Deutsche Stiftung für Recht und Informatik

UNRECHTMÄßIGE NUTZUNG VON KI-TRAININGSDATEN ALS GEFAHR FÜR NEUE GESCHÄFTSMODELLE?

COMPLIANCE-ANFORDERUNGEN BEI DER ENTWICKLUNG VON KI-SYSTEMEN

RA Jasper Siems / RA Thomas Repka

PXR Legal Berlin; Doktorand Universität Osnabrück /
NEUWERK Rechtsanwälte Hamburg

Herbstakademie 2021

Gliederung

- 1. Einleitung
- 2. Daten-Compliance und Geschäftsgeheimnisschutz
- 3. Daten-Compliance und Urheberrecht
- 4. Daten-Compliance und Datenschutzrecht
- 5. Fazit

1. Einleitung

- Trainingsdaten je nach Anwendungszweck des KI-Systems in unterschiedlicher Gestalt
- Drei Kategorien von Trainingsdaten unterscheidbar:
  - Inhalte (Bsp.: Texte oder Bilder)
  - Rohdaten (Bsp.: Maschinendaten)
  - Personenbezogene Daten (Bsp.: Bewerberdaten)
- Folge: unterschiedliche rechtliche Behandlung
- (P) KI-Entwickler = juristische Laien
  - Zu hohes Problembewusstsein und keine Verwendung „fremder“ Daten → schlechtere KI-Systeme
  - Zu geringes Problembewusstsein und Verwendung aller Daten auch problematischer Provenienz → Haftung
- Abhilfe: Daten-Compliance

1. Einleitung

- Daten-Compliance
  - Compliance an sich: systematische Organisation von Unternehmen, um Risiko von Haftung zu verringern
- Daten-Compliance bezüglich KI-Systeme
  - Ziel: vor und während der Entwicklung sowie Anwendung die Verletzung von Rechten Dritter zu verringern
  - Ansatz: verschiedene Maßnahmen etablieren
- Besonders sensible Situationen
  - Ankauf oder Lizenzierung fremder Daten
  - Einbringung von Daten durch Kooperationspartner, Auftragnehmer, Mitarbeiter etc.
- Zentrale Rechtsgebiete: Geheimnisschutz, Urheberrecht, Datenschutzrecht

2. Daten-Compliance und Geschäftsgeheimnisschutz

- Geheimnisschutz für KI-Trainingsdatensets leicht gegeben
  - Arg.: Relativer Geheimnisbegriff in § 2 Nr. 1 lit. a
  - Umgekehrt: Hohe Gefahr fremde Geschäftsgeheimnisse rechtswidrig zu nutzen
- Drittwirkung des Schutzes von Geschäftsgeheimnissen und damit Erstreckung auf mittelbare Verletzungen, § 4 Abs. 3
  - Insbesondere rechtsverletzende Produkte, § 2 Nr. 4
- (P) Vergiftung des gesamten Trainingsdatensets bei Beruhen in erheblichen Umfang auf inkriminierten Trainingsdaten

2. Daten-Compliance und Geschäftsgeheimnisschutz

- Compliance Maßnahmen im Vorfeld
  - Vermeidung von Verschulden nach § 4 Abs. 3
    - Haftung nur bei Verschulden im Sinne einer Bösgläubigkeit
      - Nachforschungspflichten
    - Etablierung von Kriterien- und Prüfungskatalog
    - Gremienentscheidung (Juristen:innen, Domainexpert:innen, KI-Expert:innen)
  - Haftungsfreistellungsklauseln bei Lizenzierung fremder Daten
    - (P) Unterlassungsanspruch
- Compliance Maßnahmen im Nachhinein
  - Einwand der Unverhältnismäßigkeit aus § 9
    - § 9 Nr. 3: insb. bei zunächst Gutgläubigkeit
    - § 9 Nr. 5: Nachweis und Protokollierung der Investitionen
  - Geldabfindungsanspruch aus § 11

3. Daten-Compliance und Urheberrecht

- Urheberrechtsschutz
  - Für einzelne Daten als Inhalte; zum Beispiel:
    - Bilder aus § 2 Abs. 1 Nr. 5 bzw. § 72
    - Texte aus § 2 Abs. 1 Nr. 1
  - Datensets über § 87a
- Relevante Verwertungshandlungen bei der Entwicklung von KI-Systemen
  - Erstmalige Speicherung bereits Vervielfältigung
  - Aufbereitung der Daten für Verarbeitung durch KI-System
  - Vervielfältigung jedenfalls im Laufe des Trainings

