

Checkliste

Verbotene Praktik Nr. 3
Soziale Bewertung oder Einstufung
i.S.d. Art 5 Abs. 1 c.) KI-VO, ErwGr 31

Praktische Relevanz



Technische Relevanz



Rechtliche Relevanz



Orga.Relevanz



Verortung im Prüfungsschema

I. Bereichsausnahmen

II. Anwendungsbereiche

III. Risikoeinordnung

a. Risikoeinordnung – Verbotene Praktiken gem. Art. 1 Abs. 2 b.) i.V.m. Art. 5 KI-VO

i. Überblick Verbotene KI-Praktiken

ii. Verbotene KI-Praktiken im Einzelnen

1. Unterschwelligen Beeinflussung und Manipulation Art. 5 Abs. 1 a.) KI-VO, 29 ErwGr

2. Ausnutzen von Vulnerabilität und Schutzbedürftigkeit Art. 5 Abs. 1 b.) KI-VO

→ **3. Soziale Bewertung oder Soziale Einstufung Art. 5 Abs. 1 c.) KI-VO ErwGr 31**

Beachte hierzu auch:

Orientierungshilfe Verbotene KI-Praktiken Überblick (Version 1.1.),

Orientierungshilfe Verbotene KI-Praktik Nr. 3 (Version 1.3.), Übersicht Verbotene KI-Praktiken (Version 1.4)

Einleitung

Bevor wir uns den Einsatz von KI-Systemen zur sozialen Bewertung nach Art. 5 Abs. 1 lit. c der KI-VO genauer anschauen, solltet ihr wissen: Diese Vorschrift verbietet KI-basiertes Social Scoring, bei dem Personen oder Gruppen anhand ihres sozialen Verhaltens oder persönlicher Eigenschaften eingestuft und dadurch ungerechtfertigt benachteiligt werden. Anders als in früheren Entwürfen, die lediglich das staatlich initiierte oder beauftragte Social Scoring in den Fokus nahmen, erfasst die finale Regelung sowohl öffentliche **als auch private Akteure**. Zudem schärft die KI-VO das bestehende Datenschutzrecht und macht deutlich: Ungerechtfertigte soziale Einordnungen und daraus resultierende Nachteile sind im europäischen Rechtsraum nicht akzeptabel.

Warum gibt es die Regelung

Hintergrund dieser klaren Verbotsnorm ist vor allem das Ziel, einer Praxis zuvorzukommen, die bereits in anderen Teilen der Welt für Unbehagen und Kritik sorgt: das staatsnahe oder privatwirtschaftliche Einstufen von Personen nach vermeintlichem sozialen Wert. Indem die KI-VO diese Regelungen auf alle Akteure ausdehnt, unterstreicht sie zudem den Vorrang des Datenschutzrechts und macht unmissverständlich deutlich, dass Entscheidungen über Personen nicht ohne tragfähige Rechtsgrundlage, klare Zwecke und einen angemessenen Ausgleich ihrer Interessen getroffen werden dürfen.

Definition KI-VO

Art. 5 Abs. 1 c.) KI-VO, 31 ErwGr definiert die Praktik folgendermaßen:

(...) das Inverkehrbringen, die Inbetriebnahme oder die Verwendung von KI Systemen zur Bewertung oder Einstufung von natürlichen Personen oder Gruppen von Personen über einen bestimmten Zeitraum auf der Grundlage ihres sozialen Verhaltens oder bekannter, abgeleiteter oder vorhergesagter persönlicher Eigenschaften oder Persönlichkeitsmerkmale, wobei die soziale Bewertung zu einem oder beiden der folgenden Ergebnisse führt:

- 1. Schlechterstellung oder Benachteiligung bestimmter natürlicher Personen oder Gruppen von Personen in sozialen Zusammenhängen, die in keinem Zusammenhang zu den Umständen stehen, unter denen die Daten ursprünglich erzeugt oder erhoben wurden;*
- 2. Schlechterstellung oder Benachteiligung bestimmter natürlicher Personen oder Gruppen von Personen in einer Weise, die im Hinblick auf ihr soziales Verhalten oder dessen Tragweite ungerechtfertigt oder unverhältnismäßig ist;*

Checkliste Verbotene KI-Praktik Nr. 3 i.S.d. Art. 5 Abs. 1c.) KI-VO, ErwGr 31

Zur besseren Orientierung und für ein formalisiertes Vorgehen wurde die folgende Checkliste entwickelt, um festzustellen, ob eine eine verbotene KI-Praktik nach Art. 5 Abs. 1 c) KI-VO, ErwGr 31 vorliegt.

Voraussetzungen

(1) Objektiver Tatbestand

(a) KI-Technologie

- (i) KI-System (Art. 3 Nr. 1 KI-VO)

(b) Zielgruppe

Betroffen sind natürliche Personen oder Gruppen von Personen, deren persönliches Verhalten, Eigenschaften oder Persönlichkeitsmerkmale zum Gegenstand einer sozialen Bewertung gemacht werden

- (i) Personen
- (ii) Personengruppen

(c) Handlungen

- (i) Handlung I: Liegt mind. einer der rollenspezifischen Handlungen vor?

1. Inverkehrbringen (Art. 3 Nr. 9 KI-VO)
2. Inbetriebnahme (Art. 3 Nr. 11 KI-VO) oder
3. die Verwendung (Art. 3 Nr. 4 KI-VO)

- (ii) Handlung II: Liegt einer der verbotenen Praktiken vor?

