

Verarbeitete Datenarten: Interne Arbeitsdokumente, Aufgabeninformationen, Prozessdaten, E-Mail- oder Nachrichtentexte, Termin- und Statusinformationen, interne Fachinformationen und gegebenenfalls begrenzt personenbezogene Beschäftigtendaten im Rahmen gewöhnlicher Arbeitsprozesse.

Werkzeugzugriffe oder Schnittstellen: Zugriff auf definierte interne Werkzeuge wie Dokumentenablagen, Aufgabenmanagementsysteme, Kalender, interne Kommunikationssysteme oder einfache Prozessschnittstellen. Kein eigenständiger Zugriff auf Zahlungs-, Vertrags-, Personal-, Kundenentscheidungs- oder sicherheitskritische Systeme im bewerteten Nutzungskontext.

Mögliche Außenwirkung: Primär interne Wirkung innerhalb einer Fachabteilung oder eines abgegrenzten organisatorischen Bereichs. Fehlerhafte Handlungen können Arbeitsabläufe, Informationsflüsse, Fristen, Zuständigkeiten oder Dokumentenstände beeinflussen.

D. Bewertung der ASEI-Dimensionen

Autonomie A = 5; Normierter Autonomiewert A_n = 3,0. Das System geht über reine Assistenz hinaus und kann innerhalb eines definierten Rahmens operative Handlungen ausführen. Es kann Aufgaben anstoßen, Informationen weiterleiten, interne Nachrichten erzeugen oder Prozessschritte beeinflussen. Es bleibt jedoch auf abgegrenzte Funktionen beschränkt und ist nicht dauerhaft ereignisgesteuert aktiv.

Severität S = 3. Der Einsatz betrifft interne Unternehmensprozesse mit möglicher organisatorischer und wirtschaftlicher Relevanz. Fehler können Arbeitsabläufe stören, Fristen verändern, Informationen falsch verteilen oder zusätzlichen Korrekturaufwand verursachen. Da keine Zahlungen, Verträge, Personalmaßnahmen, Kundenentscheidungen oder sicherheitskritischen Aktionen eigenständig ausgeführt werden, ist S3 ausreichend.

Exposition E = 3. Die Wirkung betrifft eine Organisationseinheit oder mehrere interne Prozessbeteiligte. Das System ist nicht nur auf eine Einzelperson beschränkt, entfaltet aber auch keine öffentliche, organisationsweite oder systemische Reichweite.

Irreversibilität I = 3. Fehlerhafte Prozesshandlungen können grundsätzlich erkannt und korrigiert werden, verursachen aber unter Umständen Nacharbeit, Verzögerungen, Fehlkommunikation oder organisatorische Folgewirkungen. Die Folgen sind nicht vollständig trivial, bleiben im beschriebenen Nutzungskontext aber grundsätzlich rückholbar.

E. Prüfung korrelativer Beziehungen zwischen den Dimensionen

Es bestehen Wechselwirkungen zwischen Autonomie, Exposition und Irreversibilität. Die begrenzte agentische Handlungsausführung kann interne Prozesse beeinflussen und Fehler in Prozessketten fortsetzen. Dadurch wird die Rückholung erschwert, obwohl die Folgen grundsätzlich korrigierbar bleiben. Die Dimensionen bleiben jedoch getrennt begründbar: Autonomie betrifft die operative Handlungstiefe, Exposition die betroffenen internen Prozesse und Irreversibilität den Korrekturaufwand bei bereits ausgelösten Prozesswirkungen.

F. Scorebildung

$R = 3{,}0 \cdot 3 \cdot 3 \cdot 3 = 81{,}0$ ⇒ ASEI-10-Score: ≈ 7,1

Risikoklasse: RK5 — hohes Risikopotenzial.

G. Prüfschwellen und Mindestanforderungen

Ausgelöste Prüfschwellen: Die Autonomiedimension erreicht A5, weil das System innerhalb definierter Grenzen operative Handlungen ausführen kann. Damit wird eine agentische Prüfschwelle berührt. Zusätzlich sind mittlere Anforderungen aus S3, E3 und I3 relevant, weil Fehler nicht nur einen einzelnen Text betreffen, sondern interne Prozessketten beeinflussen können.

