

Der Richter, (k)ein Bot?!

Vier Thesen zur Digitalisierung des Rechts aus rechtslinguistischer Perspektive

Friedemann Vogel, Version vom 29.09.2023

The subject of the text is a critical examination of the practical possibilities and methodological, legal-political and social implications of digitalisation of the law and a (possible) use of automated systems in justice from a legal-linguistic perspective.

Prof. Dr. Friedemann Vogel ist Universitätsprofessor für Germanistische Linguistik mit Schwerpunkt auf der computergestützten Analyse von Sprache und strategischer Kommunikation in Politik, Medien und Recht an der Universität Siegen.

Das nachfolgende Manuskript ist die Ausarbeitung eines Vortrags, gehalten am 11.05.2023 auf der Tagung zur „Digitalisierung des Rechts“ an der Universität Konstanz (<https://www.jura.uni-konstanz.de/digitalisierung/digitalisierung-des-rechts/>).

Der Text erscheint demnächst in: Wörner/Wilhelmi/Glöckner/Breuer/Behrendt (Hrsg.), Digitalisierung des Rechts, Berlin 2023.

I. Einleitung: Zum Ausgangspunkt meiner Überlegungen

Ich selbst habe keinen rechtswissenschaftlichen Hintergrund und die juristische Praxis kenne ich nicht aus eigener Erfahrung, sondern nur aus der forschenden Vogelperspektive. Das sage ich deshalb noch einmal gesondert, weil ich gerade beim vorliegenden Thema sicherlich an vielen Stellen zu Pauschalisierungen tendiere, wo juristische Praktikerinnen und Praktiker stärker differenzieren würden nach Zeit, Rechtsgebiet, Verfahrenskontext, Fallgruppen etc. Als empirischer Sprachwissenschaftler und Rechtslinguist habe ich mich in den letzten zehn Jahren mit juristischer Textarbeit beschäftigt. Digitalisierung spielte und spielt dabei in dreierlei Form eine wichtige Rolle: erstens interessiere ich mich für die Folgen, die die zunehmende digitale Verdatung von juristischer Textarbeit nach sich zieht; zweitens habe ich – vor allem auch im Austausch mit juristischen Kollegen wie Ralph Christensen, Hanjo Hamann, Tonio Walter und anderen – vergleichsweise früh mit computergestützten, vor allem korpus- und computerlinguistischen Methoden experimentiert, um mit deren Hilfe in „großen“¹ Textsammlungen („Korpora“) Muster in der juristischen Fachkommunikation zu untersuchen. Aus diesen Experimenten wurden ernsthafte Studien, die sowohl zur Theorie juristischer Sprache und Fachkommunikation beigetragen haben als auch – allerdings mit sehr viel verhaltenerem oder zögerlicherem Erfolg – neue Überlegungen in der Rechtstheorie und Rechtsmethodik angeregt

¹ Vor zehn Jahren waren 30.000 Texte noch viel, heute arbeiten wir teilweise mit vielen Hunderttausenden Texten.

haben.² Eine dritte Rolle spielt Digitalisierung in meiner Arbeit als meist umstrittener Gegenstand in Rechtsdiskursen, wenn es etwa um die Frage geht, wie Gesetzgebung, Behörden, Justiz usw. Medientechnik zu Überwachungszwecken strategisch zu perspektivieren und durchzusetzen versuchen (etwa am Beispiel „Online-Durchsuchung“³).

Das sind also drei unterschiedliche, aber miteinander verbundene Aspekte von „Digitalisierung“ des Rechts: a) seine Verdatung, b) maschinelle Verfahren der Sprachanalyse und -aggregation in juristischer Methodik und Forensik, c) Kommunikationsmedien als Gegenstand von Rechtsetzung und als Objekt rechtlicher Beurteilung. Ich will im Folgenden – quasi als interessierter Zaungast – versuchen, vor allem zu den ersten beiden Aspekten vier Thesen zur Diskussion zu stellen.

II. Zwischen Effizienz und methodischer Kontrolle

Die erste These lautet: *Die Verdatung juristischer Textarbeit bringt teilweise mehr Transparenz und Effizienz in den Rechtsstaat – zugleich überfordert sie die Praxis (bislang) mangels methodischer Strukturierung.*

Rechtstexte – seien es Urteilstexte, Kommentarliteratur, rechtswissenschaftliche Fachliteratur, ja auch Gesetzgebungsmaterialien – sind heute mehr denn je Gegenstand von Verdatungsprozessen. Sie werden zunehmend nachdigitalisiert oder liegen bereits in maschinenlesbarer Form vor, werden mit Metadaten versehen und gehen in Datenbanken ein. Diese Datenbanken – Beck Online, Juris, OpenJur, DIP und viele andere – bieten diese Textdaten wiederum ihren Nutzern an, mehrheitlich im Rahmen von Lizenzmodellen, selten open access, meist mithilfe einer GUI (Graphical User Interface) mit beschränktem Funktionsumfang, sehr selten auf Basis einer offenen API (einer technischen Schnittstelle, die z. B. Zugriff auf strukturierte Datenformate erlaubt). Die Datenbanksysteme werden oftmals mit inzwischen hocheffizienten Hypertext- und Filterfunktionen verbunden, um die mittlerweile gigantischen Datenmengen zu durchsuchen, zu vernetzen und algorithmisch zu ranken, zu sortieren oder neu zu kategorisieren. Grundlage hierfür bilden Fortschritte in Rechtsinformatik, Datenbankentwicklung, Indexierungsverfahren, Natural Language Processing, Sprachmodellen (dazu unten) und vielem anderen.

Dieser immer größer werdende Pool an Rechtstextdaten bildet eine Fundgrube für die Rechtspraxis, sie vereinfacht Recherchen und Archivierung, reduziert Laufwege bei der Besorgung, sie erleichtert auch die Rechtsempirie bzw. Rechtstatsachenforschung, sie macht also Rechts-

² Vogel, in: Felder/Müller/Vogel (Hrsg.), Korpuspragmatik, 2012, S. 314–353; Vogel (Hrsg.), Zugänge zur Rechtssemantik, 2015; Hamann/Vogel BYU Law Review 2017 (2017), 1473; Vogel/Hamann/Gauer Law & Social Inquiry (LSI) 43 (2017), 1340; Vogel/Walter/Tripps (Hrsg.), Korpuslinguistik im Recht, 2022.

³ Vogel, Linguistik rechtlicher Normgenese, 2012.

praxis damit transparenter. Effizienter wird sie auch insofern, als aus den Textmengen musterhafte Textbausteine erschlossen und als Templates in Textbearbeitungsprogrammen⁴ zur Verfügung gestellt werden können. Warum sollte man jeden Standardsatz der Dogmatik immer wieder von neuem abtippen, wenn man inzwischen auch auf *copy & paste*-Vorlagen zurückgreifen kann, nicht nur für Formulare und typisierte Anschreiben, sondern auch für Urteilsentwürfe bestimmter Fallgruppen usw.⁵ Genau genommen ist das nicht nur eine Frage der Effizienz, sondern auch der Rechtssicherheit, nämlich dann, wenn gleiches Recht(sverständnis) zuverlässig auch seinen erwartbaren bzw. identischen Ausdruck im Text findet.

Schließlich: Aus dem Nutzungsverhalten – also den durch Tracking-Algorithmen nachverfolgten Textzugriffen – der Datenbank-Endverbraucher könnte man wohl rückschließen, welche Rechtsprobleme und Rechtsgebiete am häufigsten nachgefragt sind, vielleicht auch, welche Normtexte größeren Nachschlagebedarf erzeugen. Aus diesen Benutzerverhaltensstatistiken kann man gezielt die weitere Datenakquise vorantreiben, die Benutzeroberflächen verbessern, das Angebot effizienter sortieren, vielleicht sogar Rückschlüsse für eine gesetzesredaktorische Optimierung ziehen⁶.

