

# Generative KI und geistiges Eigentum - Risiken und Lösungsmodelle

Prof. Dr. Nils Heide  
Rechtsanwalt

Nabugu / Shutterstock

# New York Times vs Open AI Inc./Microsoft

UNITED STATES DISTRICT COURT  
SOUTHERN DISTRICT OF NEW YORK

THE NEW YORK TIMES COMPANY

Plaintiff,

v.

MICROSOFT CORPORATION, OPENAI, INC.,  
OPENAI LP, OPENAI GP, LLC, OPENAI, LLC,  
OPENAI OPKO LLC, OPENAI GLOBAL LLC,  
OAI CORPORATION, LLC, and OPENAI  
HOLDINGS, LLC,

Defendants.

Civil Action No. \_\_\_\_\_

**COMPLAINT**

**JURY TRIAL DEMANDED**

Plaintiff The New York Times Company (“The Times”), by its attorneys Susman Godfrey LLP and Rothwell, Figg, Ernst & Manbeck, P.C., for its complaint against Defendants Microsoft Corporation (“Microsoft”) and OpenAI, Inc., OpenAI LP, OpenAI GP LLC, OpenAI LLC, OpenAI OpCo LLC, OpenAI Global LLC, OAI Corporation, LLC, OpenAI Holdings, LLC, (collectively “OpenAI” and, with Microsoft, “Defendants”), alleges as follows:

**I. NATURE OF THE ACTION**

1. Independent journalism is vital to our democracy. It is also increasingly rare and valuable. For more than 170 years, The Times has given the world deeply reported, expert, independent journalism. Times journalists go where the story is, often at great risk and cost, to inform the public about important and pressing issues. They bear witness to conflict and disasters,

# Rechte des geistigen Eigentums

1. Schutzrechte (Patente, Marken, Designrechte etc.)
2. Urheberrechte/Leistungsschutzrechte
3. Schutz von geheimen Informationen

Abzugrenzende Rechte mit ähnlichen Schutzgegenständen und Schutzrichtungen:

- Schutz von Persönlichkeitsrechten
- Datenschutzrecht

# Einsatzformen der KI im Journalismus

1. KI-Einsatz im Rahmen der Texterstellung
  - 1.1 *Vorbereitung von Textgrundlagen*
  - 1.2 *Autonome Generierung von Texten*
    - Abfassung einfacher Berichte mit reiner Informationswiedergabe (Wetter- oder Verkehrsberichte, Wiedergabe von einfachen Sport- oder Börsendaten)
  - 1.3 *Kürzung von Artikeln durch die KI*
  - 1.4 *Erstellung synthetischer Medien zur Informationsvermittlung*
    - Avatare als Informationsvermittler oder Text-to-Speech-Anwendungen

# Einsatzformen der KI im Journalismus

## 2. KI Instrumente beim Vertrieb von Medieninhalten

2.1 *Anpassung von Inhalten und Texten an spezifische Nutzergruppen oder Nutzer*

2.2 *Newsbots als Form des Dialog-Journalismus*

## 3. KI basierte Recherchen- und Datenanalyse

3.1 *Automatisierte Trendanalysen*

3.2 *Automatisierte Faktenüberprüfung*

3.3 *Analyse und Aufbereitung großer Datenmengen*

- *Dokumentenanalyse im Bereich des investigativen Journalismus (z. B. Panama Papers)*

