www.richterzeitung.ch

Daniel Kettiger / Andreas Lienhard

Datengrundlagen der Justizforschung in der Schweiz: von der Statistik und empirischen Forschung zu «Data Mining»

Justizforschung benötigt – wie Forschung ganz allgemein – Daten. Dieser Beitrag befasst sich mit der Gewinnung von Datengrundlagen zur Beforschung der Justiz in der Schweiz, sowohl mittels herkömmlicher Methoden wie auch mittels Nutzung der neuen technologischen Möglichkeiten. Dabei werden auch Grenzen und Schranken der Datenbeschaffung und Datenbearbeitung aufgezeigt.

Beitragsart: Science

Zitiervorschlag: Daniel Kettiger / Andreas Lienhard, Datengrundlagen der Justizforschung in der Schweiz: von der Statistik und empirischen Forschung zu «Data Mining», in: «Justice - Justiz - Giustizia» 2025/4

ISSN 1661-2981, richterzeitung.weblaw.ch, Weblaw AG, info@weblaw.ch, T +41 31 380 57 77

EDITIONS WEBLAW

Inhaltsübersicht

1. Einleitung
2. Herkömmliche Datengrundlagen
   2.1. Statistiken
      2.1.1. Historisches
      2.1.2. Uneinheitlichkeit der Gerichtsstatistiken
      2.1.3. Bestrebungen zur Harmonisierung
   2.2. Empirische Sozialforschung
      2.2.1. Quantitative Erhebungen
      2.2.2. Qualitative Erhebungen
      2.2.3. Problem der kleinen Grundgesamtheiten und Stichproben
   2.3. Auftragsarbeiten als Quellen für die Wissenschaft?
3. Methoden der Informationsgewinnung mittels «Data Mining»
   3.1. Grundlage: Datencorpus
   3.2. Auswertungen mittels «klassischem» Abgleichen
   3.3. Auswertungen mittels Künstlicher Intelligenz
   3.4. Herausforderungen und Grenzen
      3.4.1. Technische Hürden für das Scraping
      3.4.2. Negation als sprachliche Herausforderung
      3.4.3. Inhärente Ungenauigkeit von LLMs
      3.4.4. Ethische Fragen: Bias in LLMs, Replizierbarkeit und Datenschutz
      3.4.5. Rechtliche Schranken: Analyseverbote und in Services «verbotene» Themen
4. Ausblick

1. Einleitung

[1] Justizforschung ist in der Schweiz – wie in ganz Europa[1] – eine relativ junge wissenschaftliche Disziplin.[2] Dementsprechend relativ gering sind die Erfahrungen mit den Datengrundlagen bzw. mit der Datenerhebung für die Justizforschung. Nachfolgend soll – teilweise ausgehend von den Forschungserfahrungen der Autoren – ein Überblick über die Entwicklung der Datenerhebung für die Justizforschung in der Schweiz und über die sich stellenden Herausforderungen gegeben werden.

[2] Bis vor wenigen Jahren boten sich in der Justizforschung – wie ansonsten in der verwaltungs- und sozialwissenschaftlichen Forschung, sieht man von elektronischer Auswertung selbsterhobener Daten bzw. automatisierten Auswertungen in Befragungstools ab – weitgehend nur die herkömmlichen Methoden der Datenerhebung an (siehe nachfolgend Ziff. 2). Methoden der Informationsgewinnung mittels «Data Mining»[3] (siehe nachfolgend Ziff. 3) sind ziemlich neu und stellen auch neue Herausforderungen (siehe Ziff. 4). Diese neuen Forschungsmöglichkeiten sind – entsprechend der technologischen Entwicklung – raschen Veränderungen unterworfen, so dass vorgenommene Feststellungen und Wertungen schon wenige Monate nach der Publikation dieses

1 Siehe dazu Daniel Kettiger/Andreas Lienhard/Philip Langbroek/Marco Fabri, Court Management – A Young Field of Public Management, in: Edoardo Ongaro (Hrsg.), Public Administration in Europe, Governance and Public Management, Cham 2019.

2 Zur Geschichte und Entwicklung der Justizforschung in der Schweiz ausführlich Andreas Lienhard/Daniel Kettiger, 20 Jahre Justizforschung in der Schweiz, «Justice – Justiz – Giustizia» 2025/2.

3 Der Begriff von «Data Mining» wird hier weit verstanden, etwa in folgendem Sinne: Ein Sammelbegriff für Verfahren und Prozesse welche das systematische und (halb-) automatische Entdecken und Strukturieren unbekannter Informationen aus grossen Mengen von Daten ermöglichen.

Beitrags veraltet sein können. In diesem Sinne ist der vorliegende Beitrag als Momentaufnahme zu betrachten.

2. Herkömmliche Datengrundlagen

2.1. Statistiken

2.1.1. Historisches

[3] Seit dem 19. Jahrhundert informieren die obersten kantonalen richterlichen Instanzen in der Schweiz meist jährlich oder zweijährlich die kantonalen Parlamente über die Entwicklung der Rechtspflege.[4] Auch wenn der primäre Zweck dieser Rechenschaftsberichte – wie der Name bereits sagt – darin bestand und weiterhin besteht, Rechenschaft gegenüber dem Parlament abzulegen, leisteten diese tatsächlich viel mehr. Sie sind als jener Ort zu betrachten, an dem sich die Wissensproduktion über das Gerichtswesen vollzog – und immer noch vollzieht. Die Rechenschaftsberichte sind als bedeutsame historische Quelle für die Erforschung der Geschichte und Entwicklung der kantonalen Justiz zu betrachten. Dabei fand und findet man im statistischen Teil der Geschäftsberichte jeweils viele Tabellen und Zahlen vor, die etwa Auskunft geben über Art, Menge, Behandlung und Erledigung der Geschäfte der verschiedenen Gerichtsinstanzen und Justizbehörden. Augenfällig ist, dass im Laufe der Zeit die Datenerhebungen zunehmend erweitert und ausdifferenziert wurden. Allein der Umfang der erhobenen Daten und deren Darstellungsart weisen darauf hin, welches Gewicht der Statistik und der Tabelle, welche in jüngerer Zeit ausserdem vereinzelt durch grafische Darstellungen wie Säulen-, Kreis- oder Kurvendiagramme ergänzt wurden, zugemessen wurde.

[4] In den öffentlich zugänglichen Statistiken von Gerichten und anderen Justizbehörden (z.B. Staatsanwaltschaften) findet sich somit eine erhebliche Menge von Zahlenmaterial, das grundsätzlich für die Justizforschung nutzbar gemacht werden kann. Allerdings leiden diese statistischen Quellen unter dem Umstand von Änderungen im Zeitverlauf sowie der unterschiedlichen Datenaggregation, so dass statistische Zeitreihen und damit Vergleiche im Zeitablauf aus den folgenden Gründen immer nur über eine bestimmte Zeitsequenz möglich sind:

- *Inhaltliche Änderungen der Geschäftsberichte:* Geschäftsberichte der Gerichtsbehörden sind Gegenstand von Anpassungen an die Bedürfnisse des Parlaments oder des Justizmanagements; damit können sich auch die statistischen Darstellungen ändern.
- *Reformen der Rechtsgrundlagen:* Durch Änderungen des materiellen Rechts (z.B. Strafbestimmungen), des Prozessrechts (z.B. Strafprozessrecht) oder des Justizorganisationsrechts (z.B. Zuständigkeiten, Gebietsreformen im Justizwesen) müssen zwangsläufig auch die Statistiken angepasst werden. Dies kann zu Veränderungen in den Erledigungskategorien führen.

