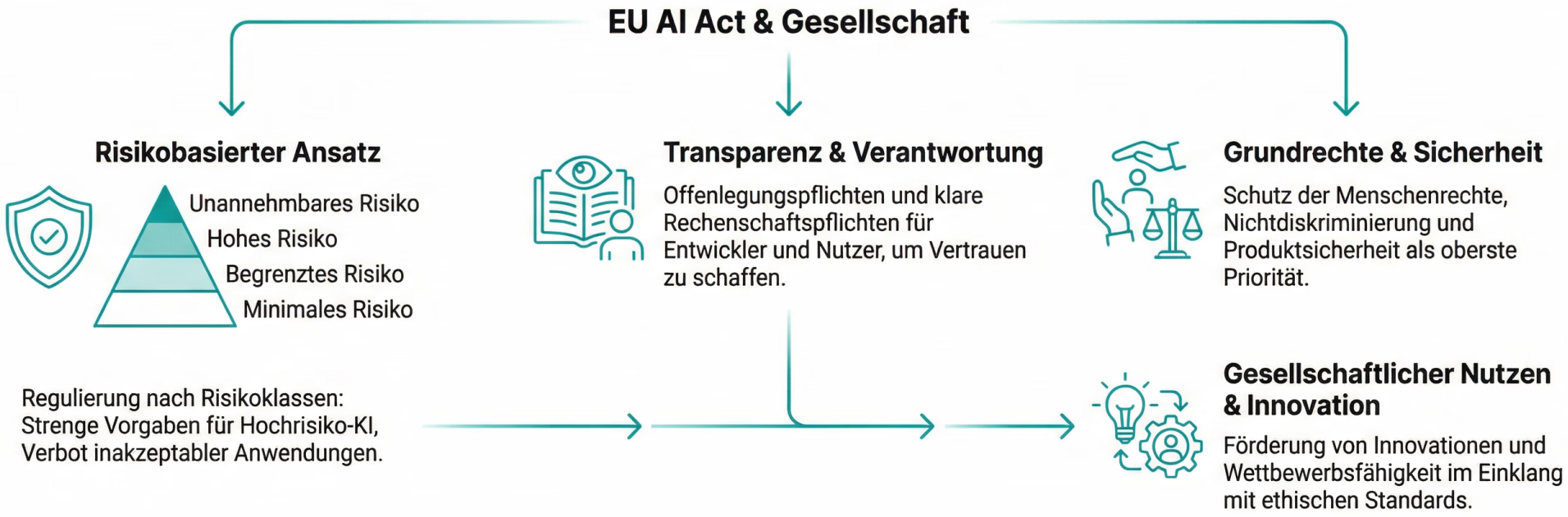
Was ist der EU AI Act?

Kapitel 6.1

Der EU AI Act ist der weltweit erste umfassende Rechtsrahmen für Künstliche Intelligenz. Er zielt darauf ab, die Entwicklung und Nutzung von vertrauenswürdiger, ethischer und sicherer KI in der Europäischen Union zu fördern und gleichzeitig Grundrechte, Demokratie und Rechtsstaatlichkeit zu schützen, mit einem besonderen Fokus auf gesellschaftliche Auswirkungen.



webconsulting

Kapitel 6.1: Was ist der EU AI Act?

webconsulting.at/blog/ki-kompendium-2025#61-was-ist-der-eu-ai-act

Infografik: KI-generiert

Was ist C2PA?

Kapitel 6.2 - Ethik und Gesellschaft

Standard für digitale Echtheitsnachweise.

C2PA schafft eine transparente Provenienz für digitale Inhalte, um deren Ursprung und Integrität nachzuweisen.

Der C2PA-Workflow & Gesellschaftlicher Wert

Nachvollziehbarkeit schaffen: Etabliert eine unveränderliche digitale Kette von Informationen über die Entstehung von Inhalten.

Verifizierung ermöglichen: Ermöglicht es Nutzern, die Echtheit von Medien (Bilder, Videos, Dokumente) unabhängig zu überprüfen.

Vertrauen fördern: Stärkt das Vertrauen in digitale Medien und hilft, die Verbreitung von Fehlinformationen zu bekämpfen.