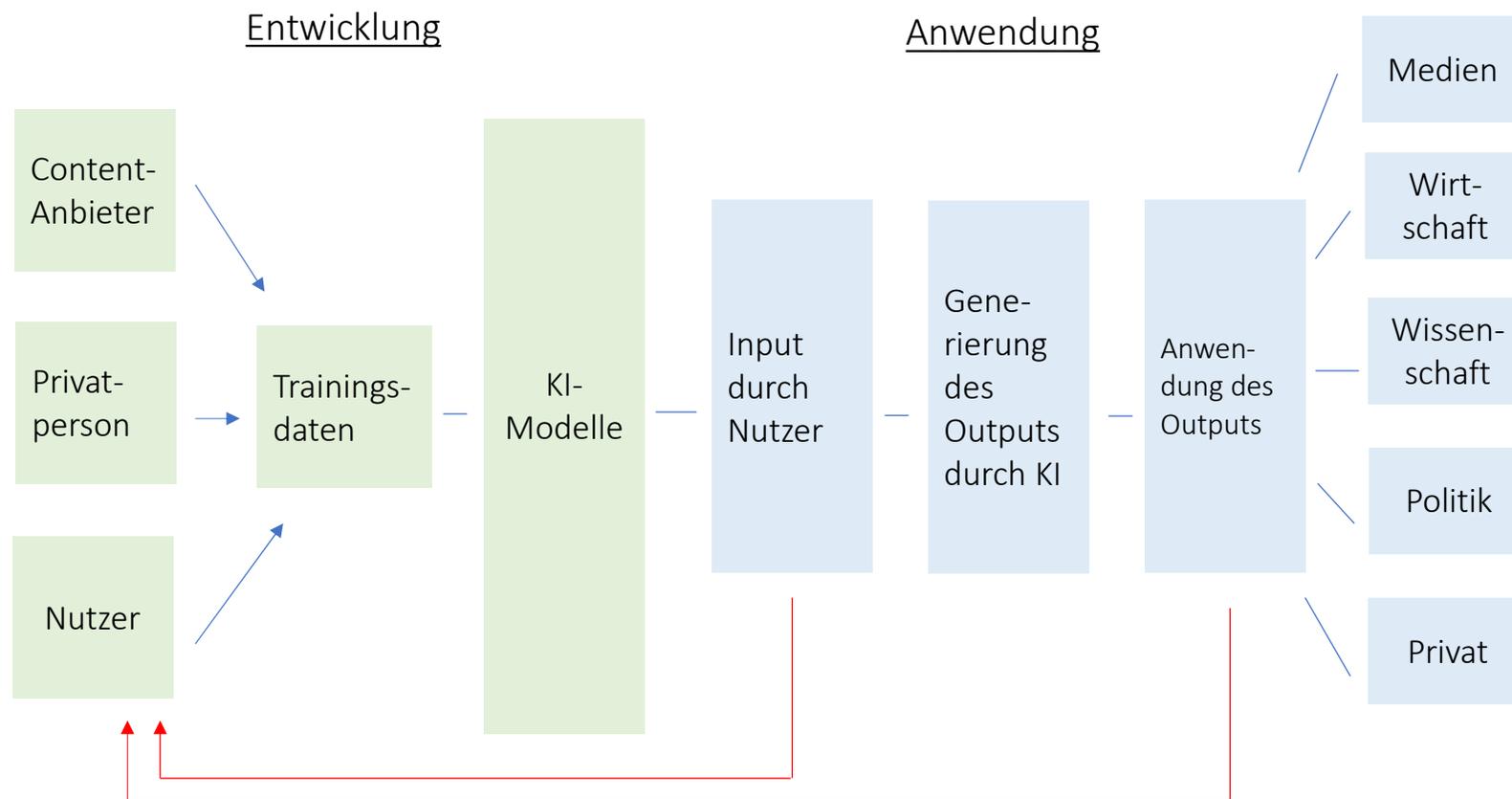
# Besonderer rechtlicher Regelungsrahmen im Journalismus

1. Grundrechtliche Presse- und Rundfunkfreiheit nach Art. 5 Abs. 1 GG führt zu besonderen Schutzmechanismen und Privilegien
2. Pressegesetze und anerkannte journalistische Grundsätze führen zu erhöhten Sorgfaltsanforderungen
3. Strittige Diskussion zur Anwendung presserechtlicher Grundsätze auf den sogenannten „Roboter-Journalismus“
4. Bedeutung des Schutzes des geistigen Eigentums wird in allen Richtlinien im journalistischen Bereich hervorgehoben (*siehe z. B. dpa, DJV, Paris-Charta zu KI und Journalismus etc.*)