Abgeleitete Mindestanforderungen: Erforderlich sind ein klares Rollen- und Rechtekonzept, Begrenzung der Toolzugriffe, definierte Freigabepunkte für kritische Vorgänge, Protokollierung agentischer Handlungen, Monitoring wiederkehrender Fehler, Eskalationsregeln, Rückhol- oder Korrekturmechanismen sowie eine klare Verantwortungszuordnung. Der Agent sollte nicht ohne erneute Bewertung für Zahlungen, Verträge, Personalmaßnahmen, Kundenentscheidungen oder sicherheitsrelevante Aktionen eingesetzt werden.

H. Neubewertung

Anlass für Neubewertung: Eine Neubewertung ist erforderlich, wenn der Agent dauerhaft ereignisgesteuert aktiv bleibt, eigenständig auf neue Systemzustände reagiert, Prozesse über längere Zeit ohne erneute Nutzerinteraktion überwacht, organisationsweit ausgerollt wird oder zusätzliche Berechtigungen für Zahlungen, Verträge, Personalmaßnahmen, Kundenentscheidungen oder sicherheitsrelevante Aktionen erhält.

Spätester Zeitpunkt der erneuten Bewertung: Vor jeder Erweiterung von Toolzugriffen, vor organisationsweiter Ausrollung, vor Einbindung in produktive Kernprozesse mit Außenwirkung und vor jeder Freigabe eigenständiger kritischer Handlungsmöglichkeiten.

Dokumentationshinweise: Die Bewertung gilt ausschließlich für einen begrenzten internen Unternehmensagenten mit definierten Toolzugriffen und ohne eigenständige kritische Entscheidungs- oder Ausführungsbefugnisse. Wird die Handlungstiefe erweitert, sind insbesondere A6, S4, E4 und I4 erneut zu prüfen.

Bewertungsbogen zu Fallbeispiel 5: KI-System zur Personalvorauswahl

A. Grunddaten der Bewertung

Bezeichnung des KI-Systems: KI-System zur Personalvorauswahl.

Bewerteter Nutzungskontext: Einsatz eines KI-Systems zur Analyse, Sortierung, Priorisierung oder Vorauswahl von Bewerbungen innerhalb eines Recruitingprozesses. Das System unterstützt Personalverantwortliche bei der Sichtung von Bewerbungsunterlagen, trifft aber im bewerteten Nutzungskontext keine endgültige Einstellungs- oder Ablehnungsentscheidung.

Bewertungsdatum: Juni 2026.

Bewertende Person / Stelle: Beispielhafte ASEI-10-Bewertung im Rahmen des Fallbeispiels.

Verantwortliche Stelle: Unternehmen bzw. zuständige Personalabteilung, Recruiting-Verantwortliche oder HR-Governance-Stelle.

B. Bewertungsmodus und berücksichtigte Begrenzungen

Bewertungsmodus: ☒ Bewertung des systemimmanent begrenzten Systemdesigns

Berücksichtigte Schutz- und Begrenzungsmaßnahmen: Das System trifft keine endgültige Personalentscheidung. Die finale Auswahlentscheidung verbleibt formal bei menschlichen Personalverantwortlichen. Das System dient der Vorstrukturierung, Priorisierung oder Empfehlung innerhalb des Bewerbungsprozesses. Automatische Absagen ohne menschliche Prüfung, verbindliche Ausschlüsse oder eigenständige Einstellungsentscheidungen sind im bewerteten Nutzungskontext nicht vorgesehen.

Begründung der Scorewirksamkeit: Die menschliche Letztentscheidung begrenzt die Autonomie und verhindert eine Einstufung als vollständig autonomes Entscheidungssystem. Gleichwohl bleibt die Vorstrukturierung des Bewerbungsprozesses risikorelevant, weil die KI-Auswahl die Wahrnehmung, Priorisierung und Entscheidungspraxis der Personalverantwortlichen faktisch prägen kann. Die Schutzwirkung der menschlichen Letztentscheidung ist daher begrenzt und reduziert Severität, Exposition und Irreversibilität nicht wesentlich.

C. Systemprofil und Wirkungskontext

Nutzergruppen: Personalverantwortliche, Recruiter, Fachvorgesetzte, gegebenenfalls externe Dienstleister im Bewerbungsprozess.