Alles eine schöne neue Datenwelt? Durchaus, aber man sollte sich keinen Illusionen hingeben:

- a) Bis heute bleibt beim Großteil der Datenbanken für den Endnutzer völlig ungewiss, wer die ihm vorliegenden Daten warum ausgewählt und wie aufbereitet hat. Die Veröffentlichung von Gerichtsurteilen folgt schon ohne Digitalisierung mehr oder weniger dem Zufallsprinzip, um nicht zu sagen der Willkür von AutorInnen bzw. Datenbankbetreibern. Der absolute Großteil gelangt nie in eine öffentlich zugängliche Datenbank: „Mindestens 99 von 100 streitigen Entscheidungen der ordentlichen Gerichte bleiben unveröffentlicht. Ein halbes Jahrhundert technischen Fortschritts hat daran so gut wie nichts geändert.“⁷ Es fehlt an Konventionen und auch an Vorschriften für diese Praxis. Die Hoffnung auf eine größere Transparenz des Rechtssystems bleibt damit weit hinter den Erwartungen und den technischen Möglichkeiten zurück.
- b) Schon 2014 haben wir auf einer Freiburger Tagung diskutiert (2015 als Buch erschienen⁸), welche Folgen ein unreflektierter Datenbankgebrauch in der Rechtspraxis mit sich bringen könnte. Etwas schlaglichtartig formuliert: Was an Argumentationsarbeit oder auch an theoretischen und methodischen Beiträgen nicht digital in Datenbanken vorliegt, droht mittel-

⁴ Beispiele damit verbundener, spezialisierter Textverarbeitungsprogramme sind etwa forumSTAR-Text, bk.text (offenbar in Entwicklung), codefy, NoRA und andere unzählige Programme für den Einsatz in Kanzleien.

⁵ Das ist vermutlich auch der Grund für die zunehmende Textlänge von Gerichtsentscheidungen.

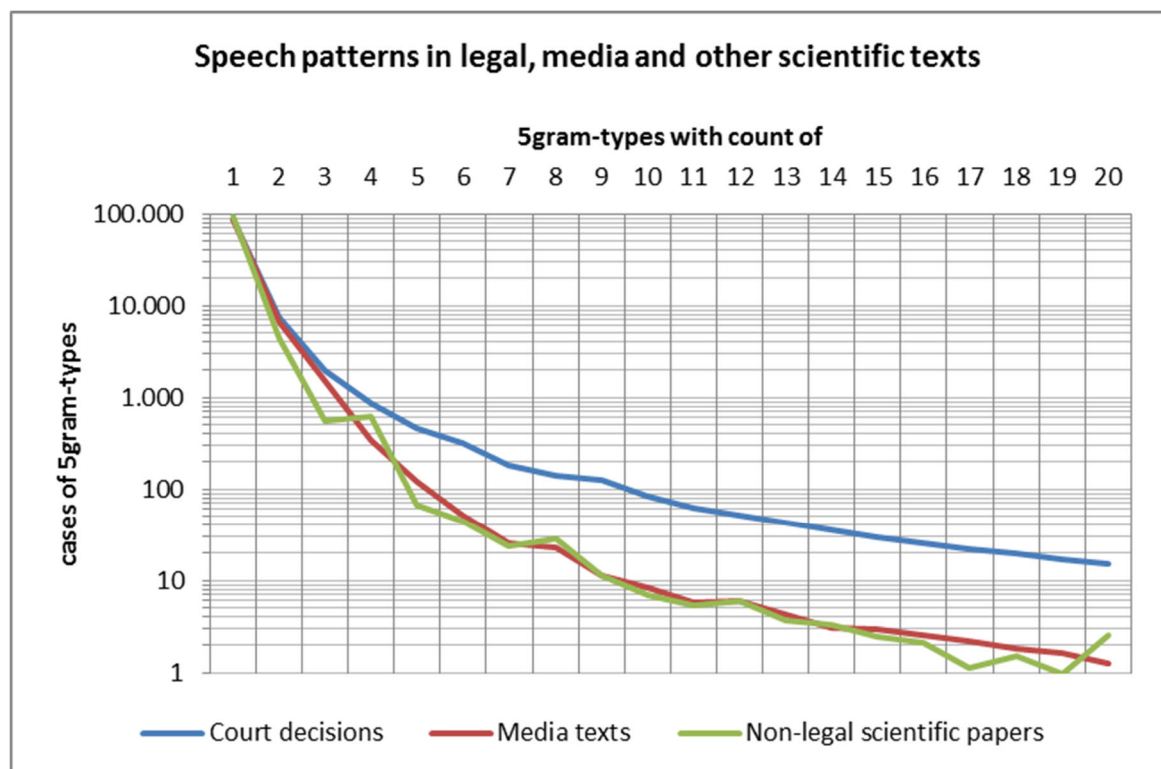
⁶ Vgl. zur gesetzesredaktorischen Arbeit im Bundesjustizministerium: *Vogel/Deus/Luth/Schmallenbach/Felder*, Gesetzesverständlichkeit aus rechtslinguistischer Perspektive, 2022.

⁷ *Hamann* JuristenZeitung (JZ) 76 (2021), 656, 658.

⁸ *Vogel* (Fn. 2).

und langfristig aus der Diskussion zu verschwinden oder gar nicht erst aufgegriffen zu werden⁹. Ohne methodische Strukturierung (im Sinne der strukturierenden Rechtslehre¹⁰) könnten die Algorithmen der maschinellen Textverarbeitung und die erzeugten Rankings und Sortiermaschinen unbeabsichtigt dogmatische Präferenzen konstruieren. Nutzungsstatistiken¹¹ werden heute in fast allen kommerziellen Plattformen genutzt, um damit Userprofile zu modellieren, zu vergleichen und Empfehlungssysteme zu füttern. „Empfehlung“ klingt toll und klingt nach Arbeitserleichterung. Empfehlung bedeutet aber auch Vorentscheidung – im Alltag (zum Beispiel bei der Suche nach Kleidern oder kompatibler Hardware) ist das manchmal praktisch, in der Rechtspraxis aber droht dann leicht ein Vor-Urteil. Wer entscheidet über diese Verfahren, wer kontrolliert, wie Nutzungsstatistiken, Datenaufbereitung und -präsentation und Rechtspraxis miteinander konfundieren? Empirische Forschung zu diesen Zusammenhängen ist mir nicht bekannt und Einsichten in halb- oder illegale Praktiken großer Datenunternehmen oftmals „Zufällen“ (Leaks) geschuldet.

- c) Für einen weiteren Punkt muss ich etwas ausholen: Die Rechtssprache ist musterhafter als etwa die Mediensprache oder zum Beispiel auch als die Fachsprache der Linguistik. Das ist nicht nur intuitiv so, sondern lässt sich auch empirisch zeigen: der Anteil an hochfrequenten, mehr oder weniger feststehenden Mehrworteinheiten ist in der Rechtssprache größer¹²:



⁹ *Schuhr*, in: *Vogel* (Hrsg.), *Zugänge zur Rechtssemantik*, 2015, S. 93–123; *Morlok*, in: *Vogel* (Hrsg.), *Zugänge zur Rechtssemantik*, 2015.

¹⁰ *Müller*, F., *Strukturierende Rechtslehre*, 2. Aufl. 1994; *Müller*, F./*Christensen*, *Juristische Methodik*, 11. Aufl. 2013.

¹¹ Mitunter auch mithilfe von Drittanbietern: *Mayer/Mitchell*, *Security and Privacy (SP)*, 2012 IEEE Symposium on, 2012, S. 413–427.

¹² *Vogel*, in: *Giltrow/Stein* (Hrsg.), *The Pragmatic Turn in Law*, 2017, S. 287–306.