# Ambivalenz des Einsatzes der generativen KI im Journalismus aus der Perspektive des geistigen Eigentums?

1. „Ausbeutung“ geschützter Werke durch das Training der generativen KI
  - *Größere gerichtliche Verfahren eingeleitet: Authors Guild vs Open AI, Getty Images vs Open AI, New York Times vs. Open AI, Thomson Reuters vs Ross etc.*
2. Risiken der Rechtsverletzung beim Einsatz der generativen KI in der journalistischen Arbeit
3. Geschäfts-/Lizenzmodelle zur Verwertung von Texten oder sonstigen Inhalten
  - z.B. Kooperation Axel Springer und Open AI, welche unter anderem die Bereitstellung von Axel Springer-Daten als Trainingsdaten für Open AI umfassen soll

# Mögliche betroffene Rechte des geistigen Eigentums beim Einsatz der generativen KI





# „Infizierung“ durch rechtliche Probleme hinsichtlich der Rohdatengewinnung

1. Fortwirkung möglicher rechtlicher Probleme der Trainingsdaten in der Anwenderkette  
⇒ „Infizierung“ von KI basierten journalistischen Texten
  
2. Probleme einer rechtlich restriktiven Regelungspraxis
  - 2.1 Bedeutung von Trainingsdaten
    - Datenarmut führt zur Verschlechterung von KI-Systemen (Datenquantität und Datenqualität)
    - „Unumkehrbarkeit“ des KI Einsatz des im Journalismus(?)
  
  - 2.2 Schaffung rechtlich geregelter Anreize zur Datenbereitstellung
    - Vermeidung des Daten-„Raubrittertums“

# Rechtliche Grundsätze zur Verwertung von Rohdaten

1. Soweit kein rechtlicher Schutz der Daten besteht, ist eine Nutzung von Daten möglich
  - Grundsatz der Freiheit von Fakten und Informationen
  
2. Besonderer rechtlicher Schutz der Daten?
  - (1) Vertragliche Bindungen
  - (2) Schutzrechte (Patente, Marken etc.)
  - (3) Geschäftsgeheimnisgesetz
  - (4) Urheberrechtlicher Schutz / Leistungsschutzrechte
  - (5) Persönlichkeitsrechte nach § 823 BGB
  - (6) Recht am eigenen Bild nach dem Kunsturhebergesetz (KUG) und § 201a StGB
  - (7) Spezieller Unterlagenschutz
  - (8) Datenschutz

# Urheberrechtlicher Schutz von Rohdaten und Trainingsdaten

1. Text-, Bild-, Video- oder Audiodateien als Werke nach § 2 UrhG
  - 1.1 Anforderungen an die Schöpfungshöhe im Urheberrecht
  - 1.2 Wirtschaftlicher oder sonstiger Wert der in den Daten verkörperten Informationen ist für die urheberrechtliche Beurteilung nicht maßgeblich
2. Handhabung von Datensammlungen als Datenbankwerke nach § 4 UrhG?
  - Persönliche geistige Schöpfungsleistung im Rahmen der systematischen oder methodischen Anordnung der Elemente

# Leistungsschutzrecht für Datenbankhersteller

1. Schutz von Datenbanken nach § 87a UrhG
  - 1.1 Sammlung von Werken, Daten oder anderen unabhängigen Elementen, „*die systematisch oder methodisch angeordnet ...*“ sind
  - 1.2 „*deren Beschaffung, Überprüfung und Darstellung eine nach Art und Umfang wesentliche Investition erfordert*“
2. Kriterium „wesentlichen Investition“ (strittig)
  - Herrschende Meinung: Investitionen in die Erzeugung von Daten werden nicht erfasst, sondern nur Kosten in die Beschaffung und Strukturierung

# Schutz von geheimen Informationen

## 1. Schutzanforderungen nach § 2 Nr. 1 Geschäftsgeheimnisgesetz

- (1) Informationen sind geheim
- (2) Information haben wirtschaftlichen Wert
- (3) Angemessene Schutzmaßnahmen