C2PA dient als technisches Fundament für ethischen Journalismus und informierte Öffentlichkeit.

webconsulting

Kapitel 6.2: Was ist C2PA?

webconsulting.at/blog/ki-kompendium-2025#62-was-ist-c2pa

Infografik: KI-generiert

Was ist "Alignment"?

Kapitel 6.4: Ethische Grundbegriffe

Alignment bezeichnet im Kontext von Ethik und Gesellschaft die **Übereinstimmung der Ziele, Werte** oder Verhaltensweisen eines Systems (z.B. KI, Organisation) mit den menschlichen und gesellschaftlichen Normen, Absichten und ethischen Prinzipien.

Werte-Übereinstimmung (Value Congruence)

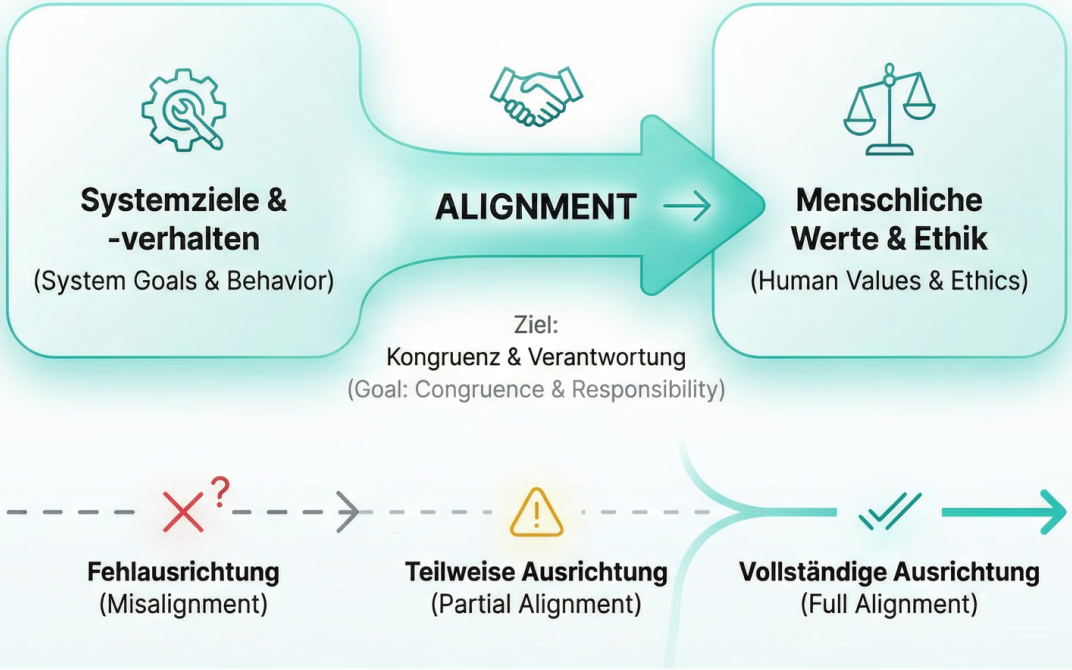
Sicherstellung, dass die fundamentalen Werte des Systems die der betroffenen Menschen und Gemeinschaften widerspiegeln.

Vermeidung von Schäden (Harm Avoidance)

Verhinderung negativer Auswirkungen und unbeabsichtigter Konsequenzen, die durch Fehlausrichtung entstehen können.

Förderung des Gemeinwohls (Beneficence for Society)

Ausrichtung des Systems auf positive Beiträge zur gesellschaftlichen Entwicklung und zum allgemeinen Wohlbefinden.



webconsulting

Kapitel 6.4: Was ist "Alignment"?

webconsulting.at/blog/ki-kompendium-2025#64-was-ist-alignment

Infografik: KI-generiert

Was ist "Constitutional AI"?

Kernkonzept: Constitutional AI ist ein Ansatz, bei dem eine KI durch explizite, menschengeschriebene Prinzipien – eine Art „Verfassung“ – gelenkt wird, um ethische Richtlinien einzuhalten und schädliche Verhaltensweisen zu vermeiden, anstatt nur durch Training auf Daten.

Verankerung von Werten & Ethik

Definiert Grundsätze wie Fairness, Sicherheit und Nutzen als unantastbare Leitlinien.

