# The European AI Act

Eine Zusammenfassung mit den wichtigsten Punkten

# So wird KI im European AI Act definiert

Im Europäischen AI Act wird ein KI-System wie folgt definiert:

Die Vorstellung eines KI-Systems in dieser Verordnung sollte klar definiert und eng mit der Arbeit internationaler Organisationen, die sich mit künstlicher Intelligenz befassen, abgestimmt sein, um Rechtssicherheit zu gewährleisten, internationale Konvergenz und breite Akzeptanz zu fördern und gleichzeitig die Flexibilität zu bieten, um die schnellen technologischen Entwicklungen auf diesem Gebiet zu berücksichtigen.

Darüber hinaus sollte es auf Schlüsselmerkmalen von KI-Systemen basieren, **die sie von einfacheren traditionellen Software-Systemen oder Programmieransätzen unterscheiden** und sollte nicht Systeme abdecken, die allein auf von natürlichen Personen definierten Regeln basieren, um Operationen automatisch auszuführen. Ein Schlüsselmerkmal von **KI-Systemen ist ihre Fähigkeit zur Inferenz**. Diese Inferenz bezieht sich auf den Prozess der Erzeugung von Ausgaben wie Vorhersagen, Inhaltsempfehlungen oder Entscheidungen, die physische und virtuelle Umgebungen beeinflussen können, und auf die Fähigkeit von KI-Systemen, Modelle und/oder Algorithmen aus Eingaben/Daten abzuleiten.

Die Techniken, die Inferenz beim Aufbau eines KI-Systems ermöglichen, umfassen maschinelle Lernansätze, die aus Daten lernen, wie bestimmte Ziele erreicht werden können; und logik- und wissensbasierte Ansätze, die aus kodiertem Wissen oder der symbolischen Darstellung der zu lösenden Aufgabe schlussfolgern.

Die Kapazität eines KI-Systems zur Inferenz geht über die grundlegende Datenverarbeitung hinaus und ermöglicht **Lernen, Schlussfolgern oder Modellieren**. Der Begriff „maschinenbasiert“ bezieht sich darauf, dass KI-Systeme auf Maschinen laufen. Der Verweis auf explizite oder implizite Ziele unterstreicht, dass KI-Systeme gemäß explizit definierten Zielen oder impliziten Zielen operieren können. Die Ziele des KI-Systems können von dem beabsichtigten Zweck des KI-Systems in einem spezifischen Kontext abweichen.

Im Rahmen dieser Verordnung sollten Umgebungen **als die Kontexte verstanden werden, in denen die KI-Systeme operieren**, während die von dem KI-System erzeugten Ausgaben verschiedene Funktionen widerspiegeln, die von KI-Systemen ausgeführt werden, und Vorhersagen, Inhaltsempfehlungen oder Entscheidungen umfassen.

KI-Systeme sind so konzipiert, dass sie mit unterschiedlichen Autonomiegraden arbeiten, was bedeutet, dass sie einen gewissen Grad an Unabhängigkeit von menschlicher Beteiligung und die Fähigkeit haben, ohne menschliches Eingreifen zu operieren. **Die Anpassungsfähigkeit, die ein KI-System nach der Implementierung aufweisen kann, bezieht sich auf Selbstlernfähigkeiten, die es dem System ermöglichen, sich während des Einsatzes zu verändern**. KI-Systeme können eigenständig oder als Komponente eines Produkts verwendet werden, unabhängig davon, ob das System physisch in das Produkt integriert ist (eingebettet) oder die Funktionalität des Produkts dient, ohne darin integriert zu sein (nicht eingebettet)

Die Definition eines KI-Systems im Europäischen AI Act umfasst mehrere Schlüsselmerkmale und Prinzipien, die es von traditionellen Software-Systemen unterscheiden. Hier eine übersichtliche Darstellung der wichtigsten Punkte:

1. **Klar definiert und international abgestimmt:** Die Definition ist so gestaltet, dass sie Rechtssicherheit bietet, internationale Konvergenz und breite Akzeptanz fördert, und flexibel genug ist, um die schnellen technologischen Entwicklungen auf dem Gebiet der KI zu berücksichtigen.

2. **Schlüsselmerkmale von KI-Systemen:** Ein entscheidendes Merkmal von KI-Systemen ist ihre Fähigkeit zur Inferenz, d.h., sie können aus Daten Ausgaben wie Vorhersagen, Empfehlungen oder Entscheidungen ableiten, die physische und virtuelle Umgebungen beeinflussen können.