## 2. Kriterium der Geheimhaltung

- Strukturierte Datensets und derivative Daten können auch bei bekannten Rohdaten geheim sein ( z. B. Betriebsdaten einer Anlage bei Berichterstattung über Emissionen)

## 3. Gesetzlicher Geheimhaltungsschutz für bestimmte Sektoren

- Ermittlungsakten im strafrechtlichen Bereich
- Quellenschutz im journalistischen Bereich

## 4. Geheimhaltungsvereinbarungen

# Schutz von personenbezogenen Informationen

1. Datenschutz (DSGVO etc.) = gesonderter Rechtsbereich
2. Schutz von Persönlichkeitsrechten und Rechte am eigenen Bild
  - 2.1 Bild-, Text- oder Audiodateien, welche individualisierbare Merkmale eines Menschen erkennen lassen
  - 2.2 Verletzung von Persönlichkeitsrechten nach § 823 BGB durch eine unberechtigte Verwendung oder Verwendung im nicht autorisierten Kontext
  - 2.3 Schutz des Rechtes am eigenen Bild nach § 22 KUG

# Urheberrechtsverletzung bei der Gewinnung von Trainingsdaten

1. Soweit urheberrechtlich geschützte Texte, Bilder oder sonstige Werke betroffen sind, dürfte zumindest das Vervielfältigungsrecht nach § 16 UrhG betroffen sein:  
= Urheberrechtsverletzung
  
2. Schrankenregelung für das Data Mining
  - 2.1 zu wissenschaftlichen Zwecken nach § 60d UrhG
  
  - 2.2 für kommerzielle und sonstige Zwecke nach § 44b UrhG:  
erlaubt ist die „*automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen*“

# Urheberrechtsverletzung bei der Gewinnung von Trainingsdaten

1. Grenzen des § 44b UrhG
  - 1.1 „Opt-out“ Möglichkeit nach § 44b Abs. 3 UrhG
    - Text- und Datamining ist nur dann zulässig, wenn sich der Rechteinhaber diese Rechte nicht ausdrücklich durch ein Opt-Out vorbehalten hat
    - Bei online verfügbaren Werken muss der Nutzungsvorbehalt in *maschinenlesbarer Form* erfolgen, damit der Opt-Out für automatisierte Crawler erkennbar ist
  - 1.2 Löschungspflicht nach Wegfall der Erforderlichkeit nach § 44b Abs. 2 UrhG
2. § 44b UrhG schafft nur für das Text- und Data Mining ab dem 07.07.2021 eine Ausnahme von der Zustimmungspflicht (*mögliche Infizierung von Daten für zuvor erstellte KI-Systeme?*)

# Fortbestehende Streitfragen zur Nutzung von urheberrechtlich geschützten Werken bei der Erstellung von Trainingsdaten

1. Noch keine höchstrichterliche Klärung zur hinreichenden rechtlichen Grundlage für das Training generativer KI-Anwendungen
  - § 44b UrhG
  - „Fair Use“ - Prinzip (USA)
2. Zeitlicher Wettlauf zwischen Technik und Recht?
  - Rechtsfindung im Instanzenzug
3. Nationale Gesetze/Urteile im digitalen „Global Village“?

# Verletzung von Rechten durch Output Nutzung

1. „Infizierung“ des Outputs mit Rechten Dritter
2. Nutzungsbedingungen der Anbieter generativer KI
  - 2.1 Übertragung nur der in der eigenen Inhaberschaft des KI Anbieters befindlichen Rechte
  - 2.2 Verantwortungszuweisung für Verwendung des Outputs an Nutzer
    - (1) Keine Freistellungsverpflichtung der KI-Anbieter
    - (2) Teilweise Änderungen in den Nutzungsbedingungen
      - Microsoft und Google (Duet AI, Vertex AI etc.)



