Deutsche Stiftung für Recht und Informatik

MANIPULATIONEN DURCH KÜNSTLICHE INTELLIGENZ UND DEEPFAKES

HERAUSFORDERUNGEN IM BEREICH DER REGULATORIK UND DES ZIVILPROZESSES

Yannick Zirnstein
Clyde & Co Europe LLP

Herbstakademie 2024

Übersicht

EINLEITUNG

Begriffsklärung & Missbrauchspotenzial

Was ist unter KI-Manipulation, bzw. Deepfake zu verstehen?

Keine eigene Definition einer KI-Manipulation

Begriffsannäherung: Medieninhalte, deren Informationsgehalt oder Erscheinungsform durch die Nutzung von KI-Systemen verändert wurden.

Deepfake in der KI-VO

“einen durch KI **erzeugten** oder **manipulierten Bild-, Ton- oder Videoinhalt**, der wirklichen Personen, Gegenständen, Orten, Einrichtungen oder Ereignissen **ähnelt** und einer Person **fälschlicherweise als echt oder wahrheitsgemäß erscheinen würde**”
(Art. 3 Nr. 60 KI-VO)

Welche Missbrauchsmöglichkeiten gibt es?

Falschdarstellungen

Betrug,
Erpressung,
Vermögensschäden

Fake News,
Wahlmanipulation

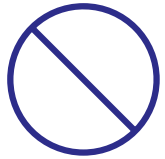
REGULATORISCHE HERAUSFORDERUNG

Anwendungsbereich der KI-VO

Relevante Begriffsbestimmungen

KI-System (Art. 3 Nr. 1 KI-VO)	Anbieter (Art. 3 Nr. 3 KI-VO)	Betreiber (Art. 3 Nr. 4 KI-VO)
„ein **maschinengestütztes System**, das für einen in unterschiedlichem Grade **autonomen Betrieb** ausgelegt ist und das nach seiner Betriebsaufnahme **anpassungsfähig** sein kann und das aus den **erhaltenen Eingaben** für explizite oder implizite **Ziele ableitet**, wie Ausgaben wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können“	„eine natürliche oder juristische Person, Behörde, Einrichtung oder sonstige Stelle, die ein **KI-System** oder ein KI-Modell mit allgemeinem Verwendungszweck **entwickelt oder entwickeln lässt** und es unter ihrem eigenen Namen oder ihrer Handelsmarke **in Verkehr bringt** oder das KI-System unter ihrem eigenen Namen oder ihrer Handelsmarke **in Betrieb** nimmt, sei es entgeltlich oder unentgeltlich“	„eine natürliche oder juristische Person, Behörde, Einrichtung oder sonstige Stelle, die ein **KI-System** in **eigener Verantwortung verwendet**, es sei denn, das KI-System wird im Rahmen einer **persönlichen und nicht beruflichen Tätigkeit verwendet**“

Einordnung des KI-Systems in Risikokategorien



Verbotene KI-Praktiken (Art. 5 KI-VO)

"*Techniken der unterschwelligen Beeinflussung außerhalb des Bewusstseins einer Person oder absichtlich manipulative oder täuschende Techniken*" (Art. 5 Abs. 1 lit. a) KI-VO)

✗ KI-Manipulationen/Deepfakes beeinflussen nicht "unterschwellig", bzw. "außerhalb des Bewusstseins".

Hochrisiko-KI-Systeme (Art. 6 ff. KI-VO)

✗ Keine Qualifizierung aufgrund Harmonisierungsvorschriften (Art. 6 Abs. 1 lit. a) i. V. m. Anhang I KI-VO)

✗ Keine Qualifizierung aufgrund von Bereichsbestimmung (Art. 6 Abs. 2 i. V. m. Anhang III KI-VO), da KI-Systeme zur Erstellung von KI-Manipulationen/Deepfakes wahrscheinlich nicht „*bestimmungsgemäß*" hierfür entworfen wurden.

Sonstige KI-Systeme (insbesondere: Art. 50 KI-VO)

✓ Insbesondere: Transparenzpflichten für KI-Systeme mit minimalen / keinem Risiko

Transparenzpflichten

Anbieter (Art. 50 Abs. 2 KI-VO)

„Anbieter von KI-Systemen (…) die **synthetische Audio-, Bild- Video- oder Textinhalte erzeugen**, stellen sicher, dass die Ausgaben des KI-Systems in einem **maschinenlesbaren Format gekennzeichnet** und als **künstlich erzeugt** oder **manipuliert erkennbar** sind.

Die Anbieter sorgen dafür, dass - soweit technisch möglich - ihre **technischen Lösungen wirksam, interoperabel, belastbar und zuverlässig** sind (...).“

Betreiber (Art. 50 Abs. 4 KI-VO)

„**Betreiber** eines KI-Systems, das Bild-, Ton- oder Videoinhalte erzeugt oder manipuliert, die ein **Deepfake** sind, müssen **offenlegen**, dass die Inhalte **künstlich erzeugt** oder **manipuliert** wurden.