Selbstreflexives Lernen

Das System nutzt die Verfassung, um sein eigenes Verhalten kontinuierlich zu bewerten und zu verbessern.

Gesellschaftliche Verantwortung

Zielt darauf ab, positive gesellschaftliche Auswirkungen zu fördern und Risiken zu minimieren.

webconsulting

Kapitel 6.5: Was ist "Constitutional AI"?

webconsulting.at/blog/ki-kompendium-2025#65-was-ist-constitutional-ai

Infografik: KI-generiert

Was ist "Red Teaming"?

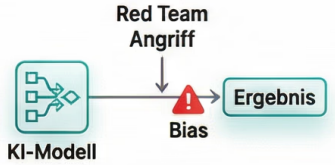
Im Kontext von Ethik und Gesellschaft

Kernkonzept: Proaktive Sicherheits- und Ethikprüfung

Red Teaming ist die Praxis der systematischen, kritischen und ethischen Überprüfung von KI-Systemen und gesellschaftlichen Strukturen durch **simulierten Angriff** und **Stress-Tests**, um **Schwachstellen, Vorurteile** und **potenzielle negative Auswirkungen** zu identifizieren, bevor sie real werden.

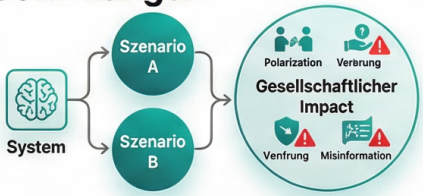
1. Ethische Schwachstellen-Analyse

Gezielte Suche nach ethischen blinden Flecken, algorithmischer Voreingenommenheit (Bias) und unvorhergesehenen diskriminierenden Ergebnissen in Modellen und Entscheidungsprozessen.



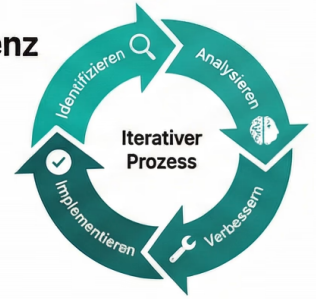
2. Simulation gesellschaftlicher Auswirkungen

Testen von Systemen unter realitätsnahen, extremen und unerwarteten gesellschaftlichen Szenarien, um die **Widerstandsfähigkeit** gegenüber **Fehlinformationen** und Manipulation zu bewerten.



3. Kontinuierliche Verbesserung & Resilienz

Die gewonnenen Erkenntnisse werden genutzt, um Sicherheitsmaßnahmen zu **stärken**, **ethische Leitlinien** zu verfeinern und die **gesellschaftliche Robustheit** langfristig zu erhöhen.



Kapitel 6.6

webconsulting

Kapitel 6.6: Was ist "Red Teaming"?

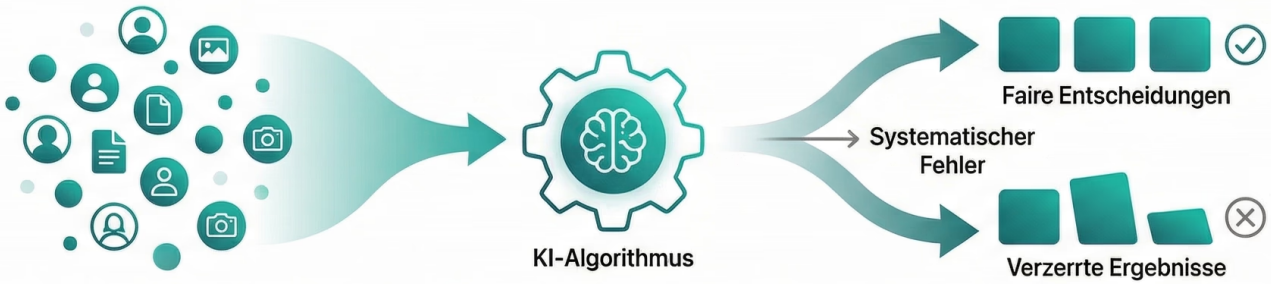
webconsulting.at/blog/ki-kompendium-2025#66-was-ist-red-teaming

Infografik: KI-generiert

ETHIK & GESELLSCHAFT

Kapitel 6.7

Was ist Bias in KI?