Gorodenkoff / Shutterstock

# Verletzung von Rechten durch die Output Nutzung

1. Aufgrund der autonomen Textsynthese erfolgt zumeist keine identische Übernahme der beim Training verwendeten Daten im Output, jedoch kann die Verwendung von Bild- oder Textfragmenten zu einer Rechtsverletzung führen
2. Auch die Verwertung von Bearbeitungen oder Umgestaltungen eines urheberrechtlich geschützten Werkes kann eine Urheberrechtsverletzung darstellen, soweit kein „hinreichender Abstand“ vom Originalwerk nach § 23 Abs. 1 UrhG vorliegt

# Bestimmung eines „hinreichenden Abstandes“ nach § 23 Abs. 1 Satz UrhG

1. Interdependenz zwischen der *Schöpfungshöhe des Ausgangswerkes* und *Umfang der notwendigen Bearbeitung*
2. Geringeres Risiko einer Urheberrechtsverletzung bei *reinen Sachtexten*, welche weniger durch die Individualität des Sprachwerkes als durch den inhaltlichen Gehalt geprägt werden
3. Nachkontrolle des Outputs durch einen Plagiatsscanner und/oder durch menschliche Nachbearbeitung und Kontrolle

# Darlegungslast

*Order on Motion to Dismiss vom 12. Februar 2024 – Tremblay v OpenAI, Inc., 3:23-cv-03223, (N.D.Cal. Feb 12, 2024) ECF Nr. 104*

"Distinctly, Plaintiffs here have not alleged that the ChatGPT outputs contain direct copies of the copyrighted books. Because they fail to allege direct copying, they must show a substantial similarity between the outputs and the copyrighted materials" [...] Plaintiffs' allegation that "every output of the OpenAI Language Models is an infringing derivative work" is insufficient. [...] For the foregoing reasons, the Court GRANTS in part and DENIES in part Defendants' motions to dismiss. Plaintiffs' amended complaint must be filed by March 13, 2024."

# Rechtsverletzung durch Eingabe von Inputs

1. Rechte an den Inputs gehören Nutzer, jedoch Nutzungsvorbehalte der KI-Anbieter
  - *Mittlerweile Ausnahmen in Nutzungsbedingungen einzelner Anbieter*
2. Verstoß gegen das Geschäftsgeheimnisgesetz und Geheimhaltungsvereinbarungen
3. Verstoß gegen sonstige Rechte (personenbezogene Informationen etc.)

# Kann der Nutzer für die Outputs selbst eigene Rechte beanspruchen?

## 1. Nutzungsbedingungen der KI-Anbieter

### 1.1 Rechteeinräumung am Output

### 1.2 Beschränkungen

#### (1) Nutzungsvorbehalte der KI-Anbieter

- Mittlerweile Ausnahmen in Nutzungsbedingungen einzelner KI-Anbieter

#### (2) Verweis auf mögliche übereinstimmende Outputs = keine exklusiven Rechte



Serhii Yevdokymov / Shutterstock

# Kann der Nutzer für die Outputs selbst eigene Rechte beanspruchen?

2. Outputs generativer KI als urheberrechtlich geschütztes Werk?
  - 2.1 Persönliche geistige Schöpfung eines Menschen nach § 2 Abs. 2 UrhG
    - Voraussetzung noch angemessen im KI Zeitalter?
  - 2.2 Begründung eines menschlichen Schöpfungsanteils
    - (1) Durch Prompts?
    - (2) Durch Programmierung?
    - (3) Durch Nachbearbeitung?



NMStudio789 / Shutterstock

# Schutz journalistischer Inhalte durch das UrhG

1. Journalistische Texte sind zumeist als Sprachwerke nach § 2 Abs. 1 Nr. 1 UrhG geschützt, soweit diese eine geistige Schöpfungshöhe nach § 2 Abs. 2 UrhG aufweisen
  - (1) Kein urheberrechtlicher Schutz für reine Nachrichteninhalte, soweit sich diese auf eine reine Informationswiedergabe beschränken
    - Urheberrechtlicher Schutz indes bei Berichten, Reportagen und Interviews
  - (2) Kein urheberrechtlicher Schutz für die von der KI autonom erzeugten Werke
    - Nachbearbeitung durch einen menschlichen Journalisten begründet dann einen urheberrechtlichen Schutz von KI-generierten Texten, wenn diese selbst eine gewisse Schöpfungshöhe erreicht.
    - *Bloße Kontrolle* eines KI-generierten Textes ist *nicht* hinreichend