Abbildung 1: Mehrgliedrige Sprachgebrauchsmuster (hier: 5-Wort-Einheiten) finden sich in Gerichtsentscheidungen deutlich häufiger als in Presse- oder sprachwissenschaftlichen Texten (aus Vogel 2017a).

Die Wahrscheinlichkeit dafür, dass eine bestimmte Mehrworteinheit (hier: 5-Wort-Einheiten wie z.B. *Grundrecht auf Gewährleistung der Vertraulichkeit* usw.) genau einmal vorkommt, ist in Gerichtsurteilen genauso hoch wie in Presstexten oder sprachwissenschaftlichen Fachtexten. Dass eine bestimmte Mehrworteinheit über einen größeren Zeitraum hinweg aber zum Beispiel 10 mal verwendet wird, ist in Gerichtsurteilen um den Faktor 10 wahrscheinlicher als in Presse- oder anderen nicht-juristischen (hier: sprachwissenschaftlichen) Fachtexten. Der Grund für die Musterhaftigkeit liegt darin, dass Mehrworteinheiten in der Rechtssprache Ausdruck – und auch Ausweis – von Dogmatik-Bezügen sind. Man signalisiert durch die korrekte Verwendung solcher Mehrworteinheiten (auch ohne sie als Zitate zu markieren) die Kenntnis und die (idealerweise argumentative) Berücksichtigung höchstrichterlicher Rechtsprechung und damit verbundener Konzepte¹³. Dieser Zusammenhang gilt meines Erachtens unabhängig vom Medienwandel. Die Digitalisierung des Rechts aber, also die Möglichkeit einer verlustfreien und nahezu zeitlosen Reproduzierbarkeit von Kommunikationszeichen, könnte – so eine Vermutung schon vor 10 Jahren – die ausdrucksseitige Musterhaftigkeit des Rechts noch befördern. Klickbare Texttemplates und *Copy & Paste* könnten über die Zeit erstens dazu beitragen, dass die Länge juristischer Texte kontinuierlich über die Zeit zunimmt. Letzteres ist nach meinen Daten der Fall und stellt auch die Frage nach der langfristigen Bewältigbarkeit dieser Textmassen: in einem Textkorpus aus allen über OpenJur¹⁴ im Mai 2023 verfügbaren deutschen Gerichtsentscheidungen ausgewählter Instanzen¹⁵ nimmt die durchschnittliche Textlänge (Durchschnitt der Mediane) seit den späten 1990er Jahren tendenziell zu.¹⁶ Vor allem bei den unteren Instanzen, wo das rechtliche ‚Alltagsgeschäft‘ erledigt wird, legen die Textlängen teils deutlich zu. Bei den Bundes- und Oberlandesgerichten gab es bis Ende des 20. Jahrhunderts offenbar einen deutlichen Rückgang der Textlängen, im Anschluss aber auch hier Zuwächse, am deutlichsten beim BFH, für den hier nur Daten seit 2007 vorliegen. Zum Vergleich: Bei juristischen Fachzeitschriften – hier untersucht auf Basis des Juristischen Referenzkorpus (JuReko)¹⁷ – zeigt sich angesichts (wandelnder) redaktioneller Richtlinien ein anders und heterogeneres Bild.

¹³ Vogel, in: *Felder/Müller/Vogel* (Fn. 2).

¹⁴ <https://openjur.de/> (27.09.2023).

¹⁵ Insg. enthält das in meiner Forschungsgruppe aufbereitete OpenJur-Korpus 583.449 Texte aus Gerichtsentscheidungen zugrunde.

¹⁶ Auch für Schweizer Rechtsprechungstexte ist eine deutliche Textlängenzunahme ab den 1990er Jahren belegt: *Abegg/Perić*, *Sprache und Sprachgebrauch des Rechts*, 2021, S. 72ff.

¹⁷ *Vogel/Hamann/Gauer* *Law & Social Inquiry* (LSI) 43 (2017), 1340.

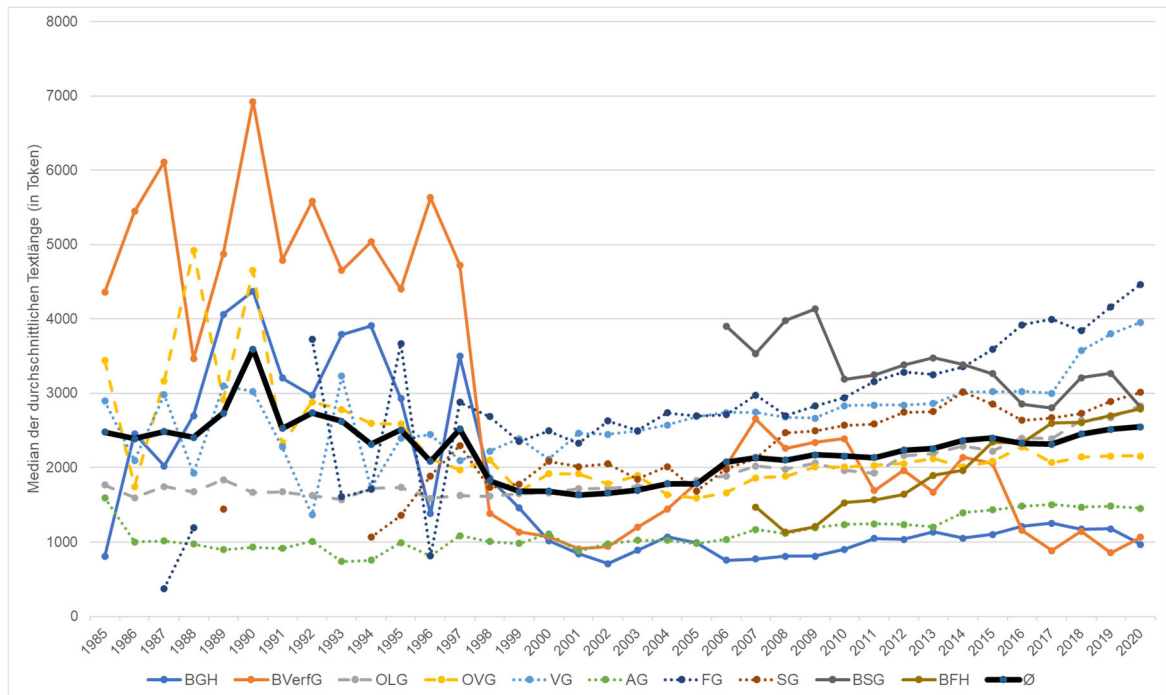


Abbildung 2: Durchschnittliche Textlängen (Median) in Token bei Gerichten über die Zeit; Datengrundlage: Ausschnitt des OpenJur-Korpus mit insg.