3. **Techniken zur Inferenz:** Die Definition umfasst maschinelles Lernen (das aus Daten lernt, bestimmte Ziele zu erreichen) und logik- sowie wissensbasierte Ansätze (die aus kodiertem Wissen oder symbolischen Darstellungen schlussfolgern).

4. **Lernfähigkeit, Schlussfolgern und Modellieren:** Die Kapazität eines KI-Systems, über die grundlegende Datenverarbeitung hinaus zu lernen, zu schlussfolgern oder zu modellieren, ist ein Kernaspekt.

5. **Maschinenbasiert:** KI-Systeme funktionieren auf Maschinen, was ihre operationelle Grundlage darstellt.

6. **Ziele von KI-Systemen:** KI-Systeme können nach explizit definierten oder impliziten Zielen operieren, die sich vom beabsichtigten Zweck des KI-Systems in

einem spezifischen Kontext unterscheiden können.

7. **Umfeld und Ausgaben:** Die Umgebungen, in denen KI-Systeme operieren, sowie die von ihnen erzeugten Ausgaben (z.B. Vorhersagen, Empfehlungen, Entscheidungen) sind Teil ihrer Definition.

8. **Autonomie und Adaptivität:** KI-Systeme sind für einen gewissen Grad an Unabhängigkeit von menschlicher Beteiligung und die Fähigkeit konzipiert, ohne menschliches Eingreifen zu operieren. Ihre Adaptivität ermöglicht es ihnen, sich während des Einsatzes durch Selbstlernfähigkeiten zu verändern.

9. **Einsatzmöglichkeiten:** KI-Systeme können eigenständig oder als Komponente eines Produkts verwendet werden, unabhängig davon, ob sie physisch in das Produkt integriert (eingebettet) oder nicht integriert (nicht eingebettet) sind.

Diese Punkte reflektieren den umfassenden und technologieneutralen Ansatz des Europäischen AI Acts, um sicherzustellen, dass die Definition von KI-Systemen die Vielfalt und Komplexität der Technologien und ihre Anwendungen abdeckt

# Diskussion und Ideen, die zum endgültigen AI Act geführt haben

1. **Risikobasierter Ansatz:** Der AI Act klassifiziert KI-Systeme nach dem Risiko, das sie für die Gesellschaft darstellen, in vier Kategorien: unannehmbares Risiko, hohes Risiko, begrenztes Risiko und minimales Risiko. Diese Klassifizierung bestimmt das Ausmaß der regulatorischen Anforderungen.

2. **Verbotene KI-Praktiken:** KI-Systeme, die als ein unannehmbares Risiko für die Sicherheit, die Grundrechte oder die Werte der EU betrachtet werden, sind verboten. Dazu gehören unter anderem Praktiken, die manipulatives oder ausbeuterisches Verhalten fördern oder die Privatsphäre der Menschen erheblich gefährden.

3. **Hohe Risiken und Anforderungen:** Für KI-Systeme, die als hohes Risiko eingestuft werden (z.B. in kritischen Infrastrukturen, Bildung, Beschäftigung, wesentliche private und öffentliche Dienste), legt der Act strenge Anforderungen fest. Diese umfassen Transparenz, Datenqualität, Überwachung, Sicherheit und

Robustheit.

4. **Transparenzpflichten:** Bestimmte KI-Systeme, auch solche mit begrenztem Risiko, müssen ihre Nutzer darüber informieren, dass sie mit einem KI-System interagieren. Dies betrifft vor allem Anwendungen, die menschliche Emotionen erkennen oder generieren und solche, die personalisierte Inhalte erzeugen.

5. **Datengovernance und Dokumentation:** Entwickler und Anwender von KI-Systemen müssen sicherstellen, dass die verwendeten Trainings-, Test- und Validierungsdatensätze den Qualitätsstandards entsprechen. Außerdem sind umfangreiche Dokumentationen erforderlich, um die Einhaltung der Vorschriften zu überprüfen.

6. **Marktaufsicht und Compliance:** Der Act sieht ein System der Marktaufsicht vor, einschließlich der Einrichtung nationaler Aufsichtsbehörden und eines EU-weiten KI-Aufsichtsgremiums. Unternehmen müssen sich an die Vorschriften halten und können bei Nichteinhaltung mit erheblichen Strafen rechnen.

