AI Act: Quick Guide

Kurze Zusammenfassung des wohl finalen Textes, der wichtigsten Inhalte und des Zeitplans für die Umsetzung

Was ist der AI Act?

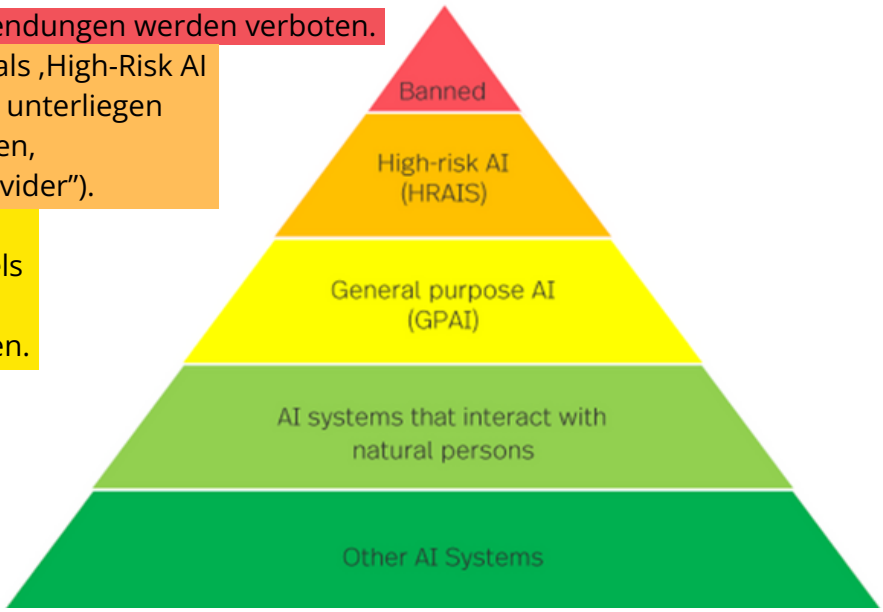
Der EU AI Act (**AIA**) ist das Flaggschiff der neuen EU-Gesetzgebung zu Künstlicher Intelligenz (KI). Der wohl finale Text des AI Acts wurde am 2. Februar 2024 vom Ausschuss der Ständigen Vertreter der Mitgliedstaaten bei der EU sowie am 13. Februar 2024 von den relevanten Ausschüssen des Europäischen Parlaments (IMCO, LIBE) angenommen. Nach seiner formellen Verabschiedung auch durch das Plenum des Europäischen Parlament und seinem Inkrafttreten wird der AIA Auswirkungen auf alle Unternehmen haben, die KI in der EU und unter Umständen auch in Drittländern entwickeln oder nutzen.

Der AIA wird Unternehmen, die an der Entwicklung, der Nutzung, dem Vertrieb oder dem Import von KI-Systemen in der EU beteiligt sind, risiko- und technologiebasierte Verpflichtungen auferlegen, die bei Nichteinhaltung mit Geldbußen (bis zu 35 Millionen Euro oder 7% des weltweiten Jahresumsatzes) geahndet werden können.

Welchen Anwendungsbereich hat der AI Act?

Die Anwendung des AIA hängt von der jeweiligen KI-Technologie, dem konkreten Anwendungsfall und der Rolle des Akteurs ab. Der Ansatz ist grundsätzlich risikobasiert:

- KI-Systeme für bestimmte Anwendungen werden verboten.
- Bestimmte KI-Systeme werden als ‚High-Risk AI Systems (HRAIS)' eingestuft und unterliegen umfangreicheren Verpflichtungen, insbesondere für Anbieter ("Provider").
- Für 'General Purpose AI (GPAI)', einschließlich Foundation Models und generativer KI, wird es besondere Bestimmungen geben.
- Andere KI-Systeme werden als risikoarm eingestuft. Sie unterliegen nur einer eingeschränkten Transparenzpflicht, wenn sie mit Menschen interagieren.



Ab wann gilt der AI Act?