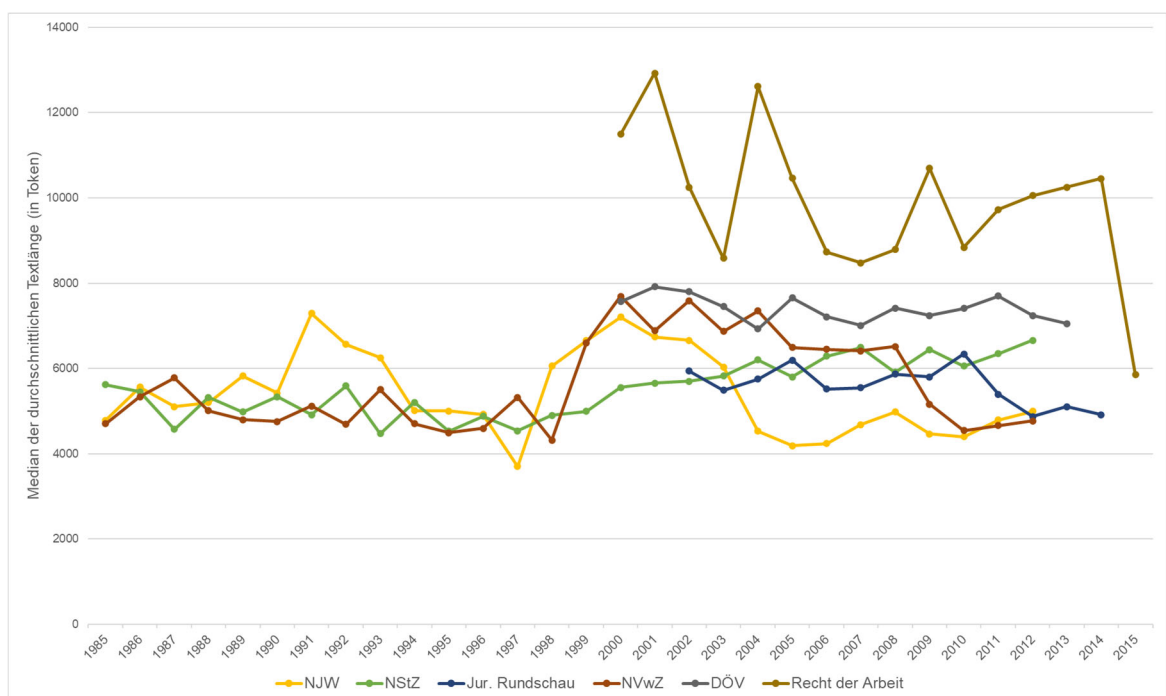


Abbildung 3: Durchschnittliche Textlängen (Median) in Token bei ausgewählten juristischen Zeitschriften über die Zeit; Datengrundlage: Ausschnitt des JuReko-Korpus mit insg.

Über die bloße Textlänge hinaus könnte zweitens auch die sprachliche Redundanz und damit quasi-dogmatische Verknöcherung im Rechtsdiskurs zunehmen, und umgekehrt die Argumentationstiefe sowie der Einzelfallbezug abnehmen (so einst die Vermutung von

Martin Morlok¹⁸). Für eine Zunahme an Redundanz gibt es bei allen Vorbehalten fehlender Grundgesamtheit¹⁹ in meinen Daten tatsächlich Indizien: der relative Anteil einmalig vorkommender Mehrworteinheiten²⁰ sinkt diachron in einem Korpus von etwa 60.000 Verwaltungsgerichtsurteilen (2000 bis 2015), der relative Anteil von mehrfach vorkommenden Mehrworteinheiten steigt dagegen über die Zeit kontinuierlich an. Mit anderen Worten: der Anteil an hochfrequenten identischen Formulierungen nimmt über die Zeit hinweg zu²¹.

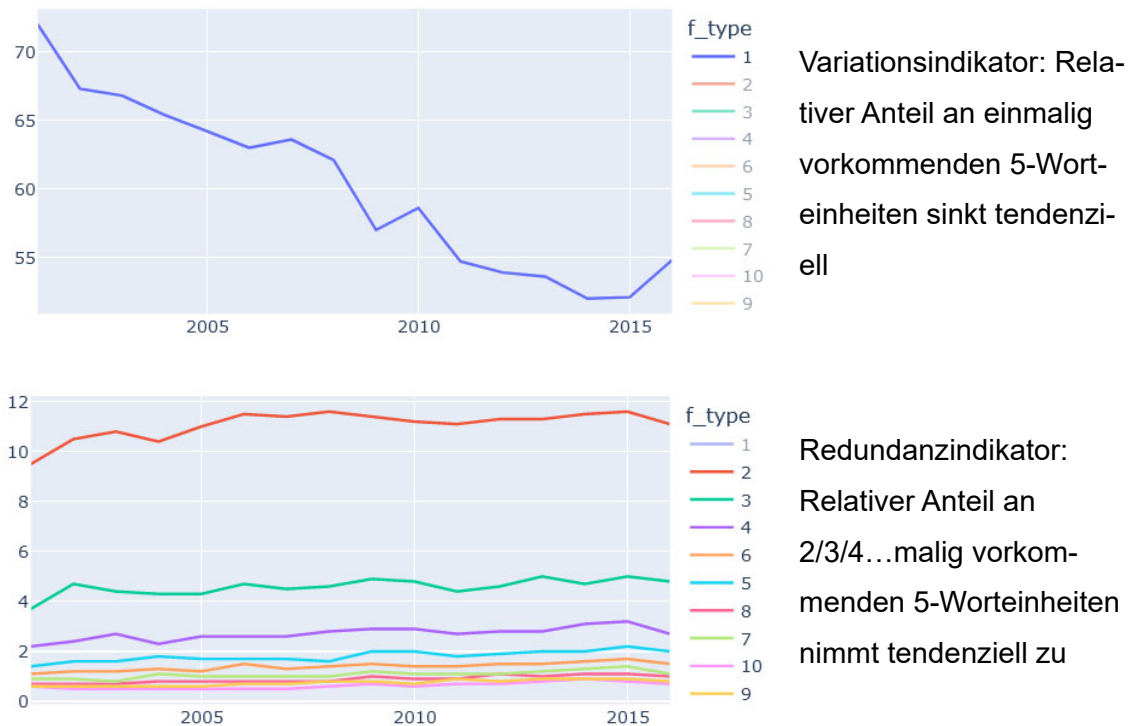


Abbildung 4: Variations- und Redundanzindikatoren in Texten der Verwaltungsgerichtsbarkeit 2000-2015.

Damit ist natürlich noch nicht gesagt, dass die zuvor behaupteten negativen Effekte ebenso zutreffen, aber ich denke, es lohnt sich, dem theoretisch und empirisch weiter nachzugehen.

Was ich mit dieser ersten These sagen möchte: die Verdattung des Rechts hat viele verschiedene Effekte auf die Rechtspraxis. Nach meiner Wahrnehmung wird über damit verbundene

¹⁸ Morlok, in: Vogel (Fn. 9).

¹⁹ Die Analysen können hier nur Tendenzen zeigen; für eine genauere Prüfung bedürfte es einer noch größeren Datengrundlage respektive Zugriff auf die Grundgesamtheit aller Gerichtsentscheidungen zumindest ausgewählter Gerichte oder Instanzen. Das ist bislang nicht einmal im Ansatz möglich: *Hammann* JuristenZeitung (JZ) 76 (2021), 656.

²⁰ Diversitäts- und Redundanzmaße (wie TTR oder MTLD, vgl. *McCarthy/Jarvis* Behav Res Methods 42 (2010), 381) beziehen sich in der Regel auf die Einwortebene und berechnen das Verhältnis der Häufigkeit verschiedener Worttypen (Types) und der Gesamtzahl aller Wörter (Token) in einem Text oder Teilkorpus. Hier interessiert aber nicht die lexikalische Diversität, sondern die Häufigkeit, in der Gerichtstexte mehrere Wörter in derselben Reihenfolge (musterhafter Gebrauch von Mehrworteinheiten) verwenden.

²¹ Zum Vergleich: in einem Korpus aus allen *Der Spiegel*-Presstexten im vergleichbaren Zeitraum (2000-2015) bleibt der Anteil einmalig vorkommender 5-Worteinheiten über die gesamte Zeit relativ konstant (bei etwa 93%).

Fragen durchaus diskutiert, aber von einer rechtstheoretischen oder methodischen Durchdringung oder gar Konventionalisierung im Rahmen juristischer Methodenausbildung sind wir noch weit entfernt. Auch für die empirische Rechtslinguistik gibt es noch viel Arbeit.