4 Der Text von Ziffer 2.1.1 stammt – stark zusammengefasst – weitgehend aus Stephan Aerschmann/Christof Schwenkel/Stefan Rieder/Michele Luminati, Umfeld (Teilprojekt 1), in: Andreas Lienhard/Daniel Kettiger (Hrsg.), Justiz zwischen Management und Rechtsstaat, Schriftenreihe für Justizforschung Band 5, Bern, 2016, S. 36–38; ausführlich zum historischen Aspekt von Gerichtsstatistiken Stephan Aerschmann, Von der Macht der Zahlen, Schriftenreihe für Justizforschung Band 14, Bern 2017.

2.1.2. Uneinheitlichkeit der Gerichtsstatistiken

[5] Die Schweiz verfügt heute noch über keine einheitlichen kantonalen Gerichtsstatistiken – trotz einheitlichem materiellem Recht und Verfahrensrecht auch nicht in Zivil- und Strafsachen. Dies ist eine Folge der föderalistisch ausgestalteten Rechtspflegeorganisation erster und oberer Instanz in Zivil- und Strafsachen sowie in der kantonalen Verwaltungsrechtspflege. Eine Harmonisierung wäre teilweise denkbar, weil es auch in der Zivil- und Strafrechtspflege möglich ist, in bestimmten Bereichen einheitliche Fallkategorien bzw. Erledigungsarten zu definieren, die von der Organisation der Justizbehörden unabhängig sind. Es bestehen aber auch einige strukturelle Grenzen. Ein Beispiel sind die Handelsgerichte (Art. 6 ZPO[5]), die in der kantonalen Organisation der ersten Instanz in Zivilsachen fakultativ sind (Art. 6 Abs. 1 ZPO: «Die Kantone können ... »). Mit der Schaffung eines Handelsgerichts als dritte Gerichtsbehörde erster Instanz in Zivilsachen (neben dem ordentlichen Zivilgericht einerseits und der einzigen kantonalen Instanz nach Art. 5 ZPO sowie der Möglichkeit der direkten Klage beim oberen Gericht nach Art. 8 ZPO andererseits) verändern sich die Zuständigkeiten in Handelssachen und in gesellschaftsrechtlichen Fällen derart, dass ein einfacher Vergleich der Gerichtsstatistik in diesem Bereich mit der Statistik eines Kantons ohne Handelsgericht ohne besonderen Aufwand kaum möglich ist.

2.1.3. Bestrebungen zur Harmonisierung

[6] Die wesentlichen Bestrebungen zur Harmonisierung der Gerichtsstatistiken und anderer Daten über die Gerichte in der Schweiz ging von der European Commission for the Efficiency in Justice (CEPEJ[6]), einer Fachkommission des Europarats aus. Im Rahmen ihres Gesamtauftrags für eine effiziente Justiz wirkte die CEPEJ auch hinsichtlich einer harmonisierten Darstellung der Organisation und Aktivitäten der Justizbehörden in allen Mitgliedstaaten des Europarats.[7] Es obliegt den Korrespondentinnen und Korrespondenten der Kommission in den Mitgliedstaaten, die statistischen Daten ihrer Länder zu sammeln, nach den einheitlichen Vorschriften der CEPEJ aufzubereiten und an das Sekretariat der Kommission zu liefern – so auch dem Korrespondenten für die Schweiz.[8] Die CEPEJ veröffentlicht dann die aufbereiteten Daten im Internet und verwendet die gesammelten Daten nicht nur zu ihrem eigenen Gebrauch, sondern stellt auf Anfrage auch die Rohdaten für wissenschaftliche Zwecke zur Verfügung.[9]

[7] Die Aktivitäten für die CEPEJ in der Schweiz[10] fanden ihre Entsprechung und Vertiefung in den Aktivitäten des Vereins ESEHA-ESSSA.[11] ESEHA-ESSSA entwickelt und stellt *CHstat*, die Datenbank der Schweizer Kantone und Städte («Föderalismus in Zahlen») zur Verfügung. Diese

---

5 Schweizerische Zivilprozessordnung (Zivilprozessordnung, ZPO) vom 19. Dezember 2008, SR 272.

6 Das Kürzel steht für die französische Bezeichnung der Kommission: Commission européenne pour l'efficacité de la justice.

7 Siehe https://www.coe.int/en/web/cepej/cepej-stat (alle Weblinks in diesem Beitrag zuletzt besucht am 2. Dezember 2025).

8 Diese Funktion wird seit Jahren durch Dr. Jacques Bühler wahrgenommen.

9 Siehe das «Protocol for the transmission and use of CEPEJ data», https://rm.coe.int/cepej-gt-eval-2025-15rev-data-transmission-protocol-en/488028b4e1.

10 Vgl. zu den Aktivitäten der CEPEJ und den entsprechenden Entwicklungen in der Schweiz auch den Beitrag von Christophe Koller, La CEPEJ et le développement des statistiques judiciaires suisses : premiers résultats de «La Justice en chiffres», in dieser Ausgabe von «Justice – Justiz – Giustizia».

11 https://www.chstat.ch/de/2153/Das%20Portal%20CHstat.

enthält *sJustice*, das Portal der Justizstatistiken mit 1500 Variablen und Indikatoren zur Organisation der Gerichte und Staatsanwaltschaften, zu den Akteuren der Justiz (Richter/innen und Staatsanwält/innen, Personal, Weiterbildung, Sanktionen) sowie zu den Aktivitäten nach Bereich und Instanz.[12] Weiter erstellt ESEHA-ESSSA derzeit periodisch einen *eJustice Barometer*.[13]

2.2. Empirische Sozialforschung

2.2.1. Quantitative Erhebungen

[8] Für spezifische Studien der Justizforschung war die Wissenschaft schon immer auf quantitative Methoden der Datenerhebung angewiesen. Ein Beispiel für quantitative Erhebungen sind die Erhebungen, die für Geschäftslaststudien notwendig sind.[14]

[9] Es wird vorliegend nicht weiter auf die einzelnen Methoden quantitativer Erhebungen[15] eingegangen, da diese als bekannt vorausgesetzt werden können[16] und empirische Sozialforschung als solche nicht das Hauptthema dieses Beitrags ist.

2.2.2. Qualitative Erhebungen

[10] Je nach Forschungsfrage sind in der Justizforschung auch qualitative Erhebungen erforderlich, oft in der Form von teilstrukturierten Interviews. Es sind auch andere Formen qualitativer Forschung denkbar, beispielsweise in der Form von Expertenworkshops.[17]

[11] Es wird vorliegend nicht weiter auf die einzelnen Methoden qualitativer Erhebungen eingegangen, da diese als bekannt vorausgesetzt werden können[18] und empirische Sozialforschung als solche nicht das Hauptthema dieses Beitrags ist.