3. Daten-Compliance und Urheberrecht

- Ansatz: keine Vermeidung urheberrechtlich relevanter Handlungen, sondern rechtmäßiges Vorgehen
- Anknüpfungspunkt: neue TDM-Schranke aus § 44b
  - Wortlaut unklar, aber Sinn und Zweck spricht für Anwendung auf KI-Training, arg.: Erwgr. 18 DSM-RL
  - Ansatz: Bei Sammeln von Informationen, insb. im Internet, Mechanismus etablieren, der maschinenlesbare Vorbehalte erfasst und Sammlung dann unterlässt
- Bei Verletzung: Unterlassungsanspruch aus § 97 Abs. 1 ist weniger weitreichend
  - Nur bezüglich Trainingsdaten
  - Nicht möglich: Unterlassung der weiteren Nutzung der auf inkriminierten Trainingsdaten beruhenden KI-Systeme

4. Daten-Compliance und Datenschutzrecht

- Spannungsfeld zwischen Datenschutz und Entwicklerinteresse, möglichst viele Daten zu nutzen
  - Notwendigkeit von psb. Daten, Bsp.: Gesichtserkennung
  - Auch im Übrigen Interesse an psb. Daten, um KI-System qualitativ besser zu machen
- Ausgangspunkt: Trainingsdaten = personenbezogen?
- Legaldefinition: Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen
- Abgrenzung zu anonymisierten Daten
  - Vorgang der Anonymisierung ist Datenverarbeitung
  - (P) Re-Identifizierung durch technische Weiterentwicklung
  - Verantwortlichkeit des Datenverarbeiters zur Prüfung der Anonymisierung

4. Daten-Compliance und Datenschutzrecht

- Infizierung des KI-Systems durch ein datenschutzwidriges Datum, Möglichkeit des Herausfilterns eines Datums technisch noch nicht abschließend erforscht
- Rechtlicher Gestaltungsspielraum durch Vorgaben insbesondere der DSGVO eingeschränkt → Einhaltung der Vorgaben der DSGVO ist oberstes Ziel
- Maßnahmen der Daten-Compliance
  - Prüfung der Erforderlichkeit psb. Daten
  - Nutzung sog. synthetischer Daten möglich?
  - Prüfung der datenschutzrechtl. Rolle (Verantwortlicher, gemeinsame Verantwortlichkeit, AVV?)

4. Daten-Compliance und Datenschutzrecht

- Denkbare Rechtsgrundlage der Datenverarbeitung(en)
  - Einwilligung (Art. 6 Abs. 1 S.1 lit. a; Art. 9 Abs. 1), (P) freie Widerrufbarkeit
  - Vertrag (Art. 6 Abs. 1 S. 1 lit. b)
  - Berechtigte Interessen (Art. 6 Abs. 1 S. 1 lit. f)
- Datenschutz-Folgenabschätzung erforderlich?
- Datenschutzkonzept
- Dokumentation (insb. Herkunft, (P) Zweckänderung)
- Implementierung eines Melde- und Compliance-Systems bei Bekanntwerden von Verstößen

- (P) Einkauf von Trainingsdaten
  - Umfangreiche Prüfung der Herkunft der Daten
  - Absicherung des Erwerbers über vertragliche Sicherungsmechanismen

5. Fazit

- Daten-Compliance für KI-Trainingsdaten zentrales Compliance-Thema für Tech-Unternehmen
- Besonders weitreichend: Geschäftsgeheimnisschutz, da sehr weitgehender Unterlassungsanspruch, Haftungsbegrenzung über Unverhältnismäßigkeitseinwand
- Urheberrecht: berechtigter Nutzer im Rahmen der TDM-Schranke, weniger weitgehender Unterlassungs-/Beseitigungsanspruch
- Datenschutzrecht: wenig rechtliche Gestaltungsoptionen, Kontrolle der Datenherkunft essentiell, Dokumentation und Meldesystem zur Begrenzung von Sanktionen
  - Vorschlag: Privilegierungstb. für KI-Entwicklung