7. **Schutz der Grundrechte:** Der Schutz der Grundrechte, insbesondere der Datenschutz und die Wahrung der Menschenwürde, ist ein zentrales Anliegen des Acts. KI-Systeme dürfen nicht in einer Weise verwendet werden, die diese Rechte verletzt.

Der Europäische AI Act ist ein ambitioniertes Vorhaben, um den Einsatz von KI-Technologien zu regulieren, das Vertrauen der Öffentlichkeit in diese Technologien zu stärken und gleichzeitig Innovation und Wettbewerbsfähigkeit in der EU zu fördern.

# RisikoGruppen

Die risikobasierte Kategorisierung im Europäischen AI Act teilt KI-Systeme in verschiedene Risikogruppen ein, basierend auf dem potenziellen Risiko, das sie für die Gesellschaft und individuelle Rechte darstellen. Hier sind die allgemeinen Risikogruppen, wie sie in den Diskussionen über den Act häufig erwähnt werden:

- **Unannehmbares Risiko:** Diese Kategorie umfasst KI-Anwendungen, die aufgrund ihres inhärenten Risikos, fundamentale Rechte zu verletzen oder die Sicherheit der Menschen zu gefährden, in der EU verboten sind.

- **Hohes Risiko:** KI-Systeme in dieser Kategorie unterliegen strengen Anforderungen in Bezug auf Transparenz, Datenqualität, Zuverlässigkeit, Überwachung und die Notwendigkeit menschlicher Aufsicht. Diese Kategorie umfasst Anwendungen in kritischen Bereichen wie Gesundheitswesen, Polizeiwesen, Verkehr, Justiz und Bildung.

- **Begrenztes Risiko:** Für KI-Systeme mit begrenztem Risiko werden spezifische Transparenzverpflichtungen vorgesehen, wie die klare Kennzeichnung von KI-generierten Inhalten oder die Information über den Einsatz von KI in bestimmten Dienstleistungen.

- **Minimales Risiko:** Die überwiegende Mehrheit der KI-Anwendungen fällt in diese Kategorie, für die keine oder nur minimale regulatorische Anforderungen vorgesehen sind. Diese Systeme gelten als sicher und positiv für die Gesellschaft und erfordern keine strengen Kontrollen.

Das Dokument, das Sie hochgeladen haben, enthält spezifische Beispiele und Definitionen für "High-Risk AI Systems", welche detailliert aufzeigen, in welchen Bereichen und unter welchen Umständen KI-Systeme als hochriskant eingestuft werden

# Hochriskante Anwendungsfälle

Der Europäische AI Act führt eine risikobasierte Kategorisierung von KI-Systemen ein, wobei bestimmte Anwendungen als "hochriskant" eingestuft werden und somit strengeren Regulierungsanforderungen unterliegen. Hier ist eine Übersicht der im Act definierten hochriskanten KI-Systeme:

**Biometrie:**

- Fernbiometrische Identifikationssysteme.
- KI-Systeme zur biometrischen Kategorisierung basierend auf sensiblen oder geschützten Merkmalen.
- KI-Systeme zur Emotionserkennung.

**Kritische Infrastruktur:**

- KI-Systeme als Sicherheitskomponenten im Management und Betrieb kritischer digitaler Infrastrukturen, Straßenverkehr und Versorgung mit Wasser, Gas, Heizung und Elektrizität.

**Bildung und berufliche Bildung:**

- Zugang oder Zulassung zu Bildungs- und Ausbildungseinrichtungen.
- Bewertung von Lernergebnissen.
- Überwachung und Erkennung verbotenen Verhaltens von Studierenden während Tests.

**Beschäftigung, Arbeitsmanagement und Zugang zur Selbstständigkeit:**

- Rekrutierung oder Auswahl von Personen, Platzierung gezielter Stellenanzeigen, Analyse und Filterung von Bewerbungen und Bewertung von Kandidaten.
- Entscheidungen, die die Arbeitsbedingungen, Beförderung und Beendigung von Arbeitsverhältnissen betreffen.

**Zugang zu und Genuss von wesentlichen privaten und öffentlichen Diensten und Leistungen:**

- Bewertung der Berechtigung natürlicher Personen für wesentliche öffentliche Unterstützungsleistungen und Dienste.
- Bewertung der Kreditwürdigkeit natürlicher Personen oder Festlegung ihrer Kreditwürdigkeit.