# Schutz journalistischer Inhalte durch das UrhG

2. Sonstige urheberrechtliche Schutzgegenstände der journalistischen Arbeit
  - 2.1 Fotos und Lichtbildwerke sind nach § 2 Abs. 1 Nr. 5 UrhG bzw. als Lichtbilder nach § 72 Abs. 1 UrhG geschützt
  - 2.2 Filmwerke nach § 2 Abs. 1 Nr. 6 UrhG bzw. Laufbilder nach § 95 UrhG
  - 2.3 Bestimmung menschlicher Beiträge bei KI generierten Bild- und Filmwerken durch die Nachbearbeitung
  
3. Bestimmung des urheberrechtlichen Schutzes bei KI basierten Formatwechseln
  - 3.1 Text-to-speech Anwendung
  - 3.2 Text-to-picture/video
    - Open AI Sora
  - 3.3 KI basierte Synthesen (z.B. Avatare)

# Schutz journalistischer Werke durch das UrhG

## 4. Größere Bedeutung von Leistungsschutzrechten bei KI-generierten Texten?

### 4.1 Schutz journalistischer Datenbank nach § 87a UrhG?

- Problem?: Investitionen in die Generierung von Daten (journalistische Texterstellung etc.) stellen nach der Rechtsprechung keine nach § 87a UrhG zu berücksichtigende Investitionen dar

### 4.2 Schutz des Presseverlegers nach § 87f Abs. 1 UrhG – strittig

- (1) Teilweise wird ein Leistungsschutzrecht für *computergenerierte journalistische Inhalte* verneint, weil es sich nicht um journalistische Beiträge handelt
- (2) Gute Gründe für eine Bejahung des Leistungsschutzrecht des Presseverlegers, da § 87f UrhG nicht an die schöpferische Leistung, sondern an die wirtschaftliche, technische und organisatorische Leistung des Presseverlegers anknüpft

# Schutzwürdigkeit eigener Datenbanken und KI-Modelle im journalistischen Bereich

1. Importierung von strukturierten Daten aus Datenquellen mit Hilfe der Data Mining-Technologie
  - 1.1 Datenquellen
    - Redaktionseigene Datenbanken oder Datenbanken Dritter
  - 1.2 Datengenerierung
    - (1) Aktuelle Informationen zum Ereignis, das Gegenstand des Berichtes ist (z.B. Tod einer prominenten Person, Zinssenkung etc.)
    - (2) Kombination mit allgemein verfügbaren Zusatzinformationen zum Hintergrund
2. Analyse der Daten nach vorgegebenen Regeln zur Gewichtung des Nachrichtenwertes der Informationen
3. Erstellung einer Sprachfassung auf Grundlage von Natural Language Generation (NLG)

# Schutzwürdigkeit eigener Datenbanken und K- Modelle im journalistischen Bereich

1. Aufbau und Pflege von Datengrundlagen
  - 1.1 *Auswahl und Sammlung von Inhalten*
  - 1.2 *Gewichtung von Inhalten*
  - 1.3 *Prüfung auf sachliche Richtigkeit und Vereinbarkeit mit ethischen und rechtlichen Maßgaben (Diskriminierungsfreiheit etc.)*
2. Darstellungsqualität, Sprachstile und Formate
3. Investitionen in Erstellung, Sammlung, Strukturierung, Aufbereitung

# Verantwortung und Haftung für mögliche Verletzungen von Rechten des geistigen Eigentums

1. Keine Rechts- oder Teilrechtsfähigkeit der KI
2. Haftung des die KI einsetzenden Medienunternehmens
  - 2.1 Innerorganisatorische Haftungszuweisung (Redaktionen, Einzeljournalisten etc.)?
  - 2.2 Innerorganisatorische Organisationsverantwortung und Richtlinien als Anknüpfungspunkte für die Haftung

