

Kommentar ohne Autor

GPT-Online-Kommentar

Johannes Kruse*

A. Aufbau und Funktionsweise	74	II. Entstehungsprozess	79
B. Technischer Hintergrund	77	C. Bewertung	80
I. Große Sprachmodelle als rechtswissenschaftliche Werkzeuge	77	D. Ausblick	83

Können uns große Sprachmodelle (Large Language Models) das juristische Kommentieren abnehmen? Dies war die entscheidende Frage hinter dem ersten LLM-generierten juristischen Kommentar, der ausdrücklich als Prototyp konzipiert ist. Der online frei zugängliche Kommentar ohne Autor liegt inzwischen sowohl zu Art. 8 GG (kommentar-ohne-autor.coll.mpg.de) als auch, in englischer Sprache, zu Art. 11 EMRK (professor-gpt.coll.mpg.de) vor. GPT kann und soll von Menschen verfasste Kommentare zwar noch (!) nicht vollständig ersetzen. Bereits heute können uns LLMs aber einen substantiellen Teil des Kommentierens abnehmen. Sie können große Mengen von Gerichtsentscheidungen zusammenfassen, analysieren sowie entlang einer vorgegebenen Dogmatik systematisch einordnen.. Dabei deckt der Prototyp die Rechtsprechung nicht nur viel umfassender ab, als alle bisherigen Kommentare (die Art. 11 EMRK Version enthält mehr als 12.000 Rechtsprechungsverweise), sondern ist auch deutlich aktueller (Möglichkeit der automatischen Aktualisierung, sobald eine neue Entscheidung veröffentlicht wird) sowie präziser (es wird stets auf RN verwiesen, wobei die jeweilige Passage verlinkt ist).

A. Aufbau und Funktionsweise

Betrachten wir zunächst Aufbau und Funktionsweise des Kommentars ohne Autor. Am Beispiel der Versammlungsfreiheit aus Art. 8 GG bzw. Art. 11 EMRK wurde erprobt, inwieweit man die Arbeit der Sammlung und Sichtung einschlägiger Entscheidungen, und ihre zusammenfassende Darstellung, dem Computer überlassen kann. In die endgültige Kommentierung sind 126 Entscheidungen des BVerfG bzw. 691 Entscheidungen des EGMR eingeflossen. Beide Kommentare sind ausdrücklich als Prototyp konzipiert. Der gesamte Code (inklusive sämtlicher Prompts) wird über *github* zugänglich gemacht. Auf dieser Grundlage können sich Rechtswissenschaftler*innen künftig schon mit Python-Grundlagenkenntnissen einen eigenen Kommentar ohne Autor erstellen. Vor diesem Hintergrund ist das Projekt auch als Teil einer Open Science Infrastruktur und als Beitrag zu einer offeneren Rechtswissenschaft zu sehen. Die nachstehende Abbildung zeigt die Start- bzw. Übersichtsseite des Kommentars ohne Autor zu Art. 8 GG.

* Dr. Johannes Kruse ist Senior Research Fellow am Max-Planck-Institut zur Erforschung von Gemeinschaftsgütern (Bonn).