2.2.3. Problem der kleinen Grundgesamtheiten und Stichproben

[12] Die Gerichtslandschaft in der Schweiz ist wegen der föderalen Strukturen sehr kleinräumig und kleinteilig. Die Gerichte und teilweise auch die Staatsanwaltschaften sind im internationalen Vergleich eher klein und weisen demzufolge oft nur wenige Personen mit ein- und derselben

---

12 https://www.chstat.ch/de/data-theme/1201/2-Justiz.

13 https://www.chstat.ch/de/data-theme/1400/eJustice.

14 Siehe dazu in Kürze Andreas Lienhard/Daniel Kettiger, Geschäftslaststudien: Wie geht man mit kleinen Gerichten um?, «Justice – Justiz – Giustizia» 2023/3, Rz. 2–7; ausführlich Daniela Winkler, Fallgewichtung an schweizerischen Gerichten: Methodik von Studien der gewichteten Geschäftslast, Schriftenreihe zur Justizforschung Band 18, Bern 2020.

15 Siehe zum Thema quantitativer Studien in den Rechtswissenschaften und der Rechtswirkungsforschung auch Tilmann Altwicker/Yoan Hermstrüwer, Quantitative Legal Studies: The Science of Measuring the Law, in: Tilmann Altwicker et al. (Hrsg.), Recht – Jurisprudenz – Wissenschaft, Festgabe aus Anlass des Schweizerischen Juristentags 2025 in Zürich, S. 41 ff.; Corinna Coupette/Andreas Martin Fleckner, Quantitative Rechtswissenschaft, Sammlung, Analyse und Kommunikation juristischer Daten, JuristenZeitung (JZ) 8/2018, 379–389.

16 Siehe jüngst etwa Lee Epstein, Making Empirical Legal Research More Relevant, in: Tilmann Altwicker et al. (Hrsg.), Recht – Jurisprudenz – Wissenschaft, Festgabe aus Anlass des Schweizerischen Juristentags 2025 in Zürich, S. 1 ff.

17 Als Beispiel siehe Daniel Kettiger/Andreas Lienhard/Thomas Myrach/Matthias Stürmer/Magda Chodup/Joel Niklaus, Open Justice vs. Privacy: Ergebnisse und Empfehlungen, «Justice – Justiz – Giustizia» 2025/1, Rz. 33 und 36.

18 Siehe etwa Epstein (Fn. 16), S. 1 ff.

Funktion auf. Für bestimmte Kategorien der Fallerledigung liegen zudem in einer bestimmten Zeiteinheit oft nur sehr wenige Fälle vor. Dies führt insbesondere bei quantitativen Erhebungen zu eher kleinen Grundgesamtheiten und kleinen Stichproben, was die Repräsentativität von Studien infrage stellen kann.

[13] Die Problematik sei am Beispiel von Geschäftslaststudien dargestellt:[19] Bei Geschäftslaststudien, bei denen das Fallgewicht jeder Fallkategorie nach der Methode der Ermittlung/Berechnung der genauen Werte (z.B. durch Selbstaufschreibung) ermittelt wird, findet das so genannte zentrale Grenzwerttheorem, eine Gesetzmässigkeit der statistischen Mathematik, Anwendung, welches besagt, dass die Mittelwerteverteilung für beliebige Verteilungsformen des Merkmals in der Population bereits dann hinreichend normal ist, wenn «n» (Anzahl der Fälle) 30 ist.[20] Im Sinne einer vorsichtigen Interpretation werden deshalb nur ermittelte Geschäftslastwerte als statistisch valide bezeichnet, bei welchen der auf eine bestimmte Funktion bezogene Geschäftslastwert einer Fallkategorie auf der Auswertung von mindestens 30 Fällen beruht, die während der Erhebungsperiode eröffnet und abgeschlossen wurden. Bei kleinen Gerichten oder Jurisdiktionen weisen oft nicht alle Fallkategorien pro Jahr 30 solcher Fälle auf. Selbst beim erstinstanzlichen Strafgericht des Kantons Basel-Stadt mussten die Autoren feststellen, dass bei dem ursprünglich gewählten feingliederigen und detaillierten System von Fallkategorien bei einem Jahr Erhebungszeit nicht bei allen Fallkategorien mindestens 30 ausgewertete Fälle vorlagen.[21]

[14] Aber auch bei qualitativen Erhebungen kann die kleine Grundgesamtheit von betroffenen Personen bzw. möglichen Interviewpartnern ein Problem sein. Möchte man eine Studie zur Frage der Relevanz und Geeignetheit von Rätoromanisch als Gerichtssprache machen, so stünden wohl nur knapp 20 Gerichtspersonen zur Verfügung, die dieser Sprache im beruflichen Kontext mächtig sind und über entsprechende berufliche Erfahrungen verfügen.

2.3. Auftragsarbeiten als Quellen für die Wissenschaft?

[15] Untersuchungen in der schweizerischen Justiz wurden und werden oft nicht als wissenschaftliche Forschungsprojekte im engeren Sinn, d.h. beispielsweise im Rahmen von Förderprojekten oder von Dissertationen initialisiert, sondern haben ihren Ursprung in Aufträgen der betroffenen Justizbehörden selber oder der betreffenden Aufsichtsorgane. So wurden alle bisherigen Studien zur Ermittlung einer gewichteten Geschäftslast und auch weitere Geschäftslaststudien als Auftragsarbeiten durchgeführt.[22] Auch Organisationsüberprüfungen von Justizbehörden können al-

19 Vgl. Lienhard/Kettiger (Fn. 14), Rz. 10.

20 Vgl. Jürgen Bortz, Statistik für Sozialwissenschaftler, 5. Aufl., Berlin, 1999, S. 93 f.; Sheldon M. Ross, Introductory Statistics', Section 7.4. Academic Press, 2017. Siehe auch https://www.investopedia.com/terms/c/central_limit_-theorem.asp: «According to the central limit theorem, the mean of a sample of data will be closer to the mean of the overall population in question, as the sample size increases, notwithstanding the actual distribution of the data. In other words, the data is accurate whether the distribution is normal or aberrant. As a general rule, sample sizes of around 30–50 are deemed sufficient for the CLT to hold, meaning that the distribution of the sample means is fairly normally distributed. Therefore, the more samples one takes, the more the graphed results take the shape of a normal distribution. Note, however, that the central limit theorem will still be approximated in many cases for much smaller sample sizes, such as n=8 or n=5».

21 Vgl. Andreas Lienhard/Daniel Kettiger/Hans Peter Uster/Daniela Winkler, Geschäftslast sowie Aufbau- und Ablauforganisation der Gerichte und der Staatsanwaltschaft im Kanton Basel-Stadt: Schlussbericht vom 12. Februar 2015, S. 33 und Anhang 4 S. 72 f.; https://www.kettiger.ch/fileadmin/user_upload/Dokumente/Downloads/BS_Justiz_Schlussbericht_def_150212.pdf.