II. Korpuslinguistische Zugänge zur juristischen Semantik: Fluch und Segen für die Rechtspraxis?

Die zweite These lautet: *Die Entwicklung und Popularisierung von korpus- und computerlinguistischen Textanalyseverfahren hat eine juristische Methodendiskussion ausgelöst, aber es ist überhaupt nicht gesagt, dass damit die juristische Praxis auch besser würde.*

Die These mag vielleicht jene überraschen, die auch meine eigenen Gehversuche auf diesem Feld kennen. Seit 2010 habe ich versucht, computergestützte Verfahren der Sprachmusterberechnung auf rechtsdiskursanalytische Fragen anzuwenden. Die Grundidee eines ersten Aufsatzes 2012 war, dass die Quantifizierung von rekurrenten Sprachgebrauchsmustern in tausenden Gerichtsentscheidungen des Bundesverfassungsgerichts und anderer Gerichte Hinweise auf sedimentierte Konzepte, Denkmuster, Mentalitäten und Ideologien in der juristischen Dogmatik geben kann²². Denn wenn wir (in einem sprachpragmatischen Verständnis²³) Bedeutungen als Gebrauchsmuster von sprachlichen Ausdrücken verstehen, dann liegt es auf der Hand, mit dem Computer in großen Textsammlungen solche Muster automatisiert auszählen zu lassen und daraus Rückschlüsse im Hinblick auf die diesem Gebrauch zugrundeliegenden kognitiven Schemata (als Teil der Semantik) zu ziehen.

²² Vogel, in: *Felder/Müller/Vogel* (Fn. 2).

²³ Damit ist keine „praktische Sprache“ gemeint, sondern eine Theorie- und Denktradition in Gefolge der Sprachphilosophie (Wittgenstein, Austin, Searle u.v.a.), die Sprache weniger auf ihre immanenten Strukturen hin untersucht, sondern als Gegenstand oder Medium sozialer Handlungen interpretiert.

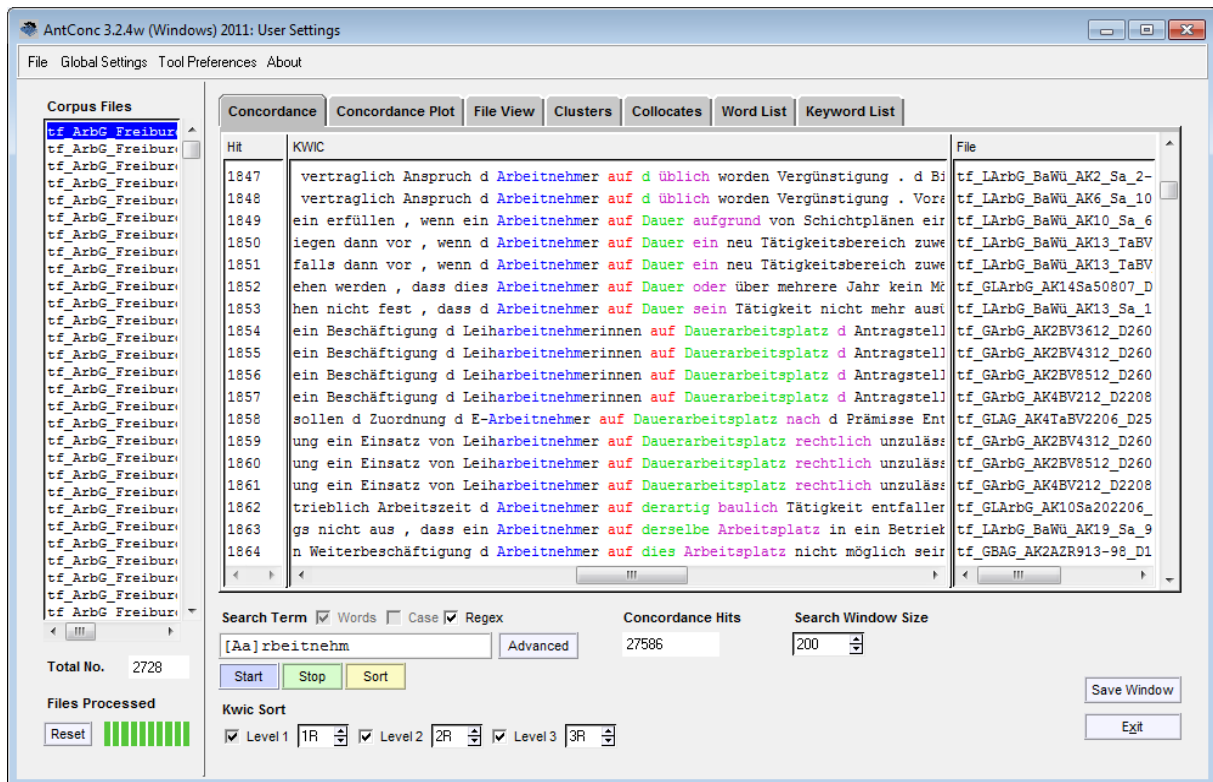


Abbildung 5: Screenshot des für Forschungszwecke frei verwendbaren korpuslinguistischen Toolkits „AntConc“²⁴, das ohne Programmierkenntnisse einfache Textanalysen erlaubt.

Eine etwas andere Zielsetzung verfolgt der Einsatz sprachstatistischer Verfahren bei der Frage nach Bedeutungsvarianten von sprachlichen Ausdrücken vor allem als Teil der Alltags- bzw. Gebrauchssprache: Was ist die im sog. Gemeinsprachgebrauch „übliche“ und damit auch in Gesetzgebung und/oder Rechtsprechung antizipierbare Bedeutung eines Wortes oder einer Phrase im Normtext? Was heißt die im amerikanischen Strafrecht verwendete Phrase „carry a firearm“ aus Sicht der amerikanischen Alltagssprache? – Eine methodische Frage, deren Diskussion in den USA seit 2011 eine geradezu steile Karriere hingelegt hat²⁵? Oder wie verstehen juristische Laien prototypisch den Ausdruck „geschäftsmäßig“, wie er in §217 StGB verwendet wird – und welche Folgen hat das für die rechtliche Beurteilung der Sterbehilfe-Praxis?²⁶

– Die korpusbasierte, maschinelle Sprachempirie verspricht mehr Transparenz, mehr interpretativen Nachvollzug als die bisherigen Standardpraktiken, als etwa die Konsultation der eigenen, höchst fehleranfälligen Sprachintuition oder der unsachgemäße Blick in Wörterbücher. Zunächst vor allem in den USA und mit Wirkung bis in die obersten Gerichte²⁷, aber seit etwa fünf Jahren auch hierzulande ist daher das Thema „Korpuslinguistik im Recht“ – salopp gesagt

²⁴ Anthony Professional Communication Conference. IPCC 2005. Proceedings 2005, 729; Anthony, AntConc (3.2.4w) [Computer Software], <http://www.antlab.sci.waseda.ac.jp/>, 22.05.2013.

²⁵ Mouritsen Brigham Young University Law Review 2010, 1915; Mouritsen International Journal of Language & Law (JLL) 2017, 67.

²⁶ Vogel/Bäumer/Deus/Rüdiger/Tripps LeGes 30 (2020).

²⁷ Mouritsen/Lee The Yale Law Journal 2018, 788.