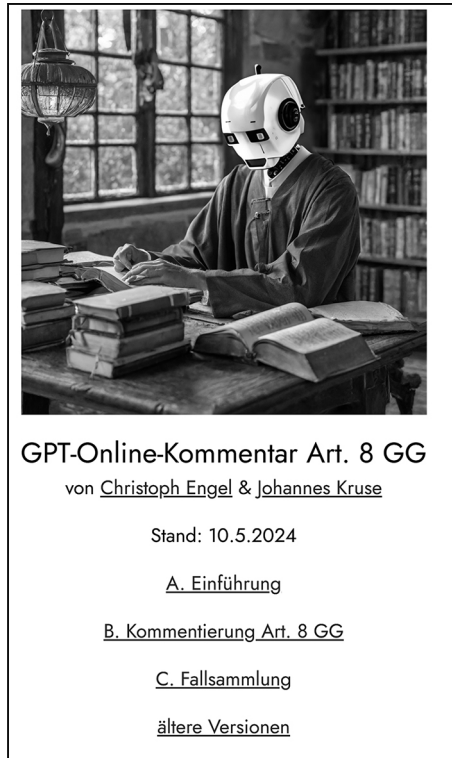


Abbildung 1: Kommentar ohne Autor: Übersicht: Art. 8 GG

Beide Kommentierungen folgen im Aufbau der einschlägigen Dogmatik, also etwa Schutzbereich, Eingriff sowie Rechtfertigung (um die oberste Gliederungsebene der Art. 8 GG Kommentierung zu nennen). Der Kommentar fasst die Rechtsprechung zusammen und ordnet sie entsprechend der grundrechtlichen Dogmatik ein. Abbildung 2 zeigt den Teil der Kommentierung, der sich mit dem Prüfungspunkt Eingriff befasst. Jeder Nachweis besteht aus einem Link auf die zitierte Passage der Entscheidung des BVerfG (siehe Abbildung 3).

Eingriff

Aktualisiert am 14.05.2024

Zusammenfassung

a) Allgemeines

Das Bundesverfassungsgericht versteht unter einem Eingriff in das Grundrecht der Versammlungsfreiheit nach Art. 8 GG jede staatliche Maßnahme, die die Ausübung der Versammlungsfreiheit unmöglich macht oder wesentlich erschwert. Dies umfasst nicht nur Verbote oder Auflösungen von Versammlungen, sondern auch Beschränkungen in der Art und Weise ihrer Durchführung. Der Schutzbereich der Versammlungsfreiheit ist betroffen, wenn eine Versammlung verboten oder aufgelöst wird oder die Art und Weise ihrer Durchführung durch staatliche Maßnahmen beschränkt wird (57_14, 64_15, 84_11). Eingriffe in die Versammlungsfreiheit sind nur zum Schutz gleichgewichtiger anderer Rechtsgüter unter strikter Wahrung der Verhältnismäßigkeit zulässig (118_6, 125_7, 127_7).

b) Anwendungsfälle

Das Bundesverfassungsgericht hat in einer Vielzahl von Fällen Eingriffe in die Versammlungsfreiheit festgestellt. Zu den Eingriffen zählen unter anderem:

- Die Anwendung eisenbahnrechtlicher Vorschriften, die das Grundrecht der Versammlungsfreiheit einschränken, wenn sie einer generell bestehenden Gefahr entgegenwirken und nicht von der Rechtmäßigkeit eines Versammlungsverbots abhängen (1_10).
- Die Annahme eines Verstoßes gegen das Versammlungsgesetz ohne Berücksichtigung der Rechtmäßigkeit eines Versammlungsverbots (2_13).
- Die Verurteilung aufgrund von eisenbahnrechtlichen Vorschriften, die das Grundrecht der Versammlungsfreiheit einschränken, ohne dass die Rechtmäßigkeit eines Versammlungsverbots berücksichtigt wurde (1_10).
- Die polizeiliche Räumung von Sitzblockaden ohne vorherige Auflösungsverfügung (7_13).
- Die Anwendung von Strafvorschriften auf Versammlungsteilnehmer (8_5, 9_3).
- Die Festsetzung von Personen nach einer Versammlung (63_25).
- Die Durchführung von Identitätsfeststellungen und Durchsuchungen im Vorfeld von Versammlungen (64_15).
- Die Verweigerung von Ausnahmegenehmigungen für Versammlungen während der COVID-19-Pandemie (103_7, 106_12, 107_9, 109_6, 110_4, 112_8, 113_6, 114_15, 119_5, 120_9).
- Die Anwendung von Infektionsschutzmaßnahmen auf Versammlungen (103_7, 106_12, 107_9, 109_6, 110_4, 112_8, 113_6, 114_15, 119_5, 120_9).
- Die Erhebung von Gebühren für die Nutzung öffentlicher Räume (55_35).
- Die Anwendung von Versammlungsgesetzen ohne Berücksichtigung der Rechtmäßigkeit eines Versammlungsverbots (2_13).
- Die Verhängung von Versammlungsverböten ohne ausreichende Rechtsgrundlage (59_16, 61_7, 66_5, 68_62, 69_36, 72_10, 77_20, 78_12, 80_9, 81_9, 82_4, 83_3, 84_11, 85_17, 88_14, 90_22, 92_8, 93_7, 97_136, 98_62, 99_17, 100_10).