Der AIA wird voraussichtlich im zweiten Quartal 2024 nach der Abstimmung im Europäischen Parlament formell verabschiedet und tritt mit der Veröffentlichung im Amtsblatt in Kraft.

Nach einer Umsetzungsfrist von grundsätzlich zwei Jahren gelten die meisten Bestimmungen des AIA. Während dieser Zeit werden verschiedene delegierte Rechtsakte, Leitlinien und Standards veröffentlicht, um die Einhaltung des AIA zu erleichtern.

Es gibt einige wichtige Ausnahmen von dieser Zweijahresfrist: Die **Verbote** für bestimmte KI-Systeme treten nach **6 Monaten** in Kraft, während die Anforderungen für **GPAI** nach **12 Monaten** in Kraft treten.

Definition KI-System

Die meisten Verpflichtungen im Rahmen des AIA beziehen sich auf KI-Systeme. Die Definition von KI-Systemen lautet:

> a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

Verbotene KI-Systeme

Der AIA wird den Einsatz bestimmter Arten von KI-Systemen verbieten. Die Verbote umfassen (unter anderem):

- Bestimmte KI-Systeme für die **biometrische Kategorisierung und Identifizierung**, einschließlich solcher für das ungezielte Auslesen von Gesichtsdaten aus dem Internet.
- KI-Systeme, die **unterschwellige Methoden** verwenden, Schwachstellen ausnutzen oder menschliches Verhalten manipulieren, um Grundrechte zu umgehen oder physischen oder psychischen Schaden zu verursachen.
- KI-Systeme zur **Emotionserkennung** am Arbeitsplatz und im Bildungswesen.
- KI-Systeme, die natürliche Personen oder Gruppen von natürlichen Personen über einen bestimmten Zeitraum auf der Grundlage ihres Sozialverhaltens bewerten oder klassifizieren (**Social Scoring**).

High-Risk AI Systems (HRAIS)

Die umfangreichsten regulatorischen Anforderungen im Rahmen des AIA gelten für *High-Risk AI Systems* (HRAIS). Dazu gehören KI-Systeme in Bereichen, die unter bestehende EU-Produktsicherheitsvorschriften fallen (ANNEX II), sowie KI-Systeme für bestimmte Zwecke, insbesondere in den folgenden Bereichen (ANNEX III):

- KI-Systeme als Sicherheitskomponenten für die Verwaltung und den Betrieb wichtiger öffentlicher Infrastrukturen wie **Wasser-, Gas- und Stromversorgung** (kritische Infrastruktur).
- KI-Systeme, die eingesetzt werden, um über den Zugang zu **Bildungseinrichtungen** zu entscheiden oder Studierende zu bewerten, z.B. KI-Systeme zur Bewertung von Prüfungen.
- KI-Systeme, die bei der **Personaleinstellung und -beschäftigung** eingesetzt werden, z. B. bei der Veröffentlichung von zielgerichteten Stellenanzeigen, der Bewertung von Bewerber:innen oder der Analyse von Bewerbungen, bei Beförderungs- oder Entlassungsentscheidungen oder bei der Überprüfung der Arbeit.
- KI-Systeme, die in den Bereichen **Migration, Asyl und Grenzkontrolle** sowie in verschiedenen anderen Bereichen der Strafverfolgung und der Justiz eingesetzt werden.
- KI-Systeme, die eingesetzt werden, um das Ergebnis **demokratischer Prozesse** oder das Wahlverhalten von Wähler:innen zu beeinflussen.
- KI-Systeme die eingesetzt werden, um die **Kreditwürdigkeit** von natürlichen Personen zu bewerten bzw. eine Risikoeinstufung von natürlichen Personen im Bereich der Lebens- und Krankenversicherung vorzunehmen.

Die Liste der High-Risk AI Systems ist nicht erschöpfend und kann in Zukunft ergänzt werden.