– ein echter Renner. Grund für diese Popularisierung ist meines Erachtens nur bedingt ein Mehr an Effizienz und Transparenz bei der Betrachtung von Sprachmustern. Vielmehr verspricht der Einsatz computergestützter Sprachanalyse vor allem in der Justiz einen willkommenen szientistischen Autoritätstopos bei der Durchsetzung von interessegeleiteten und so präferierten Lesarten (z. B. zugunsten eigener und entgegen anderer Mandanten). Publikationen zu diesem Feld haben enorm zugenommen, die Probleme allerdings auch: Oft wird nämlich ignoriert oder zumindest unterschätzt, dass eine maschinelle Analyse und Quantifizierung von Sprachgebrauch die Rechtspraxis nicht einfach nur „leichter“ macht, sondern lediglich anders. Das Zählen von Wörtern ersetzt nicht ihre Interpretation im Lichte einer verfassungsrechtlich fundierten Rechtshermeneutik²⁸. Dazu braucht es ein Verständnis grundlegender pragmatischer Sprachtheorie. Und es muss auch dogmatisch geklärt werden, welcher Sprachgebrauch denn überhaupt jeweils einschlägig ist für die jeweilige Recherche und Analyse: bei der strafrechtlichen Einordnung von inkriminierten Äußerungen als Beleidigung u. ä. mag der Fall relativ klar sein; aber wie sieht das bei Normtextausdrücken aus? Das sind keine methodischen, sondern dogmatische Fragen, die die neuen digitalen Verfahrensmöglichkeiten nicht lösen, sondern im schlimmsten Fall sogar verdecken.

III. ChatGPT & Co: Zwischen Entlastung der Rechtspraxis und Simulation von Rechtsstaatlichkeit

Die Problemfelder der ersten beiden Thesen kulminieren ein Stück weit in meiner dritten These und der neuesten Entwicklung hochperformanter Sprachmodelle: *Die enorme Entwicklung von Machine Learning-Verfahren und Sprachmodellen auf dem Niveau von ChatGPT eröffnen völlig neue Anwendungsszenarien in Form von Schreib-Assistenzsystemen für die Rechtspraxis in Gericht, Rechtsetzung, Verwaltung und Wissenschaft. Ihre fehlende Kontrolle und intuitive bzw. fehlerhafte Verwendung aber sind eine Bedrohung für den Rechtsstaat.*

Als ich im November 2022 – über Twitter angeregt – zum ersten Mal mit ChatGPT in Berührung kam, war ziemlich schnell klar: das verändert alles. Im Blick hatte ich dabei gar nicht zuerst das Rechtswesen, sondern alltägliche und professionelle kommunikative Praktiken in der Wissenschaft, im Call-Center oder in öffentlichen Diskursen. In den vergangenen Jahren wurde etwa regelmäßig von „Social Bot-Armeen“ berichtet, die nicht von „normalen Usern“ unterschieden werden und darum enormen Einfluss auf Wahlverläufe u. ä. haben sollten. Das behaupteten jedenfalls zahlreiche „Analysten“²⁹, die jedoch auf Nachfrage regelmäßig keinen einzigen Beleg für ihre Behauptungen vorweisen konnten³⁰. Und das war nicht überraschend,

²⁸ vgl. die Beiträge in: *Vogel/Walter/Tripps* (Fn. 2).

²⁹ *Murthy/Powell/Tinati/Anstead/Carr/Halford/Weal* International Journal of Communication 10 (2016), 20; *Bessi/Ferrara* First Monday 21 (2016).

³⁰ *Gallwitz/Kreil*, The Rise and Fall of 'Social Bot' Research, <https://ssrn.com/abstract=3814191>.

denn die sprachmodellierenden Automaten waren weit entfernt von der Qualität alltagstypischer Kommunikation. Kurz: „Social Bots“ als Bedrohung für Politik und Demokratie waren eine Chimäre und ein Topos zur Stützung von Regulierungsambitionen. Aber diese Gewissheit ist mit den Sprachmodellen der Generation GPT3, GPT4 und den zu erwartenden Nachfolgern passé.

Ohne hier in die technischen Details der Funktionsweise einsteigen zu können (dafür sind in diesem Band andere auch deutlich besser qualifiziert als ich): Für die Textpraxis allgemein, aber auch im Besonderen für die *juristische* Textpraxis eröffnen sich mit den neuen Sprachmodellen völlig neue Möglichkeiten der Human-Computer-Interaction und für Schreib-Assistenzsysteme³¹. Mit einer weiteren Verbesserung der jeweiligen Trainingsdaten – vor allem auf Basis von bislang fehlenden juristischen Texten – reichen zukünftig wenige Stichwörter, um einen *sachverhaltsbezogenen* Rohtext für verschiedene Textsorten zu generieren, seien es Urteilstexte, Normtextentwürfe oder Gesetzesbegründungen. Das gleiche gilt für Entwürfe von Zusammenfassungen auf Basis von Eingangsdaten – man denke an die Möglichkeit, Fallakten nach (aber nicht vor!) ihrem Detailstudium für das Sachverhaltsreferat oder eine Sammlung neuer Gerichtsurteile und Fachaufsätze für einen Kommentar automatisiert und abstrahierend zusammenfassen zu lassen. Für manche Rechtsbereiche und Fallgruppen könnten Spezialmodelle so gut werden, dass sie tatsächlich weitgereifte Texttemplates am Fließband erzeugen, einschließlich (simulierter) Argumentation, Anschluss an die Dogmatik und Kostenkalkulation. Zugunsten adressatenorientierter Gesetzesverständlichkeit könnten Paraphrasierungs- und Kommentierungsassistenten automatisiert Begleittexte zu Normtexten für verschiedene Adressatengruppen entwerfen und damit optionale Zugänge für eine Erstinformation zu rechtlichen Vorschriften bereitstellen. All das könnte die Rechtspraxis natürlich erheblich entlasten – sofern man die Möglichkeiten und vor allem Grenzen der Modelle und Assistenzsysteme kennt und zu beherrschen weiß. Das ist wiederum der entscheidende Knackpunkt: Denn im anderen Fall, also bei unsachgemäßer oder unreflektierter Verwendung dieser Systeme, könnte der Schaden für den Rechtsstaat beträchtlich sein. Zu dieser Einschätzung komme ich unter folgenden Annahmen:

- Zunächst: Sprachmodelle sind und bleiben immer statistische Näherungen, sie sind generell fehleranfällig und treffen im Kern ihrer Berechnungen nicht oder nur geringfügig voraussehbare bzw. kontrollierbare Filter-Entscheidungen („Blackbox“ infolge von „hidden layers“ im sog. Neuronalen Netz). Was bei der Erzeugung von Gedichten, Rezepten, Bewerbungsanschreiben oder selbst für Hausarbeiten relativ folgenlos ist, ist für die Rechtspraxis ein handfestes Problem. Nicht, weil die Arbeit ohne Assistenzsysteme fehlerfrei und allseits transparent wäre, nein: sondern weil das Sprachmodell

³¹ Allgemein zu Schreibassistenzsystemen vgl. *Steinhoff* MiDU 2023, 1; zum Einsatz von Sprachmodellen in der Justiz siehe *Mielke/Wolff* E-Zeitschrift für Wirtschaftsrecht & Digitalisierung (LRZ) 2023, Rn. 560.

selbst nicht erkennt, wann es Fehler produziert³² und die manuelle Nachkontrolle vor allem bei der Bewältigung wiederkehrender Aufgaben nicht oder nur eingeschränkt realisierbar wäre.

