

Scraping im Kontext von Massenverfahren und Künstlicher Intelligenz

Dr. Hannah von Wickede und Johannes Berchtold, LL.M.
Reed Smith LLP

I. Was ist Scraping?

II. Relevante Akteure

III. Rechtliche Bewertung von Scraping

IV. Risiken und praktische Implikationen für die Akteure

V. Zusammenfassung und Ausblick

I. Was ist Scraping?

Was ist Scraping?

- **Definition:**

“Web-Scraping, das „Aus- oder Abkratzen“ von Daten aus dem Internet, ist ein technisches Verfahren zur automatisierten Extraktion von (öffentlich) im Internet verfügbaren Daten aus Websites.”

(Quelle: Leistner/Zurth, in: Loewenheim UrhR-HdB, § 49 Leistungsschutz von Datenbanken, Rn. 137)

- **Zwecke:** Die vom Scraper subjektiv verfolgten Zwecke hängen vom Einzelfall ab.
- **Welche Rolle spielt Scraping bei Künstlicher Intelligenz und in Massenverfahren?**

Scraping als Gegenstand von Massenverfahren

- Öffentlich verfügbare personenbezogene Daten werden von Online-Präsenzen gescraped.
- Verbraucherkanzleien akquirieren “betroffene” Personen und klagen unter anderem Schadensersatz nach Art. 82 DSGVO ein.
- **Argument:** Daten hätten nicht öffentlich verfügbar sein dürfen und/oder Daten hätten gegen Scraping geschützt werden müssen.

Scraping von künftigen Trainingsdaten

- **Was ist künstliche Intelligenz?**

“Künstliche Intelligenz ist die Fähigkeit einer Maschine, menschliche Fähigkeiten wie logisches Denken, Lernen, Planen und Kreativität zu imitieren.”

(Quelle: <https://www.europarl.europa.eu/topics/de/article/20200827STO85804/was-ist-kunstliche-intelligenz-und-wie-wird-sie-genutzt>)

- Gescrapte und aufbereitete Daten können dazu dienen, ein KI - System zu trainieren (**“Trainingsdaten”**).
- Qualitativ hochwertige Trainingsdaten sind ein entscheidender Faktor für die Prognosegenauigkeit eines KI - Systems.

II. Relevante Akteure

Webseitenbetreiber



Betroffene Person



Rechtsanwalt



Scraper



Anbieter eines KI-Systems



III. Rechtliche Bewertung von Scraping

Anwendbare Rechtsnormen

Vielzahl an gesetzlichen Regelungen zu beachten, zum Beispiel:

- Immaterialgüterrecht
- Datenschutzrecht
- Zivilrecht
- Künftig: KI-VO

Scraping aus urheberrechtlicher Sicht

- UrhG enthält in §§ 44b, 60d UrhG Regelungen zum Text und Data Mining.
- Das deutsche bzw. europäische Datenschutzrecht enthält keine “Pendant-Normen”.
- Urheberrechtliche Implikationen im Datenschutzrecht?

Scraping aus datenschutzrechtlicher Sicht

Ist der sachliche Anwendungsbereich eröffnet?

- Werden personenbezogene Daten i.S.d. Art. 4 Nr. 1 DSGVO scraped?
- Ist Scraping eine Verarbeitung i.S.d. Art. 4 Nr. 2 DSGVO?
- Wer ist Verantwortlicher i.S.d. Art. 4 Nr. 7 DSGVO?

Ist der räumliche Anwendungsbereich eröffnet?

Europäische Union



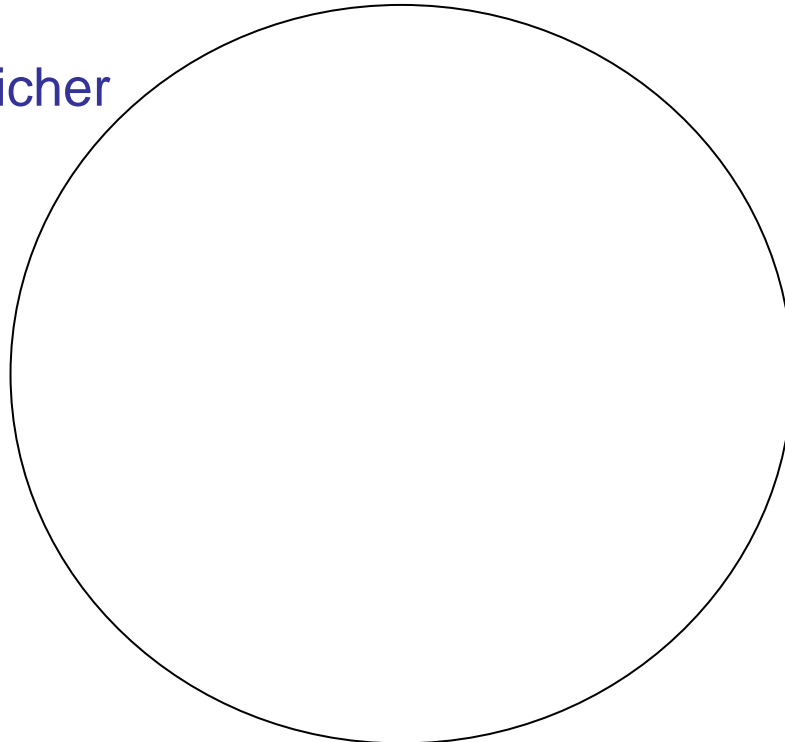
Art. 3 Abs. 1 DSGVO:

Verarbeitung im Rahmen der Tätigkeit einer Niederlassung eines Verantwortlichen oder Auftragsverarbeiters in der Europäischen Union.