Verarbeitete Datenarten: Bewerbungsunterlagen, Lebensläufe, Zeugnisse, Qualifikationsprofile, berufliche Stationen, Anschreiben, Kommunikationsdaten und gegebenenfalls weitere personenbezogene oder beschäftigungsbezogene Informationen.

Werkzeugzugriffe oder Schnittstellen: Zugriff auf Bewerbungsmanagementsysteme, Bewerbungsdatenbanken oder HR-Workflows zur Analyse, Sortierung und Priorisierung von Bewerbungen. Keine eigenständige Versendung automatischer Absagen und keine abschließende Entscheidungsbefugnis im bewerteten Nutzungskontext.

Mögliche Außenwirkung: Wirkung auf Bewerberinnen und Bewerber sowie auf interne Auswahlprozesse. Die Systemausgaben können beeinflussen, welche Bewerbungen bevorzugt geprüft, zurückgestellt oder faktisch aus dem weiteren Auswahlprozess verdrängt werden.

D. Bewertung der ASEI-Dimensionen

Autonomie A = 4; Normierter Autonomiewert A_n = 2,5. Das System analysiert Bewerbungsunterlagen und strukturiert Auswahlentscheidungen vor. Es kann Empfehlungen, Rankings oder Priorisierungen erzeugen, trifft aber im beschriebenen Nutzungskontext keine endgültigen Einstellungs- oder Ablehnungsentscheidungen und versendet keine automatischen Absagen ohne menschliche Prüfung.

Severität S = 4. Der Einsatzbereich ist hochsensibel, weil Beschäftigung, berufliche Chancen, Gleichbehandlung und Zugang zu Erwerbsmöglichkeiten betroffen sind. Fehlerhafte oder verzerrte Vorauswahl kann erhebliche Nachteile für Bewerberinnen und Bewerber verursachen.

Exposition E = 3. Die Wirkung betrifft einen organisationalen Bewerbungsprozess und regelmäßig mehrere Bewerberinnen und Bewerber. Das System ist nicht nur auf eine einzelne Person beschränkt, entfaltet aber im beschriebenen Kontext keine öffentliche oder systemische Reichweite.

Irreversibilität I = 4. Benachteiligungen im Bewerbungsprozess sind nur begrenzt rückholbar. Selbst wenn eine fehlerhafte Vorauswahl später erkannt wird, können berufliche Chancen, Zeitfenster, Vertrauen und faire Teilhabe bereits beeinträchtigt sein. Eine nachträgliche Korrektur stellt die ursprüngliche Chancengleichheit nicht vollständig wieder her.

E. Prüfung korrelativer Beziehungen zwischen den Dimensionen

Es bestehen starke sachlogische Beziehungen zwischen Severität und Irreversibilität. Der Beschäftigungskontext ist hochsensibel, weil berufliche Chancen, Gleichbehandlung und Zugang zu Erwerbsmöglichkeiten betroffen sind. Mögliche Benachteiligungen sind nur begrenzt rückholbar, selbst wenn sie später erkannt werden. Zusätzlich wirkt die Autonomie auf die Severität ein, weil die KI den Auswahlprozess vorstrukturiert. Eine Mehrfachzählung liegt nicht vor, weil Severität den betroffenen Schutzbereich beschreibt, Irreversibilität die begrenzte Korrigierbarkeit und Autonomie die faktische Vorprägung des Entscheidungsprozesses.

F. Scorebildung

$R = 2{,}5 \cdot 4 \cdot 3 \cdot 4 = 120{,}0$ ⇒ ASEI-10-Score: ≈ 7,7

Risikoklasse: RK5 — hohes Risikopotenzial.

G. Prüfschwellen und Mindestanforderungen

Ausgelöste Prüfschwellen: Die Severität erreicht S4, weil ein besonders schutzwürdiger Beschäftigungskontext betroffen ist. Die Irreversibilität erreicht I4, weil mögliche Benachteiligungen nur begrenzt rückholbar sind. Zusätzlich ist die Autonomiedimension risikorelevant, weil das System Auswahlprozesse agentisch vorstrukturiert, auch wenn die Letztentscheidung formal menschlich bleibt.