22 Vgl. Übersicht bei Andreas Lienhard/Daniel Kettiger, (Fn. 2), Rz. 23.

lenfalls Daten und Informationen enthalten, welche für andere Forschungsvorhaben interessant sein oder als Ausgangspunkt dienen können, dies insbesondere dann, wenn die Berichte zumindest teilweise nach wissenschaftlichen Kriterien gegliedert sind.[23]

[16] Die meisten Ergebnisse solcher Auftragsarbeiten werden nicht veröffentlicht. Auf der Grundlage des heute in fast allen Kantonen geltenden Öffentlichkeitsprinzips mit einem allgemeinen Recht auf Zugang zu Verwaltungsdokumenten, können sich aber interessierte Wissenschaftlerinnen und Wissenschaftler Zugang zu Studien beschaffen, welchen sie Daten entnehmen möchten.

3. Methoden der Informationsgewinnung mittels «Data Mining»

3.1. Grundlage: Datencorpus

[17] Ausgehend davon, dass «Data Mining» Verfahren und Prozesse umfasst, welche das systematische und (halb-) automatische Entdecken und Strukturieren unbekannter Informationen aus grossen Mengen von Daten ermöglichen,[24] wird als Grundlage stets ein Datencorpus mit einer grossen Zahl von Urteilen bzw. Entscheiden bestimmter Gerichte und/oder anderer Justizbehörden benötigt. Einen solchen Urteils-Korpus kann man von der entsprechenden Urteilsdatenbank bzw. den entsprechenden Urteilsdatenbanken im Internet selber erstellen. Dies erfolgt durch Web-Scraping (auch als Data/Web Harvesting bezeichnet),[25] das in einem automatisierten Auslesen von Daten aus Webseiten durch sog. «Web Crawler» besteht.[26] Wenn ein Datenkorpus mit im Internet veröffentlichten schweizerischen Urteilen erwünscht ist, so kann dieser auch beim Verein entscheidsuche.ch[27] bezogen werden (z.B. über eine API). Auf letztere Weise wurde beispielsweise der Datencorpus für ein Experiment zur De-Anonymisierung von Urteilen erstellt.[28]

[18] Oft werden dann für die Forschungsarbeiten weitere Datencorpora bzw. Datenbanken für einen Abgleich mit Zwischenergebnissen der Untersuchungen des Datencorpus der Urteile benötigt.

3.2. Auswertungen mittels «klassischem» Abgleichen

[19] Als «klassischer» Abgleich werden hier alle Methoden des «Data Mining» bezeichnet, die nicht auf der Technologie von so genannter künstlicher Intelligenz (KI) beruhen. Es können hier etwa folgende Beispiele genannt werden:

---

23 Vgl. etwa Lienhard/Kettiger/Uster/Winkler (Fn. 21), S. 37–40, insb. Ziff. 2.2.1.3, S. 38 f.; Daniel Kettiger, Organisationsüberprüfung des Kantonsgerichts Nidwalden Ergebnisse der Untersuchung, Bericht zu Handen der Justizkommission des Landrats, Version 4–0 vom 30. November 2025, S. 4 f.; https://www.nw.ch/_docn/402142/Bericht_Organisations%C3%BCberpr%C3%BCfung_Kantonsgericht.pdf.

24 Vgl. oben Fn. 3.

25 Zum Web-Scraping siehe Leonie Grob, Datenschutzrechtliche Aspekte bei der Veröentlichung von Gerichtsurteilen in sekundären Entscheid-Datenbanken, «Justice – Justiz – Giustizia» 2024/2, Rz. 17, mit zahlreichen Hinweisen.

26 Dieses Vorgehen wird beispielsweise detailliert in der folgenden Publikation beschrieben: Kerstin Noëlle Vokinger/Urs Jakob Mühlematter, Re-Identifikation von Gerichtsurteilen durch «Linkage» von Daten(banken), Jusletter 2. September 2019, Rz. 30.

27 https://entscheidsuche.ch/.

28 Vgl. Joel Niklaus/Magda Chodup/Thomas Lüthi/Daniel Kettiger, Re-Identifizierung in Gerichtsurteilen mit Simap Daten, Version v3, veröffentlicht am 22. Oktober 2023, https://zenodo.org/records/10032009, Rz. 12.

- *Empirische Analyse anhand von Bundesgerichtsbeschwerden gegen Preisfestsetzungs-Verfügungen von Arzneimitteln:*[29] In der Studie wurde untersucht, ob es mit der Methodik des «Linkage» – der Verbindung von verschiedenen, öffentlich zugänglichen Daten(banken) – möglich ist, Urteile zu re-identifizieren. In einem ersten Schritt wurden die im Datencorpus vorhandenen Bundesgerichtsurteile automatisch (entweder direkt über die Referenzangabe in der Regeste bzw. im Urteilstext oder mittels eines Cosinus-Ähnlichkeit-Algorithmus) mit dem entsprechenden Vorentscheid des Bundesverwaltungsgerichts verlinkt.[30] In einem zweiten Schritt wurde versucht, Informationen aus den jeweiligen Bundesgerichtsurteilen und/oder den korrespondierenden erstinstanzlichen Urteilen des Bundesverwaltungsgerichts mittels Datenbank-Suchabfragen mit möglichst spezifischen Informationen zu verlinken («Linkage»), um die Re-Identifikation zu ermöglichen, d.h. den Wirkstoff bzw. das Arzneimittel und die Zulassungsinhaberin zu bestimmen.[31] Materiell interessierten die Fragestellungen, welche pharmazeutischen Unternehmen zwischen 2000 und 2018 in einem Verfahren gegen (Preis-)Verfügungen des BAG vor Bundesgericht involviert und welche Arzneimittel davon betroffen waren. Es wurde eine Re-Identifikation in 84% der Fälle erreicht.
- *Re-Identifizierung in Gerichtsurteilen mit Simap-Daten*[32]:[33] Der Datencorpus mit schweizerischen Gerichtsurteilen (insbesondere auch Urteilen des Bundesverwaltungsgerichts und des Bundesgerichts) wurde nach dem Begriff «simap» durchsucht (case-insensitive: Gross- und Kleinschreibung wird nicht berücksichtigt). Darauf wurden diese Urteile mit regulären Ausdrücken[34] nach den Meldungsnummern und Projekt-Identifikatoren (fortan «Nummern») durchsucht. Neben der selbst erstellten Datenbank wurde die Plattform IntelliProcure[35] als Hilfsmittel zur Re-Identifikation verwendet. IntelliProcure lädt regelmässig die Daten der Simap Plattform herunter und bereitet diese auf. Im dritten Schritt der Analyse wurden nun die Daten des Exports nach den Nummern durchsucht, welche in den Gerichtsentscheiden gefunden wurden. Die Untersuchung hat zweierlei gezeigt: Zum einen war es durch das angewandte Verfahren des String-Matchings[36] nicht möglich, die Beschwerdeführer zu re-identifizieren; zum anderen lässt sich mittels der angewandten Methode aber in einer Vielzahl der Fälle zuordnen, bei welchen Submissionen der Rechtsweg beschritten worden ist. Die Veröffentlichung dieser Information ist gesetzlich nicht vorgesehen.
- *Bauen ausserhalb der Bauzonen: Wirksame Vollzugskontrolle durch den Bund:*[37] Es ging um die Fragestellung, ob die Beschwerdemöglichkeit von Bundesbehörden an das Bundesgericht in Fragen des Bauens ausserhalb von Bauzonen wirksam sei. Der Datencorpus mit Urteilen aus

29 Vgl. Vokinger/Mühlematter (Fn. 26).

30 Vgl. Vokinger/Mühlematter (Fn. 26), Rz. 31.

31 Vgl. Vokinger/Mühlematter (Fn. 26), Rz. 31.

32 SIMAP-Daten sind Ausschreibungsdaten und Unterlagen, die auf der Plattform simap.ch veröffentlicht werden und das öffentliche Beschaffungswesen in der Schweiz betreffen.