(…)

Ist der Inhalt Teil eines **offensichtlich künstlerischen**, **kreativen**, **satirischen**, **fiktionalen** oder **analogen Werks** oder Programms, so **beschränken** sich die in diesem Absatz festgelegten Transparenzpflichten darauf, das Vorhandensein solcher erzeugten oder manipulierten Inhalte in **geeigneter Weise offenzulegen**, die die Darstellung oder den Genuss des **Werks nicht beeinträchtigt**.“

Rechtsschutz?

Rechtsbehelfe in der KI-VO

- Recht auf Beschwerde bei einer Marktüberwachungsbehörde (Art. 85 KI-VO)
- Recht auf Erläuterung der Entscheidungsfindung im Einzelfall (Art. 86 KI-VO)
- Meldung von Verstößen und Schutz von Hinweisgebern (Art. 87 KI-VO)

Rückgriff auf bereits bekannte Ansprüche

- Rechte betroffener Personen aus der DSGVO (Art. 15-22, 82 DSGVO)
- Verletzung des Rechts am eigenen Bild (insbesondere: § 22 KUG)
- Deliktsschutz, APR (§ § 823 Abs. 1, 1004 Abs. 1 BGB)

ZIVILPROZESSUALE HERAUSFORDERUNG

KI-Manipulationen/Deepfakes als Gegenstand des Zivilprozesses

Identifizierung des Anspruchgegners

- Problem: Verbreitung von KI-Manipulationen/Deepfakes durch “anonyme” Nutzerkonten
- Auskunft über Bestandsdaten (§ 21 Abs. 2 TDDDG):

  Weiterhin ungeklärte Rechtslage über Klarnamenpflicht in sozialen Netzwerken
- Auskunft über Nutzungsdaten (§ 24 TDDDG):

  Nicht für Privatpersonen.

Entwurf der KI-Haftungs-Richtlinie

- Keine wesentliche Hilfestellung für betroffene Personen, da:
  - Offenlegung von Beweismitteln / Beweislastregelungen häufig nicht erforderlich.
  - Widerlegbare Vermutung der Kausalität zwischen Verschulden des Anspruchsgegners und Ausgabe des KI-Systems nicht passend.

Täuschung durch KI-Manipulationen/Deepfakes als Beweismittel

Einordnung als Beweismittel

Urkunde?

Nur bei KI-Manipulationen;

Nur bei erforderlicher schriftlicher Verkörperung mittels anderer technischer Hilfmittel.

Zeuge/SV?

Nur bei Vernehmung des Zeugen/Sachverständigen durch Bild- und Tonübertragung (§ 128a Abs. 2 Satz 1 ZPO).

Augenscheinsobjekt?

Elektronische Dokumente (§ 371 Abs. 1 Satz 2 ZPO):

Daten und technische Aufzeichnungen, insbesondere Grafik-, Audio- und Videodateien.

Täuschung durch KI-Manipulationen/Deepfakes als Beweismittel

Beweisverwertungsverbote und weitere Angriffspunkte

Beweisverwertungsverbot wg. Eingriff in APR

- Eingriff in die Intimsphäre führt stets zur Unverwertbarkeit des Beweismittels.
- Eingriff in die Privat-/Sozialsphäre führt zur Güterabwägung.

→ Kein Interesse am Beweis nicht wahrheitsgemäßer Abläufe!

Fehlende Authentizität / Integrität

- Authentizität: Beweismittel kann dem Urheber zugeschrieben werden.
- Integrität: Objekt stimmt mit seinem Inhalt zum Zeitpunkt der Erstellung überein.
- Beweislast hierfür liegt beim Beweisführer.

Verknüpfung mit Transparenzpflichten (Art. 50 KI-VO), kann bei Darlegung fehlender Authentizität / Integrität helfen.

FAZIT

Takeaways

1.

KI-Manipulationen und Deepfakes können sinnvolle Zwecke haben, bergen jedoch auch Missbrauchspotenzial.

Das Missbrauchspotenzial ist wegen der breiten technischen Verfügbarkeit und des leichten Einstiegs erheblich.

2.

Die Regulatorik bietet keine spezifischen Mittel zur adäquaten Begegnung dieser Risiken, kann aber im Einzelfall die Anspruchsdurchsetzung erleichtern.

Der Rückgriff auf bereits bekannte Anspruchsgrundlagen ist aus materiell-rechtlicher Perspektive möglich und bedingt wirksam.

3.

Die Anspruchsdurchsetzung ist mit Problemen verbunden, die nicht unmittelbar durch die Eigenschaften eines Deepfakes verursacht sind.

Die zivilprozessualen Regelungen zu Beweismitteln bieten belastbare Mittel zum Schutz vor Deepfakes als Beweismittel.

VIELEN DANK!

RA Yannick Zirnstein
Clyde & Co Europe LLP
yannick.zirnstein@clydeco.com