Abgeleitete Mindestanforderungen: Erforderlich sind transparente Einsatzgrenzen, menschliche und fachlich verantwortete Einzelfallprüfung, nachvollziehbare Auswahlkriterien, Dokumentation der Systemnutzung, regelmäßige Prüfung auf Diskriminierungs- oder Verzerrungseffekte, Beschwerde- und Korrekturmöglichkeiten, Datenschutzprüfung, klare Verantwortungszuordnung sowie eine besondere Kontrolle der faktischen Wirkung des Systems auf Auswahlentscheidungen. Automatische Absagen oder verbindliche Ausschlüsse ohne menschliche Prüfung sollten ohne erneute Bewertung ausgeschlossen bleiben.

H. Neubewertung

Anlass für Neubewertung: Eine Neubewertung ist erforderlich, wenn das System automatische Absagen versendet, Bewerberinnen und Bewerber ohne menschliche Prüfung aussortiert, verbindliche Rankings erzeugt, organisationsweit in allen Recruitingprozessen eingesetzt wird, zusätzliche sensible Daten verarbeitet oder stärker in HR-Entscheidungsprozesse integriert wird.

Spätester Zeitpunkt der erneuten Bewertung: Vor jeder Erweiterung zu automatisierten Absagen, vor organisationsweiter Ausrollung, vor Einbindung in verbindliche Personalentscheidungen und vor Verarbeitung zusätzlicher sensibler oder besonders schutzwürdiger Daten.

Dokumentationshinweise: Die Bewertung gilt ausschließlich für eine KI-gestützte Personalvorauswahl mit menschlicher Letztentscheidung. Entscheidend ist nicht nur, wer formal entscheidet, sondern wie stark die KI die Wahrnehmung, Priorisierung und Auswahl der Personalverantwortlichen faktisch beeinflusst.

Bewertungsbogen zu Fallbeispiel 6: KI-System in kritischer Infrastruktur und Cybersecurity

A. Grunddaten der Bewertung

Bezeichnung des KI-Systems: KI-System in kritischer Infrastruktur und Cybersecurity.

Bewerteter Nutzungskontext: Einsatz eines KI-gestützten Systems zur Überwachung, Analyse und ereignisgesteuerten Reaktion auf sicherheitsrelevante Vorgänge in kritischer Infrastruktur, etwa zur Erkennung von Angriffsmustern, Priorisierung von Sicherheitsereignissen, Auslösung definierter Schutzmaßnahmen oder Unterstützung von Incident-Response-Prozessen.

Bewertungsdatum: Juni 2026.

Bewertende Person / Stelle: Beispielhafte ASEI-10-Bewertung im Rahmen des Fallbeispiels.

Verantwortliche Stelle: Betreiber kritischer Infrastruktur, IT-Sicherheitsverantwortliche, Security Operations Center, CISO-Funktion oder zuständige Governance- und Risikostelle.

B. Bewertungsmodus und berücksichtigte Begrenzungen

Bewertungsmodus: ☒ Bewertung des systemimmanent begrenzten Systemdesigns

Berücksichtigte Schutz- und Begrenzungsmaßnahmen: Das System arbeitet dauerhaft ereignisgesteuert und kann sicherheitsrelevante Vorgänge erkennen, priorisieren und innerhalb definierter Grenzen Schutz- oder Reaktionsmaßnahmen anstoßen. Es ist jedoch im bewerteten Nutzungskontext nicht als unkontrollierbares, selbstoptimierendes oder vollständig autonomes System angelegt. Kritische Eingriffe, Eskalationen und Änderungen zentraler Schutzarchitekturen unterliegen menschlicher Aufsicht, Protokollierung und definierten Notfallverfahren.

Begründung der Scorewirksamkeit: Die dauerhafte ereignisgesteuerte Aktivität begründet A6. Die vorhandenen Begrenzungen verhindern im beschriebenen Systemzustand eine Einstufung in A7, A8 oder A9. Sie reduzieren jedoch nicht die sehr hohe Severität, systemische Exposition und kaum reversible Folgen möglicher Fehlreaktionen in kritischer Infrastruktur. Deshalb bleiben S5, E5 und I5 vollständig scorewirksam.

C. Systemprofil und Wirkungskontext

Nutzergruppen: IT-Sicherheitsverantwortliche, SOC-Analystinnen und -Analysten, Systemadministration, Krisen- oder Incident-Response-Teams, gegebenenfalls Management- oder Leitstellenfunktionen.