Bias in der Künstlichen Intelligenz (KI) bezieht sich auf systematische und wiederholbare Fehler, die zu unfairen Ergebnissen führen, indem sie bestimmte Gruppen oder Perspektiven bevorzugen oder benachteiligen. Dies entsteht oft durch voreingenommene Daten oder Algorithmen.

Daten-Bias

Wenn die zum Trainieren der KI verwendeten Daten nicht repräsentativ sind oder historische Ungleichheiten widerspiegeln, lernt das Modell diese Verzerrungen, Beispiel: Ein Gesichtserkennungssystem, das überwiegend mit Bildern einer Bevölkerungsgruppe trainiert wurde, erkennt andere weniger genau.

Algorithmus-Bias

Verzerrungen können auch durch die Gestaltung des Algorithmus selbst entstehen, etwa durch die Auswahl von Optimierungskriterien, die unbeabsichtigt bestimmte Gruppen benachteiligen, selbst bei ausgewogenen Daten.

Gesellschaftliche Auswirkungen

Bias in KI kann bestehende soziale Ungleichheiten verstärken und zu Diskriminierung in Bereichen wie Kreditvergabe, Einstellungsprozessen, Strafverfolgung und Gesundheitsversorgung führen, was Gerechtigkeit und Chancengleichheit gefährdet.

Erkennung & Minderung

Es ist entscheidend, Bias proaktiv zu erkennen und zu mindern durch diverse Datensätze, transparente Algorithmen, regelmäßige Audits und ethische Richtlinien. Werkzeuge und Frameworks helfen, Fairness zu messen und zu verbessern.

webconsulting

Kapitel 6.7: Was ist Bias in KI?

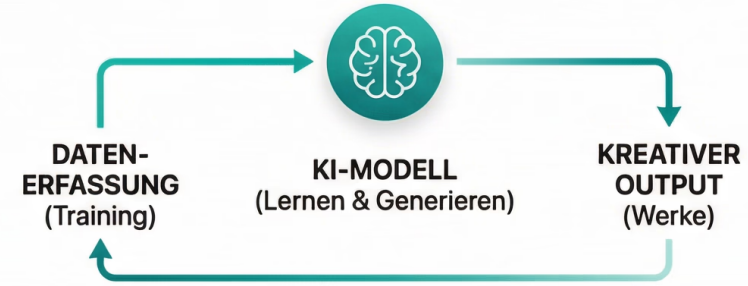
webconsulting.at/blog/ki-kompendium-2025#67-was-ist-bias-in-ki

Infografik: KI-generiert

KAPITEL 6.8: ETHIK UND GESELLSCHAFT

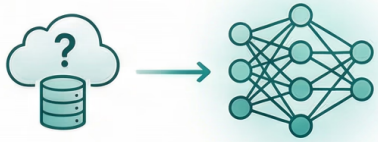
Stehlen KIs Urheberrechte?

KERNKONZEPT: Der Spannungsfeld zwischen dem Trainieren von KI-Modellen mit urheberrechtlich geschützten Daten und dem potenziellen Verstoß gegen kreative Rechte. Die ethische und gesellschaftliche Debatte dreht sich um Fairness, Vergütung und die Zukunft menschlicher Kreativität.



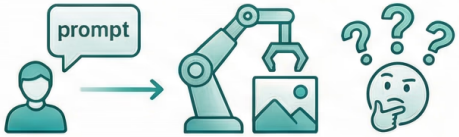
1. FAIR USE & TRAININGSDATEN

Die rechtliche Grauzone des **"Fair Use"** beim Trainieren von KIs mit riesigen Datenmengen aus dem Internet. Ist dies eine transformative Nutzung oder eine Verletzung?



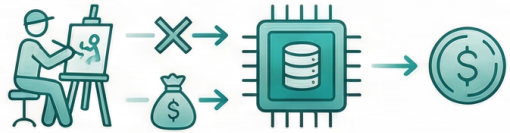
2. GENERIERTE WERKE & EIGENTUM

Wer besitzt die Rechte an KI-generierten Inhalten? Der Mensch, der den Prompt eingibt, die KI-Firma oder niemand? Die **fehlende menschliche Schöpfungshöhe** ist ein zentrales Problem.