33 Vgl. Niklaus/Chodup/Lüthi/Kettiger (Fn. 28).

34 Reguläre Ausdrücke: Eine reguläre Ausdrucksweise ist eine Schreibweise, die verwendet wird, um Muster in Zeichenketten zu finden und zu manipulieren.

35 https://intelliprocure.ch/.

36 Dieser Begriff beschreibt die Aufgabe, eine Teilzeichenkette innerhalb einer grösseren Zeichenkette zu identifizieren (also beispielsweise «Haus» in Bau*haus*, Berg*haus* usw).

37 Unveröffentlichte Studie unter Mitwirkung des Erstautors, Publikation in Vorbereitung.

entscheidsuche.ch wurde mit Hilfe von Python Parsers[38] und Regulären Ausdrücken aufbereitet. Im Rubrum des Urteils wurden mit Hilfe von Regulären Ausdrücken die Namen der Parteien und deren gerichtlicher Vertretungen extrahiert. Der Entscheid wurde dann weiter durchsucht nach Schlüsselwörtern, welche Aufschluss geben über die Art des Hauptentscheids. Ursprünglich wurden die folgenden sieben Arten unterschieden: Gutheissung, teilweise Gutheissung, Abweisung, teilweise Abweisung, Nicht-Eintreten, Abschreibung und Vereinigung. Für die vorliegende Analyse wurden der Einfachheit halber diese sieben Entscheidarten auf zwei reduziert (Gutheissung und Abweisung). Die drei letzteren Entscheidarten Nicht-Eintreten, Abschreibung und Vereinigung wurden ganz entfernt. Die teilweise Gutheissung resp. Abweisung wurde – je als Unterkategorie – in eine Gutheissung resp. Abweisung umgewandelt. Urteile, bei welchen sowohl eine Abweisung als auch eine Gutheissung gefunden wurde, wurden nicht berücksichtigt und entfernt. In einem weiteren Schritt wurden die Erwägungen der Urteile des Bundesgerichts gefiltert nach dem Vorhandensein von Art. 24b, 24c, 24dund 24e RPG.[39] In einem letzten Schritt wurde das so erhaltene Datenset unterteilt in vier Sub-Datensets (Bundesamt für Raumentwicklung ARE ist Beschwerdeführer, ARE ist nicht Beschwerdeführer, ARE ist Beschwerdegegner und ARE ist nicht Beschwerdegegner), abhängig davon, ob das ARE als Beschwerdeführer resp. Beschwerdegegner involviert war.

- *Empirische Analyse der bundesgerichtlichen Praxis der Anonymisierung von Urteilen*[40] Die Anonymisierung von Gerichtsurteilen steht insbesondere im Spannungsfeld zwischen dem Grundsatz der Justizöffentlichkeit und dem Persönlichkeitsschutz der Parteien. Anhand zweier Kohorten, bei denen sich das Verhältnis von öffentlichem Interesse und Persönlichkeitsschutz unterscheiden (Beschwerden gegen die Preisfestsetzung von Arzneimitteln und gegen Strafurteile betreffend Vergewaltigung), wurde die Anonymisierungspraxis des Bundesgerichts empirisch analysiert. Dabei zeigte sich, dass die Anonymisierung im Verlauf der Zeit zugenommen hat, unerwartet stärker bei der Kohorte «Arzneimittel» als bei der Kohorte «Vergewaltigung». Die Methode ist mehrschrittig und mit manueller Arbeit verbunden. Die Arbeiten basieren hier ausschliesslich auf dem Urteilscorpus.

3.3. Auswertungen mittels Künstlicher Intelligenz

[20] Einen Datencorpus von Urteilen – allenfalls zusammen mit weiteren Datenbanken oder Inhalten des Internets oder eines Intranets – kann man heute auch mittels Large Language Models (LLM) untersuchen und auswerten. LLMs kann man dabei – je nach Aufgabenstellung und Fachbereich – in der Form der angebotenen Chat-Versionen (z.B. ChatGPT), spezifisch, d.h. mittels Computertechnik eingebunden (z.B. GPT 4.0 via API) oder unter Mithilfe von sogenannten KI-Assistenten einsetzen. Es können hier etwa folgende Beispiele genannt werden (auf die Methoden wird nicht detailliert eingegangen, weil diese sehr komplex sind):

38 Python ist eine universelle, üblicherweise interpretierte, höhere Programmiersprache. Ein Parser bezeichnet in der Informatik ein Computerprogramm zur Übersetzung eines Inputs in ein geeignetes Format zur Weiterverarbeitung. Dabei zerlegt es zunächst die Informationen und gibt sie dann in einem geordneten System wieder aus.

39 Bundesgesetz über die Raumplanung (Raumplanungsgesetz, RPG) vom 22. Juni 1979, SR 700.

40 Vgl. Kerstin Noëlle Vokinger/David Schneider, Anonymisierung von Gerichtsurteilen, in: Jusletter 13. Dezember 2021.

- *Erkennen von Rechtsverstössen in unstrukturierten Texten:*[41] Die Verletzung von Rechtsregeln hinterlässt heute oft Datenspuren (im Internet oder anderswo). Diese Studie – ein Experiment im Bereich der Computerwissenschaften – konzentriert sich auf zwei Hauptaufgaben: erstens die Erkennung von Verstössen gegen das Recht[42] in grossen Mengen von unstrukturierten Textdaten[43] und zweitens die Zuordnung dieser Verstösse zu potenziell betroffenen Personen. Ein derartiges Aufdecken von Rechtsverstössen könnte insbesondere im Bereich von Sammelklagen interessant sein[44] – allenfalls auch in der Dunkelfeldforschung. Arbeiten im Bereich der Identifizierung von Rechtsverstössen konzentrieren sich meist auf domänenspezifische Themen, darunter Bereiche wie Compliance, Datenschutz und branchenspezifische Vorschriften.[45] Es wurden zwei Datensätze mithilfe grosser Sprachmodelle (LLMs) erstellt, die anschliessend von Fachexperten validiert wurden. Beide Aufgaben wurden speziell für den Kontext von Sammelklagen konzipiert. Das experimentelle Design umfasste das Feinabstimmen von Modellen der BERT-Familie und Open-Source-LLMs sowie die Durchführung von Few-Shot-Experimenten[46] mit proprietären LLMs. Die Ergebnisse mit einem F1-Score von 62,69% (Verfolgung von Rechtsverstössen) und 81,02% (Zuordnung von Opfern) zeigen, dass die verwendeten Datensätze und Setups für beide Aufgaben geeignet sind.
- *Identifizierung von Zusammenhängen zwischen den Umständen des Falles und dem Urteil:*[47] Der Kern der Inhaltsanalyse von Gerichtsentscheidungen in dieser Studie liegt in der systematischen Identifizierung von ansonsten nicht leicht erkennbaren Zusammenhängen zwischen den Umständen des Falles (Sachverhalt, Motive etc.) und dem Urteil. Dabei handelt es sich um einen innovativen Ansatz, der die Nutzung eines GPT-4-Sprachmodells der OpenAI-API zur Generierung der notwendigen Fakten aus unstrukturierten Textdokumenten von Gerichtsentscheidungen mit der Analyse von Assoziationsregeln kombiniert, um Muster in den Kriterien zu identifizieren, die das Gericht bei der Strafzumessung in ähnlichen Fällen berücksichtigt. Die Analyse basiert auf einer Sammlung von 10'000 Urteilen in Strafsachen aus dem Einheitlichen Register der Gerichtsentscheidungen der Ukraine. Häufige Item-Sets (Support 0,982) und starke Assoziationsregeln (Konfidenz = 0,987) wurden identifiziert. Es zeigte sich, dass die zu Freiheitsstrafen Verurteilten in den meisten Fällen Mittäter waren und/oder Vorstrafen hatten und/oder wiederholt Straftaten begangen hatten. Es zeigte sich, dass Straftäter, bei denen das Gericht milde Entscheidungen in Form von bedingten Verur-