Verarbeitete Datenarten: Sicherheitsereignisse, Logdaten, Netzwerkdaten, Systemzustände, Angriffsmuster, Zugriffsinformationen, technische Konfigurationsdaten, Alarmmeldungen, Schwachstelleninformationen und gegebenenfalls besonders schutzwürdige Betriebs- oder Infrastrukturdaten.

Werkzeugzugriffe oder Schnittstellen: Zugriff auf Monitoring-Systeme, SIEM- oder SOAR-Plattformen, Netzwerk- und Sicherheitssysteme, Alarmierungs- oder Eskalationskanäle sowie definierte technische Schutzmaßnahmen. Im bewerteten Nutzungskontext keine unkontrollierte eigenständige Veränderung der gesamten Sicherheitsarchitektur und keine unbegrenzte Selbstoptimierung.

Mögliche Außenwirkung: Potenziell systemische Wirkung auf kritische Dienste, technische Infrastruktur, Sicherheitslage, Betriebsfähigkeit, Versorgungssicherheit und nachgelagerte Organisationen oder Nutzergruppen. Fehlerhafte Reaktionen können weit über einzelne interne Prozesse hinausreichen.

D. Bewertung der ASEI-Dimensionen

Autonomie A = 6; Normierter Autonomiewert A_n = 3,5. Das System arbeitet dauerhaft ereignisgesteuert und kann auf sicherheitsrelevante Systemzustände reagieren. Es bleibt über längere Zeit aktiv, priorisiert Ereignisse und kann definierte Schutz- oder Reaktionsmaßnahmen anstoßen. Es ist jedoch nicht als verteiltes, selbstoptimierendes oder nicht zuverlässig kontrollierbares System im Sinne der Stufen A7 bis A9 einzustufen.

Severität S = 5. Der Einsatzbereich betrifft kritische Infrastruktur und Cybersecurity. Fehlerhafte Bewertungen, Fehlreaktionen oder unterlassene Reaktionen können erhebliche Auswirkungen auf Sicherheit, Betriebsfähigkeit, Versorgung, Schutz kritischer Systeme oder gesellschaftlich bedeutsame Funktionen haben.

Exposition E = 5. Die Wirkung kann über einzelne interne Prozesse oder Organisationseinheiten hinausreichen. Betroffen sein können mehrere Systeme, Infrastrukturbereiche, externe Nutzergruppen, abhängige Organisationen oder öffentliche Versorgungs- und Sicherheitsfunktionen.

Irreversibilität I = 5. Mögliche Fehlreaktionen oder unterlassene Reaktionen können schwer rückholbare Folgen verursachen. Dazu gehören Sicherheitsvorfälle, Betriebsunterbrechungen, Folgeschäden in vernetzten Systemen, Vertrauensverlust, Datenabfluss oder kritische Störungen, die nachträglich nur begrenzt korrigiert oder kompensiert werden können.

E. Prüfung korrelativer Beziehungen zwischen den Dimensionen

Es bestehen sehr starke Wechselwirkungen zwischen Severität, Exposition und Irreversibilität sowie zusätzlich mit der persistenten agentischen Aktivität des Systems. Kritische Infrastruktur weist hohe Folgenrelevanz auf; systemische Exposition kann Wirkungen über einzelne Prozesse hinaus ausbreiten; Fehlreaktionen oder unterlassene Reaktionen können kaum vollständig rückholbare Folgen erzeugen. Die hohen Werte in S, E und I sind dennoch nicht bloß Wiederholungen derselben Eigenschaft, sondern beziehen sich auf unterschiedliche Risikofragen: betroffene Schutzgüter, Wirkungsbreite und Rückholbarkeit.

F. Scorebildung

$R = 3{,}5 \cdot 5 \cdot 5 \cdot 5 = 437{,}5$ ⇒ ASEI-10-Score: ≈ 9,5

Risikoklasse: RK6 — sehr hohes Risikopotenzial.

G. Prüfschwellen und Mindestanforderungen

Ausgelöste Prüfschwellen: Die Autonomiedimension erreicht A6, weil das System dauerhaft ereignisgesteuert aktiv ist. Die Severität erreicht S5, weil kritische Infrastruktur und Cybersecurity betroffen sind. Die Exposition erreicht E5, weil systemische Wirkungen möglich sind. Die Irreversibilität erreicht I5, weil mögliche Folgen kaum vollständig rückholbar sein können.