3. VERGÜTUNG & MENSCHLICHE KREATIVITÄT

Die ethische Notwendigkeit, **Künstler und Urheber zu entschädigen,** deren Werke ohne Erlaubnis für das Training verwendet werden, um kreative Berufe nicht zu gefährden.

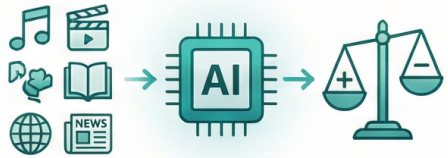


4. RECHTLICHE RAHMENBEDINGUNGEN

Die Notwendigkeit **neuer Gesetze** und **internationaler Standards**, um mit den Herausforderungen der KI-Ära Schritt zu halten und klare Regeln zu schaffen.

5. GESELLSCHAFTLICHE AUSWIRKUNGEN

Die **langfristigen Auswirkungen** auf die **Kulturproduktion**, die **Vielfalt der Inhalte** und das **Vertrauen** in digitale Medien. Eine ausgewogene Balance ist entscheidend.



webconsulting

Kapitel 6.8: Stehlen KIs Urheberrechte?

webconsulting.at/blog/ki-kompendium-2025#68-stehlen-kis-urheberrechte

Infografik: KI-generiert

Was ist der NIST AI RMF?

Ein Rahmenwerk zur systematischen Identifizierung, Bewertung und Steuerung von Risiken in KI-Systemen, um vertrauenswürdige, ethische und gesellschaftlich verantwortungsvolle Entwicklung zu fördern.

Schlüsselprinzipien (Key Principles)

Förderung von **Governance & Verantwortung**

Minimierung von **Voreingenommenheit (Bias) & Diskriminierung**

Steigerung der **Transparenz & Verständlichkeit**

Gewährleistung von **Robustheit, Sicherheit & Datenschutz**

webconsulting

Kapitel 6.9: Was ist der NIST AI RMF?

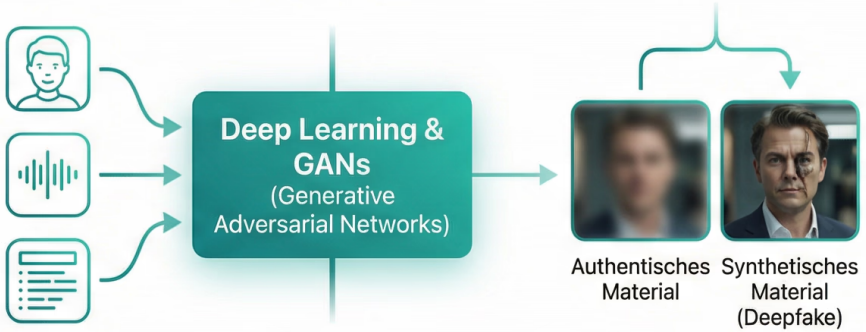
webconsulting.at/blog/ki-kompendium-2025#69-was-ist-der-nist-ai-rmf

Infografik: KI-generiert

Was ist "Deepfake"?

Synthetische Medien im Kontext von Ethik und Gesellschaft

Künstliche Intelligenz (KI) generiert realistische Fake-Inhalte, indem sie Bilder, Audio und Video manipuliert, oft ohne Zustimmung.



Technologie & Funktionsweise

Nutzt komplexe Algorithmen, um Merkmale (Gesicht, Stimme) von einer Quelle auf ein Ziel zu übertragen.

Ethik & Privatsphäre

Massive Verletzung der Privatsphäre und digitalen Identität. Gefahr von Cybermobbing und Reputationsschaden.

Desinformation & Vertrauen

Zentrale Bedrohung für die Informationsintegrität und Demokratie. Erzeugt tiefgreifenden Vertrauensverlust in Medien.

Gesellschaftliche Auswirkungen

Eskalation von Fake News, politische Manipulation und soziale Destabilisierung. Herausforderungen für die Justiz und Gesetzgebung.

Erkennung & Abwehr

Notwendigkeit fortschrittlicher Detektionstools und Medienkompetenz. Entwicklung von Authentifizierungsstandards.

webconsulting

Kapitel 6.10: Was ist "Deepfake"?

webconsulting.at/blog/ki-kompendium-2025#610-was-ist-deepfake

Infografik: KI-generiert