- Der Grad an Redundanz in Rechtstexten könnte sich selbstverstärkend erhöhen. Das liegt auf der Hand: denn die Sprachmodelle generieren musterhafte Texte auf Basis ihrer Trainings- bzw. Eingangsdaten. Sind die Eingangsdaten ihrerseits bereits teilweise muster-generiert, führt das in einen Zirkel, der es immer unwahrscheinlicher macht, dass seltenere Fallgruppen, Argumentationen oder rechtsdogmatische Konzepte zur Geltung kommen oder überhaupt berücksichtigt werden.
- Die neuen Hochleistungssprachmodelle basieren auf Datenmengen, Rechenleistungen und finanziellen Möglichkeiten³³, die bislang nahezu ausschließlich Großkonzernen zur Verfügung stehen: Alphabet (Google-Mutterkonzern), Microsoft, Apple, SAP und ähnliche. Selbst die Universitäten sind nach meiner jetzigen Kenntnis nur in sehr begrenztem Umfang in der Lage, hier Schritt zu halten. Das heißt aber: Es gibt bislang nur äußerst begrenzte Transparenz oder gar Kontrolle über die Auswahl der Trainingsdaten, der Algorithmen und manuellen Eingriffe. Zu manuellen Eingriffen gehören auch moralische Codices, die Sagbarkeitsgrenzen definieren – ohne dass diese Grenzen irgendwie demokratisch legitimiert wären. Man stelle sich vor, die Rechtspraxis nutzte Assistenzsysteme dieser Art ohne Kenntnis der Hintergründe und ihrer möglichen Folgen für die Textgenerierung, *ohne* sorgfältige qualitative Kontrolle und Rückversicherung! – Immerhin: Es finden sich immer mehr kleinere und mittelgroße Sprach- und Datenmodelle unter freien Lizenzen, auch die Effizienz der Trainingsalgorithmen hat durch die weltweite Popularität dieser Technologie innerhalb weniger Monate große Fortschritte erzielt, so dass man sich auch am heimischen Computer mittlerweile einfachere Modelle berechnen und Chatbots erzeugen kann. Diese Entwicklung außerhalb oder neben den ‚Big Playern‘ wird sicherlich noch weiter an Fahrt gewinnen, aber ich vermute, dass am Ende die Inhaber großer Ressourcen (etwa zur Erhebung und

³² Am Beispiel ChatGPT: das System gab in seinen ersten Monaten auf (fast) jede erdenkliche Frage eine Antwort und dies in aller Regel mit absolutem Geltungsanspruch. Wies man das System darauf hin, seine Antwort sei falsch (völlig gleich, ob dies der Tatsache entsprach oder nicht), entschuldigte sich das System höflich und behauptete nunmehr das Gegenteil. Mittlerweile ist das System soweit „informiert“, dass es „korrekte“ von „unkorrekten“ Antworten unterscheidet und sich dabei auch nicht mehr allzu leicht aus der Fassung bringen lässt. Doch wo verlaufen und wer definiert diese Grenzziehungen? Diese Frage beantwortet das System nicht (und kann sie auch nur bedingt abschließend definieren, weil es dem probabilistischen Grundprinzip des zugrundeliegenden Algorithmus widerspricht). „Die Modelle haben kein explizites Wissensmodell; sie wissen nichts und sie verstehen nichts.“ (Mielke/Wolff E-Zeitschrift für Wirtschaftsrecht & Digitalisierung (LRZ) 2023, Rn. 560, 573) – Mittlerweile finden sich in sozialen Medien zahlreiche Hinweise darauf, dass Anwälte oder Justizangehörige leichtfertig oder manipulativ frei erfundene Verweise auf Präjudizien und vermeintlich argumentativ stützende Fachliteratur mit ChatGPT erzeugen lassen (Beispiel: <https://reason.com/volokh/2023/05/27/a-lawyers-filing-is-replete-with-citations-to-non-existent-cases-thanks-chatgpt/>, 26.09.2023).

³³ Mielke/Wolff E-Zeitschrift für Wirtschaftsrecht & Digitalisierung (LRZ) 2023, Rn. 560, 572.

Aufbereitung von Trainingsdaten) über die Qualität der Modell-Ergebnisse und ihren produktiven Einsatz in der Breite entscheiden werden.

- Wenn die Assistenzsysteme nur online („in der Cloud“) effizient arbeiten, verschärft sich das Problem weiter – in zweierlei Hinsicht: Bereits nach kurzer Zeit des ChatGPT-Hypes mehrten sich Berichte von Angehörigen aus IT-Unternehmen oder sogar Sicherheitsbehörden, die teilweise hochsensible Daten an cloudbasierte Assistenzsysteme online übertragen haben, um damit Programmcode zu optimieren oder eigene Texte zu verbessern.³⁴ Auf Twitter werben mittlerweile unzählige Plattformen mit dem Versprechen, bei direktem Zugang zum Mailarchiv oder zum Bankkonto automatisiert Briefvorlagen oder quasi-juristische Inkasso-Schreiben aufzusetzen und damit durch das System erkannte mögliche Geldrückforderungen einzutreiben³⁵. – Für mich ist nach wie vor viel zu intransparent, was die Cloudbieter mit den Daten anfangen. Oft gehen sie in die Verbesserung der Assistenzsysteme ein, wo ggf. hochsensible Informationen dann Teil der Trainingsdaten werden und/oder ‚händisch‘ durch am oder neben dem Verarbeitungsprozess Beteiligte missbraucht werden können. Hinzu kommt ein zweiter Aspekt: Die Sprachmodelle sind bereits jetzt in der Lage, in Echtzeit Online-Daten in ihre Textmodellierung zu integrieren. Geschieht das automatisch und ohne User-Kontrolle, entsteht damit ein offenes Scheunentor für Manipulationsversuche. Angreifer – seien es militärische Propagandaeinheiten, Kriminelle oder Werbeagenturen – könnten dann an geeigneten Stellen strategische Textbausteine platzieren und damit den Textoutput der Assistenzsysteme beeinflussen. Das ist keine Hypothese mehr, erste Studien haben genau das bereits demonstriert³⁶. Neu ist das Prinzip übrigens auch nicht, man kennt es im Kleinen schon lange, etwa unter dem Namen „Google-Bombing“³⁷.

Das Risiko für diese Probleme halte ich intuitiv für ziemlich hoch, weil die Rechtspraxis erstens keine Zeit hat, sich in die technischen Hintergründe solcher Software einzuarbeiten, und zweitens, weil das offensichtliche Versprechen einer erheblichen Effizienzsteigerung viel zu verlockend ist. Denn wer solche Assistenzsysteme dann nicht einsetzt oder nicht praktisch mit ihnen umzugehen weiß, kann mit der Produktivität von konkurrierenden KollegInnen möglicherweise schnell nicht mehr mithalten.

³⁴ So etwa prominent geworden im Falle von Samsung: <https://www.golem.de/news/kuenstliche-intelligenz-samsung-ingenieure-leaken-interne-daten-an-chatgpt-2304-173220.html> (27.09.2023).

³⁵ Das Unternehmen „DoNotPay“ wirbt auf seiner Website etwa: „The DoNotPay app is the home of the world's first robot lawyer. Fight corporations, beat bureaucracy and sue anyone at the press of a button.“ (<https://donotpay.com/>, 10.05.2023).

³⁶ *Carlini/Jagielski/Choquette-Choo/Paleka/Pearce/Anderson/Terzis/Thomas/Tramèr*, Poisoning Web-Scale Training Datasets is Practical, 20.02.2023.

³⁷ *Bager* Diskursmonitor. Glossar zur strategischen Kommunikation in öffentlichen Diskursen 2022.

Angenommen, wir bekommen diese Probleme in den Griff – ist es dann nicht einfach nur eine Frage der Zeit, bis wir bestimmte oder zumindest viele Aspekte der Rechtsarbeit automatisieren können?

IV. Der Richterbot: Bald Alltag oder Dystopie?

Damit komme ich zu meiner vierten und letzten These: *Einen „Richterbot“, der nicht nur Textentwürfe generiert, sondern automatisiert Entscheidungen trifft, kann und darf es nicht geben.*

Das ist natürlich eine provokante These – zumal Prognosen darüber, was nicht gehe oder technisch unmöglich sei, regelmäßig eines Besseren belehrt werden. Aber ich will meine These kurz erläutern und begründen:

Rechtsarbeit ist Textarbeit³⁸, ist eine textbasierte Argumentations- und Entscheidungswissenschaft³⁹. Das ist eine Binsenweisheit unter RechtslinguistInnen. Die Strukturierende Rechtslehre beschreibt die Praktiken ziemlich genau⁴⁰, dazu gehören – sehr verknappt formuliert:

- das Verarbeiten von Eingangsdaten der Streitparteien, zuarbeitenden Behörden, GutachterInnen, ihre Versprachlichung teilweise mündlich, letztlich aber immer schriftlich;
- die rechtsmethodisch geleitete Sortierung und Selektion dieser Sachdaten unter Einbeziehung von juristischem Vorwissen, unter Einbeziehung der zu interpretierenden Normtexte, dogmatischen Literatur und anderen begleitenden Ko(n)texten;
- die argumentative, auf eine revisionssichere Entscheidung hinführende Perspektivierung und Zubereitung aller normativen und fallbezogenen Eingangsdaten im Urteilstext.