Abgeleitete Mindestanforderungen: Erforderlich sind eine vertiefte technische, organisatorische und sicherheitsbezogene Prüfung, strenge Begrenzung von Tool- und Systemrechten, menschliche Eingriffsmöglichkeiten, Notfall- und Abschaltverfahren, Protokollierung aller relevanten Systemhandlungen, kontinuierliches Monitoring, Red Teaming, Incident-Response-Prozesse, Rollback- und Wiederherstellungsmechanismen, unabhängige Überprüfung, klare Verantwortungszuordnung sowie regelmäßige Neubewertung. Ein produktiver Einsatz ist nur vertretbar, wenn Kontrollierbarkeit, Eingrenzbarkeit, Nachvollziehbarkeit und Notfallfähigkeit belastbar nachgewiesen sind.

H. Neubewertung

Anlass für Neubewertung: Eine Neubewertung ist erforderlich, wenn das System mehrere Agenten koordiniert, eigene Abwehrstrategien optimiert, Schutzmaßnahmen selbstständig verändert, sich über mehrere Infrastrukturbereiche hinweg verteilt, zusätzliche operative Rechte erhält oder nicht mehr zuverlässig begrenzbar, überwachbar, abschaltbar oder rückholbar ist.

Spätester Zeitpunkt der erneuten Bewertung: Vor jedem produktiven Einsatz, vor jeder Erweiterung der Systemrechte, vor jeder Anbindung an weitere kritische Systeme, vor jeder Änderung der Reaktions- oder Eskalationslogik sowie nach relevanten Sicherheitsvorfällen, Fehlreaktionen oder festgestellten Kontrollproblemen.

Dokumentationshinweise: Die Bewertung gilt ausschließlich für ein dauerhaft ereignisgesteuertes, aber weiterhin begrenztes und kontrollierbares KI-System in kritischer Infrastruktur und Cybersecurity. Sollte die Kontrollierbarkeit eingeschränkt sein oder eine selbstoptimierende, verteilte oder nicht zuverlässig abschaltbare Systemarchitektur vorliegen, wären insbesondere A7, A8 oder A9 zu prüfen. In solchen Fällen wäre vor einem produktiven Einsatz eine grundlegende Vertretbarkeitsprüfung erforderlich.

Über den Autor

Thomas H. Anhut ist Diplom-Kaufmann und freiberuflicher Dozent für betriebswirtschaftliche Themen. Er studierte Betriebswirtschaftslehre an der Freien Universität Berlin und war langjährig im Controlling, Projektmanagement und Marketing Berliner und internationaler Industrieunternehmen tätig. Diese Tätigkeit verband betriebswirtschaftliche Analyse mit der praktischen Bewertung von Entscheidungs- und Kontrollprozessen in Organisationen – mit der Frage also, anhand welcher Kriterien sich komplexe Sachverhalte strukturiert und nachvollziehbar einordnen lassen.

Seit mehr als zwanzig Jahren ist er in der beruflichen Aus- und Weiterbildung tätig und als Prüfer in der kaufmännischen Ausbildung bei der IHK Berlin bestellt; er verfügt über die berufs- und arbeitspädagogische Eignung nach AEVO. Aus der Lehr- und Prüfertätigkeit bringt er besondere Erfahrung in der didaktischen Strukturierung komplexer Sachverhalte, der Entwicklung prüfbarer Bewertungsmaßstäbe und der nachvollziehbaren Begründung fachlicher Entscheidungen mit – Kompetenzen, die unmittelbar in die Konzeption des ASEI-10-Modells eingeflossen sind.

Sein Interesse an künstlicher Intelligenz ergibt sich aus der Verbindung von betriebswirtschaftlicher Steuerung, Bildungspraxis und technologischer Entwicklung. Im Mittelpunkt steht die Frage, wie KI-Systeme in Organisationen kontrollierbar und verantwortbar eingesetzt werden können – künstliche Intelligenz also nicht allein als technologische Innovation, sondern als Management-, Governance- und Bewertungsaufgabe. Das ASEI-10-Modell überträgt dieses betriebswirtschaftliche Strukturierungs- und Bewertungsdenken auf die Einordnung des Risikopotenzials konkreter KI-Systeme im Anwendungskontext.