**Strafverfolgung (soweit gesetzlich erlaubt):**

- Risikobewertung, dass eine natürliche Person Opfer von Straftaten wird.
- Verwendung durch oder im Auftrag von Strafverfolgungsbehörden als Lügendetektoren und ähnliche Werkzeuge.
- Bewertung der Zuverlässigkeit von Beweismitteln im Verlauf von Ermittlungen oder Strafverfolgungen.

Diese Kategorisierung zeigt, wie der Europäische AI Act versucht, die Risiken von KI-Systemen zu mindern, indem er für hochriskante Anwendungen strenge Auflagen vorsieht, um Sicherheit, Transparenz und die Einhaltung von Grundrechten zu gewährleisten.

# Weitere Beispiele je nach Kategorisierung

Hier sind jeweils drei Beispiele für die verschiedenen Risikokategorien nach dem Europäischen AI Act:

## Unannehmbares Risiko

Diese Kategorie umfasst KI-Anwendungen, deren Einsatz in der EU verboten ist, weil sie fundamentale Rechte oder die Sicherheit der Bürger erheblich gefährden.

1. Sozialkreditsysteme, die das soziale Verhalten von Menschen bewerten und darauf basierend Vor- oder Nachteile gewähren.
2. Echtzeit-Gesichtserkennungssysteme in öffentlich zugänglichen Räumen für Überwachungszwecke, mit Ausnahme spezifischer, streng regulierter Ausnahmefälle.
3. KI-Systeme, die manipulatives Verhalten oder Techniken anwenden, die Personen dazu bringen können, sich selbst oder anderen unbeabsichtigt Schaden zuzufügen.

## Hohes Risiko

KI-Systeme in dieser Kategorie sind mit erheblichen potenziellen Risiken für die Gesundheit, Sicherheit oder Grundrechte verbunden und unterliegen strengen Anforderungen.

1. Kritische Infrastrukturen: KI in Verkehrssystemen oder zur Steuerung elektrischer Netze, wo Fehlfunktionen oder Ausfälle erhebliche Schäden verursachen könnten.
2. Bildung und berufliche Bildung: KI-Systeme, die für Zulassungsentscheidungen oder Bewertung von Prüfungen verwendet werden und die Bildungschancen von Personen erheblich beeinflussen können.
3. Beschäftigung, Personalwesen und Arbeitsmanagement: KI-basierte Bewerbungsauswahlverfahren oder Überwachung der Arbeitsleistung, die Diskriminierung fördern oder die Privatsphäre der Mitarbeiter verletzen könnten.

## Begrenztes Risiko

Diese Kategorie beinhaltet KI-Systeme, die bestimmte Transparenzanforderungen erfüllen müssen, um Nutzer über die KI-Interaktion zu informieren.

1. Chatbots: Nutzer müssen darüber informiert werden, dass sie mit einem KI-System interagieren, um Missverständnisse über die Art der Kommunikation zu vermeiden.
2. KI-basierte Emotionserkennungssysteme: In bestimmten Kontexten eingesetzt, müssen diese Systeme ihre Nutzer über die Emotionserkennung informieren.
3. Generative KI-Anwendungen: Systeme, die Inhalte wie Texte, Bilder oder Musik generieren, sollten Nutzer darauf hinweisen, dass die Inhalte maschinell erstellt wurden.

## Minimales Risiko

In dieser Kategorie finden sich KI-Anwendungen, die allgemein als sicher betrachtet werden und nur minimalen oder keinen regulatorischen Anforderungen unterliegen.

1. Spamfilter: KI-Systeme, die zur Filterung unerwünschter E-Mails eingesetzt werden.

2. Empfehlungssysteme: Wie jene für Bücher oder Filme, die Nutzern basierend auf ihren Präferenzen Vorschläge machen, solange diese nicht in manipulativer Weise eingesetzt werden.
3. KI-basierte Spielzeuge: Die mit Kindern interagieren können, solange diese keine Risiken für die Sicherheit oder Entwicklung der Kinder darstellen.

Diese Beispiele verdeutlichen, wie der Europäische AI Act darauf abzielt, den Einsatz von KI-Technologien zu regulieren, indem er sie nach dem Grad ihres potenziellen Risikos für die Gesellschaft kategorisiert.