Die im Folgenden zusammengefassten HRAIS-Verpflichtungen gelten in erster Linie für ***Provider*** von KI-Systemen. *Provider* sind diejenigen, die ein KI-System entwickeln oder entwickeln lassen, um es unter ihrem eigenen Namen oder ihrer eigenen Marke auf den Markt zu bringen oder in Betrieb zu nehmen.

Andere Akteure ("*Operator*"), einschließlich Betreiber ("*Deployer*"), Distributoren und Importeure unterliegen weniger strengen Verpflichtungen. *Operator* können unter bestimmten Umständen auch als *Provider* angesehen werden, z.B. wenn sie ein HRAIS wesentlich verändern oder in eigenem Namen in Betrieb nehmen.

HRAIS-Provider unterliegen in Bezug auf ihr HRAIS umfassenden Verpflichtungen, einschließlich:

- **Risikomanagementsystem:** Implementierung von Prozessen für den gesamten Lebenszyklus von HRAIS zur Identifizierung, Analyse und Minderung von Risiken.
- **Daten und Datenverwaltungsmaßnahmen:** HRAIS müssen nach strengen Datenverwaltungsmaßnahmen trainiert und getestet werden.
- **Technische Dokumentation:** Erstellung eines umfassenden „Handbuchs" für HRAIS mit spezifischen Mindestinformationen.
- **Protokollierung:** HRAIS müssen so konzipiert sein, dass Ereignisse wie Nutzungszeiten und Eingabedaten automatisch protokolliert werden. Diese Protokolle müssen vom Provider für bestimmte Zeiträume aufbewahrt werden.
- **Transparenz:** HRAIS müssen von detaillierten Informationen über ihre Eigenschaften, Möglichkeiten und Grenzen begleitet werden.
- **Menschliche Aufsicht ("*Human Oversight*"):** HRAIS müssen so gestaltet werden, dass sie von Menschen überwacht werden können, die verschiedene Anforderungen erfüllen müssen, wie z.B. das jeweilige HRAIS zu verstehen („KI-Kompetenz") und seine Nutzung zu unterbinden.
- **Genauigkeit, Robustheit und Cyber-Sicherheit:** HRAIS müssen genau sein (mit Genauigkeitsmetriken in der Bedienungsanleitung), robust gegenüber Fehlern oder Inkonsistenzen (z. B. durch ausfallsichere Plänen) und robust gegenüber Cyberangriffen.
- **Qualitätsmanagementsystem:** *Provider* müssen ein umfassendes Qualitätsmanagementsystem einrichten.
- **Überwachung nach der Markteinführung:** *Provider* müssen ein System zur Verfügung stellen, das in der Lage ist, Daten über die Leistung des HRAIS während der gesamten Lebensdauer des HRAIS zu sammeln und zu analysieren.

HRAIS-Provider unterliegen verschiedenen Verfahrenspflichten, bevor sie ihr HRAIS anbieten können:

- **CE-Kennzeichnung:** *Provider* müssen sicherstellen, dass ihr HRAIS vor der Auslieferung einem Konformitätsbewertungsverfahren unterzogen wird und dass ihre Dokumentation die CE-Kennzeichnung trägt.
- **Registrierung in EU-Datenbank:** *Provider* und öffentliche Einrichtungen, die HRAIS verwenden, müssen das HRAIS in einer EU-weiten Datenbank für KI-Systeme registrieren.
- **Meldepflichten:** *Provider* müssen schwerwiegende Vorfälle oder Störungen, die ihr HRAIS betreffen, innerhalb von 15 Tagen an die zuständige Behörde melden.

Andere *HRAIS-Operator* haben begrenztere Pflichten, wie z.B. die Durchführung von Folgenabschätzungen in Bezug auf Grundrechte, die Sicherstellung, dass das HRAIS gemäß der Bedienungsanleitung verwendet wird, die Überwachung des Betriebs des HRAIS und das Führen von Aufzeichnungen über die vom HRAIS erzeugten Protokolle (soweit sie ihrer Kontrolle unterliegen).

General Purpose AI (GPAI)

Grundsätzlich gilt, dass KI-Technologien, die nicht verboten oder als High-Risk eingestuft sind, weniger strengen regulatorischen Anforderungen unterliegen.