41 Vgl. Dor Bernsohn/Gil Semo/Yaron Vazana/Gila Hayat/Ben Hagag/Joel Niklaus/Rohit Saha/Kyryl Truskovskyi, LegalLens: Leveraging LLMs for Legal Violation Identification in Unstructured Text, arXiv:2402.04335, https://arxiv.org/abs/2402.04335.

42 Es geht um deliktisches Handeln im Sinne des Straf-, Zivil- oder Verwaltungsrechts, das sich in bestimmten Wörtern oder Textelementen manifestiert (z.B. «sich entblössen», «Exhibitionismus», «sexuelle Belästigung»), die mit Hilfe der Technik der Named Entity Recognition (NER) im Text erkannt werden können, vgl. Bernsohn et al. (Fn. 41), S. 1.

43 Beispielsweise Datacorpora erstellt aus grossen Mengen von Zeitungsartikeln oder aus Posts in Social Medias, vgl. Bernsohn et al. (Fn. 41), S. 1, insb. auch Abbildung.

44 Vgl. Bernsohn et al. (Fn. 41), S. 1.

45 Vgl. Bernsohn et al. (Fn. 41), S. 1.

46 Few-Shot-Learning ist ein Framework für maschinelles Lernen, bei dem ein KI-Modell lernt, genaue Vorhersagen zu treffen, indem es mit einer sehr kleinen Anzahl von gekennzeichneten Beispielen trainiert wird.

47 Vgl. Olha Kovalchuk/Ruslan Shevchuk/Mariia Masonkov/Anhelina Banakh, Content Analysis of Court Decisions: A GPT-4 Based Sentence-by-Sentence Data Generation and Association Rules Mining, CEUR Workshop Proceedings, https://ceur-ws.org/Vol-3716/paper5.pdf.

teilungen oder vorzeitiger Entlassung getroffen hat, ein höheres Risiko für Rückfalltaten, insbesondere Beihilfe, aufweisen und eine grössere Gefahr für die Gesellschaft darstellen. Die gewonnenen Ergebnisse können das Verständnis der Hauptfaktoren bei der Strafzumessung verbessern und eine verlässliche Informationsgrundlage für juristische Entscheidungen liefern.

- *Textanalysen in Investitionsschiedsverfahren:*[48] Das Lesen ist die primäre Methodik, die Praktiker in Investitionsschiedsverfahren[49] nutzen, um sich mit den Schriftsätzen der Parteien, früheren Schiedssprüchen, Investitionsabkommen und allen anderen schriftlichen Materialien auseinanderzusetzen, die für ein Schiedsverfahren relevant sind. Doch ein alternativer und ergänzender Ansatz zur Analyse von Texten in Investitionsschiedsverfahren gewinnt an Bedeutung, indem Algorithmen anstelle von Menschen schriftliche Materialien analysieren, um relevante Erkenntnisse zu gewinnen. Was diesem algorithmischen Ansatz an Nuancen fehlt (Computer haben Schwierigkeiten, Texte zu verstehen), macht er durch seine Effizienz wett (Computer sind hervorragend im Umgang mit Zahlen). Textanalyse kann riesige Textkorpora in Sekundenschnelle verarbeiten und stellt Schiedsexperten neue, ergänzende Erkenntnisse zur Verfügung.

[21] Textanalyse zählt möglicherweise zu den vielversprechendsten Methoden, mit denen generative KI die sozialwissenschaftliche Forschung – und damit auch die Justizforschung – verbessern und erleichtern kann.[50] Dass LLMs die Genauigkeit menschlicher Expertise im Bereich der Kodierung von Texten noch nicht erreichen können, bedeutet, dass sie die herkömmliche Textanalyse nicht so schnell ersetzen werden. LLMs ermöglichen es Sozialwissenschaftlern jedoch, Korpora von beispielloser Grösse mit ungeahnter Geschwindigkeit zu untersuchen; anstatt beispielsweise eine zufällige Stichprobe von Dokumenten zu nehmen, können Sozialwissenschaftlerinnen und Sozialwissenschaftler nun potenziell den Text eines ganzes Korpus in kurzer Zeit kodieren.[51]

3.4. Herausforderungen und Grenzen

3.4.1. Technische Hürden für das Scraping

[22] Der Eidgenössische Datenschutz- und Öffentlichkeitsbeauftragte (EDÖB) hat im August 2023 mit neun anderen nationalen Datenschutzbehörden eine gemeinsame Erklärung (Joint State-

48 Vgl. Wolfgang Alschner/Damien Charlotin, Data Mining, Text Analytics, and Investor-State Arbitration (May 31, 2021). Forthcoming in: Pietro Ortolani et al. (Hrsg.): International Arbitration and Technology, Wolters Kluwer, Ottawa Faculty of Law Working Paper No. 2021–17, Available at SSRN https://dx.doi.org/10.2139/ssrn.3857127.

49 Ein Investitionsschiedsverfahren ist ein Verfahren zur Beilegung von Streitigkeiten zwischen einem ausländischen Investor und einem Gaststaat, das meist auf Grundlage internationaler Investitionsschutzabkommen oder Freihandelsabkommen mit Investitionsschutzklauseln beruht. Ziel ist es, dem Investor zu ermöglichen, seine Rechte vor einem internationalen Schiedsgericht durchzusetzen, anstatt ausschliesslich auf die nationalen Gerichte angewiesen zu sein.

50 Vgl. Christopher A. Bail, Can Generative AI improve social science? PNAS 2024 Vol. 121 No. 21 e2314021121, S. 5; https://doi.org/10.1073/pnas.2314021121.

51 Vgl. Bail (Fn. 50), S. 5.

ment[52]) an die Social Media Plattformen zum Schutz der Personendaten gegen «Data Scraping»[53] veröffentlicht. Social Media Unternehmen und Webseitenbetreiber werden darin aufgefordert, entsprechende Massnahmen zu treffen.[54] Obwohl die Veröffentlichung von Gerichtsurteilen im Internet in anonymisierter Form erfolgt und sich damit die Frage des Datenschutzes kaum je stellt, hat mindestens ein Gericht in der Schweiz – allenfalls auch wegen der Erklärung des EDÖBs – nun bei seiner Urteilsdatenbank technische Hürden für ein Web-Scraping eingebaut. Man kann sich fragen, ob dies eine Verletzung der Justizöffentlichkeit darstellt, nachdem das Bundesgericht die Veröffentlichung von Urteilen im Internet als geeignete Massnahme der Justizöffentlichkeit bezeichnet hat.[55] Jedenfalls behindern aber solche technische Hürden allenfalls die neuen Methoden der Justizforschung.