# Verantwortung und Haftung für mögliche Verletzungen von Rechten des geistigen Eigentums

3. Haftungsbeziehung zwischen den Medienunternehmen und den KI-Anbietern
  - 3.1 Vertragliche Verantwortungszuweisung und Freistellungsregelungen berühren nicht die Position des Medienunternehmens als Anspruchsadressat
  - 3.2 Durchsetzungsrisiken bei Freistellungsregelungen

# Anpassung der eingesetzten KI Systeme an Anforderungen hinsichtlich der Rechtskonformität

1. Versionen mit unterschiedlichen rechtlichen Schutzniveaus
  - ChatGPT Free, ChatGPT Plus, ChatGPT Enterprise, ChatGPT Team etc.
  - Microsoft Copilot Versionen für kommerzielle Nutzungen
  - Google bietet mit Bard und Vertex AI-Versionenvarianten
2. Differenzierungen hinsichtlich Datenschutz, Geschäftsgeheimnissen, Urheberrechten etc.
  - Sicherheitsstandards bei der Eingabe von geheimen Daten und personenbezogenen Daten
  - Zusicherungen zur Vermeidung von Rechtsverletzungen und Freistellungsregelungen
  - Teilweise Abschluss mit EU Tochtergesellschaften und „European Terms of Use“

# Lizenzmodelle für journalistische Inhalte im Bereich der generativen KI

1. Urheberrechte und Leistungsschutzrechte für journalistische Inhalte und Datenbanken als Grundlage von Verwertungsmodellen
2. Lizenzmodelle zur Zugänglichmachung von Daten als Trainingsdatenbasis für KI Systeme
  - Vereinbarung von Associated Press (AP) mit Open AI  
(Open AI erhält Zugriff auf AP-Berichte bis 1985)
  - Vereinbarung zwischen Open AI und Axel Springer

# Lizenzmodelle für journalistische Inhalte im Bereich der generativen KI

## 3. Regelungsziele?

- 3.1 Akute Datenarmut durch verstärkte Regulierung des Datenzugangs
- 3.2 Erhöhte Anforderung an die Datenbeschaffung und Transparenz durch den AI Act
  - Interessenidentität der Verpflichteten
- 3.2 Schaffung des Zugangs zu journalistischen Datenbanken als Trainingsdaten zur Verbesserung der Datenqualität und Datenquantität
  - Datenpluralismus zur Sicherung des Meinungspluralismus
- 3.3 Wirtschaftliche Verwertungsmodelle

# Lizenzmodelle für journalistische Inhalte im Bereich der generativen KI

## 4. Regelungsbedarf

- 4.1 Sicherung der Kontrolle im Rahmen der Nutzung für KI Trainingsdaten
- 4.2 Reichweite des Nutzungsrechts für Eigennutzungen und Drittnutzungen
- 4.3 Gewährleistungen und Freistellungspflichten
- 4.4 Vergütungsmodelle

# Organisationsmaßnahmen zum Schutz des geistigen Eigentums

1. Identifizierung von Bereichen, welche für den Einsatz der generativen KI unter Abwägung aller Risiken geeignet sind
2. Auswahl der generativen KI Tools nach Risikokategorien unter Berücksichtigung der jeweiligen Schutzmaßnahmen der KI Tools (aktuelle Nutzungsbedingungen, Leistungsbeschreibung, Sitzort und Zusicherungen der Anbieter)
3. Etablierung von Richtlinien für die Handhabung
4. Unterweisung von Mitarbeitern

# Organisationsmaßnahmen zum Schutz des geistigen Eigentums

5. Nachkontrolle und Nachbearbeitung der Ergebnisse der generativen KI
  - Outputs der generativen KI unterliegen keinen systeminternen Richtigkeitskontrollen, das heißt Institutionalisierung einer menschlichen Nachkontrolle zur Erfüllung der Organisationspflicht
  - Nachbearbeitung auch zur Begründung eines urheberrechtlichen Schutzes
  
6. Auditing hinsichtlich der Entwicklungen des Einsatzes

Vielen Dank!