1. Verbotene Praktik

Ein bloßes finanzielles oder organisatorisches Rating ohne soziale Wertung stellt demnach kein „Social Scoring“ im Sinne des Verbots dar.

a. Bewertung oder

Auslegungshilfe: Nicht erfasst sind rein zweckgebundene, verkehrübliche Bewertungen bei denen es nicht zu einem allgemeinen moralischen Werturteil über die Person als Mitbürger kommt.

Ein KI-gestütztes Scoring, das durch Inferenz oder Vorhersage persönliche oder Persönlichkeitsmerkmale generiert, fällt unter das Verbot, sofern alle weiteren Voraussetzungen erfüllt sind.¹

Nicht erfasst sind Bewertungen, die sich zwar auf persönliche Merkmale beziehen, die aber kein derartiges allgemeines Werturteil beinhalten.

Kein Social Scoring liegt vor, wenn lediglich die Zahlungsfähigkeit oder Zahlungsbereitschaft natürlicher Personen bewertet wird, oder auch deren berufliche Leistungsfähigkeit, weil derartigen Bewertungen **nach der Verkehrsaufassung kein entsprechendes Werturteil (oder Unwerturteil) innewohnt**. Dabei kommt es allerdings – jedenfalls in der Variante der „Verwendung eines KI-Systems“ – **auf den tatsächlichen Verwendungskontext an**.

b. Einstufung

2. Über einen bestimmten Zeitraum

Auslegungshilfe: Diese Bewertung muss über einen bestimmten Zeitraum hinweg erfolgen, sodass eine allgemeine Aussage über eine Person oder Personengruppe getroffen wird. Ein einmaliges Ereignis oder eine singuläre Bewertung genügt nicht.

¹ Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act), S. 54-55, Abs. 159

Auch wenn die Daten nur scheinbar getrennt erhoben werden, kann eine Zusammenführung durch KI über Zeiträume hinweg zur Anwendung des Verbots führen (Stichwort: „dataveillance“).²

3. Auf der Grundlage zahlreicher Datenpunkte

a. Persönliche Eigenschaften

- i. Bekannte
- ii. Abgeleitete (vermutete)
- iii. Vorhergesagte

b. Persönlichkeitsmerkmale

c. Soziales Verhalten

2) Kausalität: Abstrakte Gefährdung („kann“ in Zusammenhang stehen (Siehe ErwGr 31)

Nach ErwGr 31 ist für das Vorliegen der Kausalität keine konkret nachweisbare Schadensverursachung erforderlich. Vielmehr genügt es, dass der Einsatz des KI-Systems objektiv geeignet ist („kann“), diskriminierende oder benachteiligende Effekte hervorzurufen. **Eine bloß potenzielle Beeinträchtigung der geschützten Interessen ist damit ausreichend**, um die Kausalität im Sinne des Art. 5 Abs. 1 c) KI-VO zu bejahen.³

Ein schädigender Effekt kann auch vorliegen, wenn der Score von einem Dritten erstellt, aber durch eine andere Stelle verwendet wird.⁴

3) Subjektiver Tatbestand

(a) Absichtsvermutung i.S.d. ErwGr 29

- (i) Die Absicht des Anbieters oder Betreibers muss sich nicht auf die Verursachung des Schadens beziehen.
- (ii) Es genügt eine objektive Wirkung, auch wenn die Manipulationsabsicht schwer nachweisbar ist.

(b) Ausnahme: Keine Absichtsvermutung i.S.d. ErwGr 29, wenn vorliegende Faktoren vorliegen:

1. Keine Absichtsvermutung-Faktoren

- a. Faktoren, welche das Verhalten wesentlich beeinflussen
- b. Faktoren, die nicht Teil des KI-Systems sind
- c. Faktoren liegen außerhalb der Kontrolle des Akteurs, weil:
 - i. Faktoren konnten vom Akteur vernünftigerweise nicht vorhergesehen werden, oder
 - ii. Faktoren konnten vom Akteur vernünftigerweise nicht gemindert werden

4) Schaden

(1) Schlechterstellung / Benachteiligung der Zielgruppe

Auslegungshilfe:

- (a) In sozialen Zusammenhängen/Kontexten die in keinem Zusammenhang zu den Umständen stehen, unter denen die Daten
 - (i) Ursprünglich erzeugt wurden
 - (ii) Ursprünglich erhoben wurden

² Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act), S. 55, Abs. 157

³ Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act), S. 56, Abs. 160

⁴ Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act), S. 56, Abs. 162

(b) „in einer Weise“

(i) Ungerechtfertigt hinsichtlich sozialen Verhaltens

1. Soziales Verhalten
2. ungerechtfertigt

(ii) Unverhältnismäßig hinsichtlich der Tragweite des sozialen Verhaltens

1. Tragweite des sozialen Verhaltens
2. Unverhältnismäßig

Eine Benachteiligung kann auch in Form von Ausschluss aus sozialen Leistungen oder Diensten erfolgen, selbst ohne direkten Schaden (z.B. Zugangsbeschränkung zu Kredit, Wohnung, Beschäftigung).⁵

(2) Inakzeptable Bewertungspraktiken gem. ErwGr 31.

Auch eine bevorzugende Behandlung bestimmter Gruppen kann als indirekt benachteiligend gegenüber anderen gelten und fällt damit unter das Verbot.⁶

⁵ Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act), S. 57, Abs. 166

⁶ Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act), S. 59, Abs. 169