3.4.2. Negation als sprachliche Herausforderung

[23] Die Negation ist ein wichtiges Merkmal der Sprache und ein wesentlicher Bestandteil der Informationsgewinnung aus Texten.[56] Sie stellt Schwierigkeiten für die kognitive und computergestützte Verarbeitung dar. Trotz vieler Fortschritte in der Textanalyse bleibt die Negationsauflösung eine aktuelle Forschungsfrage in der Verarbeitung natürlicher Sprache mittels Computer (Natural Language Processing).[57] Dies muss beim Arbeiten mit Methoden des «Data Mining» beachtet werden.

[24] Gerade in der deutschen Sprache kann die Negation verschiedene Formen haben. Im Rahmen einer – *nota bene* gescheiterten – Untersuchung der Praxis des Bundesgerichts zur Thematik der Rechtsfragen von grundsätzlicher Bedeutung zeigte sich, dass das Gericht bei der Verneinung des Vorliegens von Fragen grundsätzlicher Bedeutung verschiedenste Formulierungen verwendet:[58] «keine Fragen von grundsätzlicher Bedeutung»; «Eine Rechtsfrage von grundsätzlicher Bedeutung stellt sich in dieser Konstellation nicht, …»; «mangels Vorliegens einer Rechtsfrage von grundsätzlicher Bedeutung» Dies stellt selbst semantische Suchen vor erhebliche Herausforderungen.

52 Vgl. https://www.edoeb.admin.ch/dam/de/sd-web/TvfoxgaTWHUl/Data%20Scraping%20-Joint%20Statement%20-%20August%2024%202023.pdf.

53 Gemeint ist Web-Scraping, vgl. oben Fn. 25.

54 Vgl. https://www.edoeb.admin.ch/de/24082023-data-scraping.

55 Vgl. Urteil des Bundesgerichts 1C_123/2016 vom 21. Juni 2016, E. 3.5.1 i.V-m. 3.6.

56 Vgl. Aditya Khandelwal/Suraj Sawant, NegBERT: A Transfer Learning Approach for Negation Detection and Scope Resolution, Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 5739–5748, S. 5739.

57 Vgl. Anastassia Shaitarova/Fabio Rinaldi, Negation typology and general representation models for cross-lingual zero-shot negation scope resolution in Russian, French, and Spanish, Proceedings of NAACL-HLT 2021: Student Research Workshop, pages 15–23, S. 15; Ramona Christen/Anastassia Shaitarova/Matthias Stürmer/Joel Niklaus, Resolving Legalese: A Multilingual Exploration of Negation Scope Resolution in Legal Documents, arXiv:2309.08695, https://arxiv.org/abs/2309.08695.

58 Erkenntnisse aus der eigenen Forschung der Autoren.

3.4.3. Inhärente Ungenauigkeit von LLMs

[25] Neuere Publikationen zeigen an Untersuchungen im Bereich Legal-Tech[59], bei Anwendungen im Medienbereich[60] sowie in Studien der Computerforschung[61] auf, dass Anwendungen der generativen KI oft zu ungenauen Ergebnissen führen, weil die KI halluziniert. Das betrifft – je nach Anwendungsgebiet und Studie – zwischen 16 und 50% aller Anwendungsfälle.[62] Man kann solchen Fehlleistungen generativer KI teilweise vorbeugen durch Training mit einer besseren Datenbasis, durch besseres Prompting, durch das Arbeiten mit Retrieval-augmented Generation (RAG)[63], durch Aneinanderreihung von KI-Assistenten sowie durch automatisierte Überprüfung der ausgegebenen Zitierungen bzw. Fundstellen. Da die LLMs der generativen KI auf der Grundlage probabilistischer Methoden arbeiten, kann das Risiko von Halluzinationen aber dennoch nie ganz ausgeschlossen werden. Dies muss berücksichtigt werden, wenn mit solchen Anwendungen Forschung betrieben wird.

3.4.4. Ethische Fragen: Bias in LLMs, Replizierbarkeit und Datenschutz

[26] LLMs können auf Grund der Daten, mit denen sie trainiert wurden, Bias enthalten, was die Forschungsergebnisse verfälschen kann.[64]

[27] Ein zentraler Pfeiler der Sozialforschung – und damit auch der Justizforschung – ist, dass Forschende Studien erstellen sollten, die von anderen repliziert werden können. Es gibt indessen mehrere Gründe, warum Forschung mit generativer KI schwer zu replizieren ist.[65] Erstens sind diese Modelle probabilistisch. Selbst wenn ein einzelner Forschender identische Eingabeaufforderungen wiederholt verwendet, ist zu erwarten, dass ein LLM unterschiedliche Antworten liefert. Die Tendenz von LLMs, heterogene Ergebnisse zu erzeugen, kann teilweise durch so genannte Temperatur-Parameter gesteuert werden, die die Vorhersagbarkeit der Modellausgabe regulieren. Derzeit gibt es jedoch keine Standards dafür, welche Werte für diesen Parameter angemessen sind.[66] Die meisten LLMs liefern zudem auch im Laufe der Zeit unterschiedliche Ergebnisse. Dies liegt daran, dass viele von ihnen ständig optimiert werden, um sie effektiver zu machen oder neue

---

59 Vgl. Matthew Dahl/Varun Magesh/Mirac Suzgun/Daniel E. Ho, Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models, Journal of Legal Analysis 16, no. 1 (2024): 64–93 oder arXiv:2401.01301, https://arxiv.org/abs/2401.01301; Varun Magesh/Faiz Surani/Matthew Dahl/Mirac Suzgun/Christopher D. Manning/Daniel E. Ho, Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools, Journal of Empirical Legal Studies, 2025; 0:1–27; Joshua Yuvaraj, The Verification-Value Paradox: A Normative Critique of Gen AI in Legal Practice (October 18, 2025). The University of Auckland Faculty of Law Research Paper Series 2026, (2026) 52 Monash University Law Review (forthcoming), Available at SSRN: https://dx.doi.org/10.2139/ssrn.5621550; Li Zhang/Jaromir Savelka/Kevin Ashley, Do LLMs Truly «Understand» When a Precedent is Overruled?, arXiv:2510.20941 [cs.CL], https://doi.org/10.48550/arXiv.2510.20941.

60 Vgl. EBU/BBC (Hrsg.), News Integrity in AI Assistants, An international PSM study, Oktober 2025.

61 Vgl. Bernsohn et al. (Fn. 41), S. 9.

62 Zu dieser Bandbreite kommt man zusammenfassend aus den in Fn. 58 und 59 erwähnten Publikationen.

63 Dies ist eine KI-Technik, bei der ein generatives Sprachmodell (LLM) nicht nur sein internes (antrainiertes) Wissen nutzt, sondern auch Informationen aus externen Quellen abruft (beispielsweise aus einen zuvor erstellten Datencorpus aus Urteilen), um präzisere und aktuellere Antworten zu liefern. Dies geschieht, indem ein Abruf-System bei einer Anfrage relevante Daten findet, die dann mit der ursprünglichen Frage (Prompt) an das Sprachmodell gesendet werden, welches daraus eine Antwort generiert.