Ist der räumliche Anwendungsbereich eröffnet?

Europäische Union

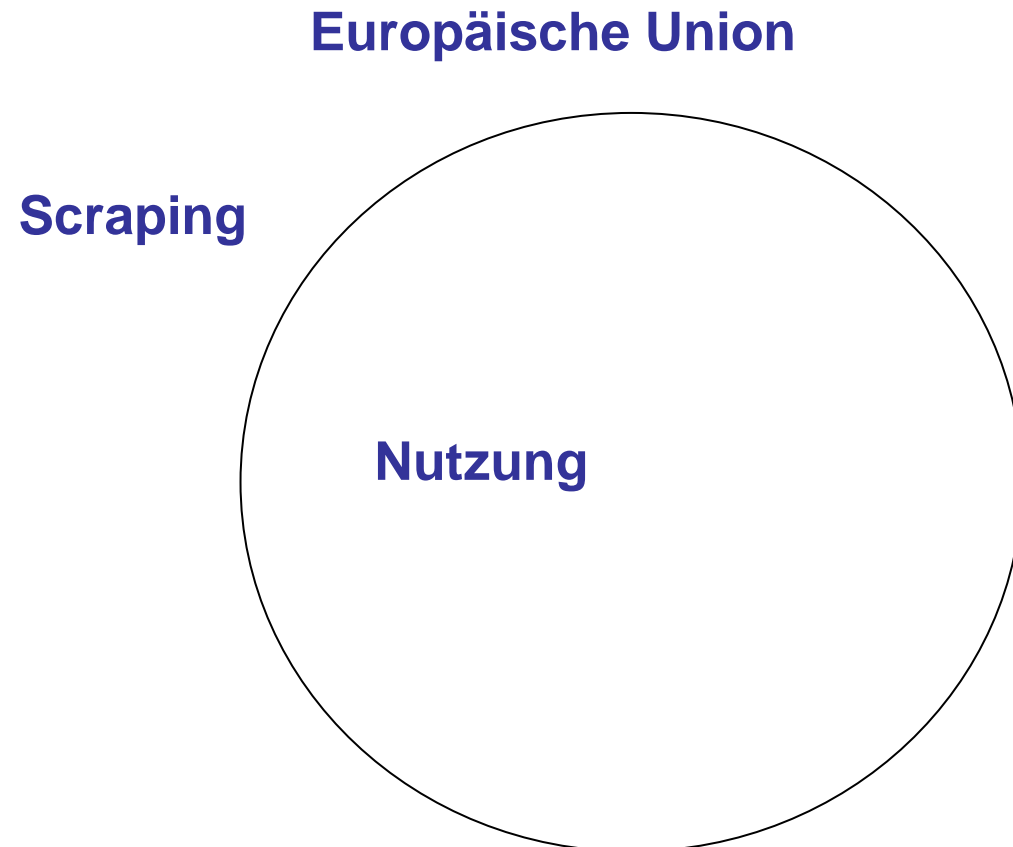
Verantwortlicher



Art. 3 Abs. 2 DSGVO:

- Angebot an EU-Bürger?
- Verhaltensbeobachtung?

Ist der räumliche Anwendungsbereich eröffnet?



Problem:

- Trainingsdatensammlung außerhalb der EU
- Nutzung des trainierten KI-Systems innerhalb der EU

Rechtsgrundlage für die Trainingsdatensammlung nach Art. 6 DSGVO

Art. 6 Abs. 1 lit. f DSGVO: berechtigtes Interesse

- Abgrenzung zu anderen Rechtsgrundlagen der DSGVO
- Interessenabwägung: Wirtschaftliche Interessen vs. informationelle Selbstbestimmung
- Erforderlichkeit der Verarbeitung

Rechtsgrundlage für die Trainingsdatensammlung nach Art. 6 DSGVO

Problemkreise

- **Problem 1:** Sensitive Datenkategorien i.S.v. Art. 9 DSGVO
- **Problem 2:** Widerspruch i.S.v. Art. 21 DSGVO
- **Problem 3:** Informationspflichten

IV. Risiken und praktische Implikationen für die Akteure

Risiken und praktische Implikationen für die Akteure

- **Betroffene Personen**
 - gesteigertes Bewusstsein für Verwendungsmöglichkeiten öffentlich verfügbarer Daten
- **Webseitenbetreiber**
 - Risiko: Beklagte in DSGVO Massenverfahren
 - Technische und rechtliche Gestaltung der Webseite
- **Scraper**
 - Gesteigertes Risiko für DSGVO Massenverfahren durch KI-Verordnung?
- **KI-Anbieter**
 - Risiko: Risiko für DSGVO Massenverfahren → noch keine typische Konstellation
 - To-Do: Entwicklung eines Datenschutzkonzepts für die Sammlung/Bezug von Trainingsdaten

V. Zusammenfassung und Ausblick

Zusammenfassung

Was folgt aus der Bewertung der datenschutzrechtlichen Zulässigkeit von Scraping zur Sammlung von Trainingsdaten?

- Ob Scraping zur Sammlung von Trainingsdaten nach DSGVO zulässig ist oder nicht, hängt von vielerlei Faktoren, insbesondere aber immer vom jeweiligen Einzelfall ab.
- Mit der zunehmenden Relevanz von KI steigt auch die Relevanz der Frage der rechtlichen Zulässigkeit der Trainingsdatenbeschaffung.
- Scraping rückt dadurch potentiell noch stärker in den Fokus von DSGVO-Massenverfahren.
- Lösung und Risikominimierung: Technische und rechtliche Ausgestaltung.



Vielen Dank!