Betrachtet man allein diesen Ausschnitt der Rechtsarbeit als Textarbeit, dann gibt es eigentlich keinen Grund zu der Annahme, dass entsprechend trainierte Sprachmodelle nach „Fütterung“ aller fallrelevanten Sprachdaten – also Akten, Normtexte, dogmatische Texte usw. – nicht zu ähnlichen Ergebnissen wie Richter gelangen könnten⁴¹.

In der Rechtsarbeit eine Rolle spielen allerdings auch andere Aspekte, die oft nicht oder nur partiell versprachlicht werden:

³⁸ *Felder*, Juristische Textarbeit im Spiegel der Öffentlichkeit, 2003; *Müller, F./Christensen/Sokolowski*, Rechtstext und Textarbeit, 1997.

³⁹ *Augsberg*, Die Lesbarkeit des Rechts, 2009, S. 1.

⁴⁰ zur Einführung: *Hamann*, in: *Felder/Vogel* (Hrsg.), Handbuch Sprache im Recht, 2017; grundlegend: *Müller, F.* (Fn. 10).

⁴¹ Skeptisch dazu auch: *Gless/Wohlens*, in: *Böse/Schumann/Toepel* (Hrsg.), Festschrift zum 70. Geburtstag von Professor Dr. Dr. h.c. mult. Urs Kindhäuser, 2019, S. 147, 153ff.

- etwa die Einordnung einer Person (Beschuldigter oder Zeuge) anhand von habituellen Markern jenseits der Sprache (z. B. Kleidungsmode, Frisur, Verhalten vor Gericht) und darauf aufbauende Annahmen über deren Glaubwürdigkeit, vergangene und zukünftige Entwicklung etc.;
- eine Rolle spielen auch Annahmen über typische gesellschaftliche Zusammenhänge, die als Inferenzen (lebensweltliches Vorwissen zur Schließung von Wahrnehmungslücken) bei der Sachverhaltsbeurteilung implizit zum Tragen kommen;
- nicht zuletzt relevant ist auch die Herbeiführung von sog. „Deals“ zur Vermeidung langer Prozesse usw.

Mit anderen Worten: Rechtsarbeit ist methodisch strukturierte Textarbeit, aber de facto eben *nicht nur* Textarbeit⁴².

Ein Modell (das nicht nur ein Sprach-, sondern ein multimodales Modell wäre), das in gleichem Maße diese Daten erfassen, ggf. in Echtzeit verarbeiten und in einer Entscheidung automatisiert verdichten könnte, ein solches Modell setzte eine totalitäre Gesellschaft voraus, in der jede soziale Regung erfasst, gespeichert, kategorisiert und berechnet würde. Diese Voraussetzung ist zumindest heute wenig wahrscheinlich.

Aber man muss gar nicht zu einer solchen Dystopie greifen: ich hielte voll-automatisierte „Entscheidungsbots“ – sofern sie funktionierten – selbst bei kleineren Fallgruppen für hoch problematisch, denn auf diese Weise entzöge sich die staatliche Herrschaft jeglicher symbolischer Verantwortungszurechenbarkeit:

„Es geht nicht um Effizienz, sondern es geht um grundlegende Konzepte dazu, anhand welcher Parameter Entscheidungen getroffen werden sollen. Letztlich geht es um eine Diskussion über die Legitimation von Richtern und der Justiz an sich.“⁴³

Ein Richter oder eine Richterin kann personal adressiert, objektiviert und zur Verantwortung bzw. Rechtfertigung verpflichtet werden, ein Sprachmodell nicht, da herrschte im Arendtschen Sinne „Niemand“⁴⁴. Welche Folgen das für das Rechtsgefühl von Rechtsunterworfenen hätte, kann jeder erahnen, der schon einmal kafkaesk mit den Mühlen klassischer Bürokratien zu kämpfen hatte oder auch nur eine Beschwerde in der Schleife eines Voicebot-Telefonats absenden wollte.

Und was wäre mit Bot-Systemen, die zwar keine Richter ersetzen, aber für Profis sowie Rechtsunterworfenen auf Basis von Trainingsdaten, akuten Unterlagen und Fallbeschreibungen Vorhersagen treffen über mögliche Klagerisiken oder -chancen (das Werbeschlagwort dafür ist zum Beispiel „Predictive Analytics“)? Angebote dieser Art finden sich bereits im Internet

⁴² Vogel (Hrsg.), *Recht ist kein Text*, 2017.

⁴³ Gless/Wohlers, in: *Böse/Schumann/Toepel* (Fn. 41), S. 165.

⁴⁴ Narr, *Niemand-Herrschaft*, 2015.

und könnten auch einmal auf den Websites der Justizbehörden Eingang finden, ich halte sie aber aus rechtsstaatlicher Perspektive nicht weniger problematisch als einen Bot-Richter. Denn zum einen ließen sich diese Systeme leicht missbrauchen (siehe oben zu Manipulationsmöglichkeiten), um Klagewege unwahrscheinlicher zu machen, zum anderen entstünde (bzw. verschärfte sich) wohl eine Mehr-Klassen-Justiz, bei der der Geldbeutel (noch stärker) darüber vor-entschiede, ob man entgegen eines Bot-Rates dennoch mit anwaltlichem Beistand vor Gericht zöge.

V. Fazit

Damit ende ich pessimistischer als ich das – als technikbegeisterter Philologe – von Haus aus bin: Die Digitalisierung verändert die juristische Textarbeit schon jetzt in vielfältiger Weise, die rechtsmethod(olog)ische Reflexion und Strukturierung aber hält bislang nicht Schritt mit der Geschwindigkeit dieser Entwicklung. Das Problem ist wie so oft nicht die Technologie, sondern ihr Einsatz. Ihre Regulierung sollte man dabei nicht dem „Silicon Valley“ oder gar nur den großen Tech-Firmen überlassen, denn dort hat man nicht das vielbeschworene „Allgemeinwohl“, sondern die Wahrung eigener Profitmargen vor Augen. Anders lässt sich nicht erklären, dass namhafte IT-Manager bei den Regierungen vorstellig werden und vor den „Gefahren“ ihrer eigenen Technologie warnen, nur um damit die bedrohlich wachsende Konkurrenz in den Griff zu bekommen⁴⁵.

Und wenn der Deutsche Richterbund 2022 fordert, man müsse das Potential von KI mehr ausschöpfen (hier zunächst „zur besseren Bewältigung von Massenverfahren“⁴⁶), ohne auch nur ein Wort über die damit verbundenen ungelösten Fragen zu verlieren, dann ist wohlwollende Skepsis und Innehalten meines Erachtens kein unvernünftiger Ansatz.

⁴⁵ <https://www.heise.de/news/Ein-Risiko-fuer-Menschheit-wie-Pandemien-und-Atomkrieg-KI-Branche-warnt-erneut-9068973.html> (30.05.2023).

⁴⁶ Deutscher Richterbund, Stellungnahme vom Mai 2022, https://www.drj.de/fileadmin/DRB/pdf/Stellungnahmen/2022/DRB_220513_Stn_Nr_1_Massenverfahren.pdf (10.05.2023).