64 Dazu ausführlich Bail (Fn. 50), S. 5 ff.

65 Ausführlich dazu Bail (Fn. 50), S. 7.

66 Vgl. Bail (Fn. 50), S. 7.

Schutzmechanismen gegen Verzerrungen oder unzulässiges Verhalten zu schaffen.[67] Neben der Beurteilung der Reproduzierbarkeit von Ergebnissen innerhalb desselben Modells sollten Justizforscherinnen und Justizforscher auch die Stabilität der Ergebnisse über verschiedene Modelle hinweg berücksichtigen. Es gibt offenbar Hinweise darauf, dass unterschiedliche Modelle wesentlich unterschiedliche Ergebnisse liefern.[68]

[28] Letztlich können sich bei der Verwendung von Personendaten in offenen Anwendungen generativer KI Datenschutzprobleme stellen.

3.4.5. Rechtliche Schranken: Analyseverbote und in Services «verbotene» Themen

[29] Im März 2019 unternahm Frankreich einen umstrittenen Schritt und verbot als erstes Land weltweit ausdrücklich die Forschung zum individuellen Verhalten von Richterinnen und Richtern.[69] Es ist nun strafbar, das Verhalten einzelner Richter zu «bewerten, zu analysieren, zu vergleichen oder vorherzusagen». Die Höchststrafe beträgt bemerkenswerte fünf Jahre Haft. Der Anwendungsbereich des Gesetzes, obwohl es auf das französische Rechtssystem abzielt, ist unklar. Hat es beispielsweise extraterritoriale Auswirkungen, d.h. verbietet es auch die Forschung über französische Richter an internationalen Gerichten oder die Forschung über französische Gerichte von der Schweiz aus? Was ist nun erforderlich, um die betreffenden Richter zu anonymisieren und rechtlich auf der sicheren Seite zu sein?

[30] Was die Analyse richterlicher Tätigkeiten in der Schweiz auf der Grundlage von Gerichtsurteilen betrifft, kann man sich an der Dissertation von Peter Bieri orientieren: So kann festgehalten werden, dass auch veröffentlichte Urteile stets die Namen der den Spruchkörper bildenden Richterinnen und Richter enthalten müssen.[70] Dies impliziert, dass grundsätzlich die Urteile auch bezüglich statistischer Angaben bearbeitet werden dürfen, welche die einzelnen Richterinnen und Richter betreffen. Vorbehalte werden lediglich hinsichtlich der Wiederwahl von Richterinnen und Richter gemacht, indem es als «absolut unzulässig» bezeichnet wird, die Urteile von Richterinnen und Richter im Vorfeld von Wiederwahlen anhand elektronischer Urteilspublikationen systematisch nach politischen Kriterien zu analysieren.[71] Daran sollten sich Wissenschaftlerinnen und Wissenschaftler halten; dass bei Volkswahlen politische Aktivistinnen und Aktivisten diese Regel missachten, lässt sich hingegen wohl kaum verhindern. Der Vollständigkeit halber sei hier auch noch erwähnt, dass Dokumente der Justizverwaltung grundsätzlich zwar dem Öffentlichkeitsprinzip unterliegen können, nicht aber der Justizöffentlichkeit i.e.S. unterliegen,[72] so dass in solchen Dokumenten Personennamen grundsätzlich zu anonymisieren sind.[73]

[31] Eine ganz andere rechtliche – oft auch einfach faktische – Schranke sind strafrechtliche Schranken und Regeln der Anbieter von KI-Tools bezüglich unerlaubter Inhalte. Urteile zu bestimmten Fällen von Pornografie oder Gewaltdarstellungen können Texte enthalten, welche als

---

67 Vgl. Bail (Fn. 50), S. 7.

68 Vgl. Bail (Fn. 50), S. 7.

69 Vgl. Malcom Langford/Mikael Rask Madsen, France Criminalises Research on Judges, VerfBlog, 2019/6/22, https://verfassungsblog.de/france-criminalises-research-on-judges/, DOI: 10.17176/20190622-232658-0.

70 Vgl. Peter Bieri, Bearbeitung von Daten über Richterinnen und Richter, Schriftenreihe für Justizforschung Band 13, Bern 2017, Rz. 423.

71 Vgl. Bieri (Fn. 70), Rz. 577.

72 Für das Bundesgericht selber differenziert in BGE 113 II 209.

73 Vgl. Bieri (Fn. 70), Rz. 455.

solche unter die Strafnormen oder die einschränkenden Regulierungen der Anbieter fallen. Wenn solche Texte nicht «on premise» sondern in KI-Anwendungen «as a service» bearbeitet werden, so kann allenfalls ein strafbares Weiterverbreiten bzw. ein unerlaubtes Bearbeiten solcher einschlägiger Inhalte vorliegen, was zur Strafverfolgung oder zur Sperrung von Nutzerkonten führen kann.

4. Ausblick

[32] Wir leben in einer Zeit mit einer rasanten technologischen Entwicklung. Wohin sich die KI-Anwendungen noch entwickeln und welche neuen Möglichkeiten sich dadurch für die Justizforschung eröffnen, ist kaum abzuschätzen.

[33] Ungeachtet dessen werden zwei Problempunkte bleiben:

- Es ist nicht immer klar, wo die Grenzen zwischen Justizforschung und Forschung in anderen Disziplinen liegen. Wie weit ist beispielsweise Forschung in Computer Science zur Anonymisierung von Gerichturteilen der Justizforschung zuzurechnen, wie weit handelt es sich um Forschung zu Legal NLP?
- Als eher Nischen-Forschungsdisziplin kann es sich die Justizforschung eigentlich nicht leisten, die gleichen Datengrundlagen mehrfach zu erheben. Es wäre deshalb wünschenswert, wenn es ein schweizweites Verzeichnis aller abgeschlossenen und laufenden Erhebungen von Daten zur Justizforschung gäbe und wenn diese Daten in einem zentralen Repositorium frei verfügbar abrufbar wären.

Daniel Kettiger, Mag. rer. publ., Rechtsanwalt, ist Justizforscher und externer Forschungsleiter am Kompetenzzentrum für Public Management (KPM) der Universität Bern, sowie freischaffender Anwalt und Berater in Bolligen (kettiger.ch). Er ist Vorstandsmitglied des Vereins entscheidsuche.ch.

Prof. Dr. iur. Andreas Lienhard, Rechtsanwalt, ist ordentlicher Professor für Staats- und Verwaltungsrecht am Kompetenzzentrum für Public Management (KPM) und am Institut für öffentliches Recht der Universität Bern.

Diese Publikation ist eine verschriftlichte und erweiterte Fassung des Vortrags, den der Erstautor an der Tagung «Die Justiz im Fokus der empirischen Forschung: aktueller Stand und zukünftige Herausforderungen» des Obwaldner Instituts für Justizforschung (IFJ) am 20. November 2025 an der Universität Luzern hielt.