Besonderheiten gelten allerdings für **General Purpose AI Modelle** (**GPAI**). Die Anforderungen an die meisten GPAI-Modelle, einschließlich Foundation Models und generativer KI-Modelle, betreffen in erster Linie die Transparenz.

Zu den Pflichten für alle *GPAI-Provider* gehören die Pflicht zur Veröffentlichung der technischen Dokumentation, die Einhaltung des EU-Urheberrechts und die Bereitstellung einer Zusammenfassung der verwendeten Trainingsdaten.

Der finale Text enthält zusätzliche Anforderungen für GPAI, die auf großen Datensätzen trainiert wurden und eine überdurchschnittliche Leistung ("*high impact capabilities*") aufweisen, basierend auf den potenziellen systemischen Risiken, die diese KI-Modelle in der gesamten Wertschöpfungskette darstellen können (**GPAI with systemic risk**).

Jedes **GPAI-Modell mit systemischem Risiko** wird **zusätzlichen Anforderungen** unterliegen, die voraussichtlich Folgendes umfassen:

- Strenge **Modellevaluierungen**, einschließlich *Adverserial Testing/Red-Teaming*.
- Bewertung und **Minderung möglicher systemischer Risiken** durch den Einsatz von GPAI.
- Verschärfung der **Meldepflichten** gegenüber den Aufsichtsbehörden, insbesondere bei schwerwiegenden Vorfällen.
- Gewährleistung einer angemessenen **Cybersicherheit** für GPAI mit systemischem Risiko.
- Bericht über die **Energieeffizienz** des GPAI.

Sonstige KI-Systeme

Abgesehen von den oben genannten Fälen und abgesehen von den Fällen, die vom Anwendungsbereich des AI Acts ausgenommen sind (Militär/Verteidigung; wissenschaftliche Forschung und Entwicklung), besteht die einzige verbindliche Anforderung für andere KI-Systeme in einer begrenzten Verpflichtung zur **Transparenz**: *Provider* müssen sicherstellen, dass KI-Systeme, die für die Interaktion mit Menschen bestimmt sind, so konzipiert und entwickelt werden, dass die einzelnen Nutzer wissen, dass sie mit einem KI-System interagieren.

Der finale Text des AIA enthält nicht die vom Europäischen Parlament vorgeschlagenen **'General Principles'** für KI, die in einem früheren Entwurf des AIA enthalten waren. Diese 'General Principles' bilden jedoch weiterhin die Grundlage für viele Bestimmungen des AIA.

Sanktionen

Die im Rahmen des AIA zu erlassenden Sanktionen richten sich nach Art des Verstoßes und der Größe des Unternehmens. Es können Sanktionen in Höhe von 7,5 Mio. EUR (oder 1,5% des weltweiten Jahresumsatzes) bis zu 35 Mio. EUR (oder 7% des weltweiten Jahresumsatzes) für das vorangegangene Geschäftsjahr anfallen.

Die Inhalte dieses Quick Guides wurden **freundlicherweise von Simmons & Simmons zur Verfügung gestellt**. Simmons & Simmons ist eine der führenden Kanzleien im Bereich KI und Partner des KI Bundesverbandes. Die internationale Kanzlei besteht in Deutschland aus rund 100 Anwalt:innen, die in verschiedenen Praxisgruppen und Rechtsgebieten tätig sind. Simmons & Simmons berät zu rechtlichen und regulatorischen Fragen im Zusammenhang mit dem Anwendungsbereich und der Anwendung des AIA, einschließlich der Erstellung von Folgenabschätzungen und Risikoanalysen.

Für Rückfragen steht euch unser **Ansprechpartner bei Simmons & Simmons** gerne zur Verfügung: **Christopher Götz, LL.M.** (christopher.goetz@simmons-simmons.com).

Der KI Bundesverband dankt Simmons & Simmons für die Bereitstellung der Inhalte.

KI BUNDESVERBAND