c) Neue Gesichtspunkte

Das Bundesverfassungsgericht hat in seiner Entscheidung klargestellt, dass bei der Prüfung von Eingriffen in die Versammlungsfreiheit grundsätzlich die Tatsachenfeststellungen und Tatsachenwürdigungen in den angegriffenen Entscheidungen zugrunde zu legen sind. Eine Abweichung hiervon ist nur dann geboten, wenn die getroffenen Tatsachenfeststellungen offensichtlich fehlsam sind oder die Tatsachenwürdigungen unter Berücksichtigung der betroffenen Grundrechtsnormen offensichtlich nicht tragen (127_7). Weiterhin erfordert eine behördliche Verfügung, die auf eine unmittelbare Gefahr für die öffentliche Sicherheit oder Ordnung gestützt ist, tatsächliche Anhaltspunkte, die bei verständiger Würdigung eine hinreichende Wahrscheinlichkeit des Gefahren Eintritts ergeben. Im Rahmen der Folgenabwägung muss das Gericht berücksichtigen, ob die für die Beurteilung der Gefahrenlage herangezogenen Tatsachen unter Berücksichtigung des Schutzgehalts des Art. 8 GG in nachvollziehbarer Weise auf eine unmittelbare Gefahr hindeuten (127_7).

d) Text aller einschlägigen Auszüge

Abbildung 2: Kommentierung: II. Eingriff

Fall 118_6
Aktenzeichen: 1 BvQ 135/20
Beck Online: COVuR 2021 93.0

Eingriff Art. 8 Abs. 1 GG schützt die Freiheit, mit anderen Personen zum Zwecke einer gemeinschaftlichen, auf die Teilhabe an der öffentlichen Meinungsbildung gerichteten Erörterung oder Kundgebung örtlich zusammen zu kommen (vgl. BVerfGE 104, 92 <104>; 111, 147 <154 f>; 128, 226 <250>). Als Freiheit zur kollektiven Meinungskundgabe ist die Versammlungsfreiheit für eine freiheitlich demokratische Staatsordnung konstituierend (vgl. BVerfGE 69, 315 <344 f.>; 128, 226 <250>). In ihrer idealtypischen Ausformung sind Demonstrationen die gemeinsame körperliche Sichtbarmachung von Überzeugungen, bei der die Teilnehmer in der Gemeinschaft mit anderen eine Vergewisserung dieser Überzeugungen erfahren und andererseits nach außen – schon durch die bloße Anwesenheit, die Art des Auftretens und die Wahl des Ortes – im eigentlichen Sinne des Wortes Stellung nehmen und ihren Standpunkt bezeugen (vgl. BVerfGE 69, 315 <345>; 128, 226 <250>). Nach Art. 8 Abs. 2 GG kann dieses Recht für Versammlungen unter freiem Himmel durch Gesetz oder auf Grund eines Gesetzes beschränkt werden. Derartige Beschränkungen sind im Lichte der grundlegenden Bedeutung von Art. 8 Abs. 1 GG auszulegen (vgl. BVerfGE 87, 399 <407>). Eingriffe in die Versammlungsfreiheit sind nur zum Schutz gleichgewichtiger anderer Rechtsgüter unter strikter Wahrung der Verhältnismäßigkeit zulässig (vgl. BVerfGE 69, 315 <349>; 87, 399 <407>). Insbesondere Versammlungsverbote dürfen nur verhängt werden, wenn mildere Mittel nicht zur Verfügung stehen und soweit der hierdurch bewirkte tiefgreifende Eingriff in das Grundrecht aus Art. 8 Abs. 1 GG auch in Ansehung der grundlegenden Bedeutung der Versammlungsfreiheit für das demokratische und freiheitliche Gemeinwesen insgesamt nicht außer Verhältnis steht zu den jeweils zu bekämpfenden Gefahren und dem Beitrag, den ein Verbot zur Gefahrenabwehr beizutragen vermag (vgl. BVerfG, Beschluss der 1. Kammer des Ersten Senats vom 30. August 2020 - 1 BvQ 94/20 -, Rn. 16).

Abbildung 3: Nachweis-Beispiel

B. Technischer Hintergrund

Damit kommen wir zum technischen Hintergrund. Da der Kommentar letztlich von einem Large Language Model (auch LLM oder großes Sprachmodell) verfasst wurde, wollen wir zunächst diese neueste computertechnische Innovation aus einer spezifisch rechtswissenschaftlichen Perspektive betrachten.

I. Große Sprachmodelle als rechtswissenschaftliche Werkzeuge

Im Unterschied zu computergestützten Analyseverfahren, die auf regulären Ausdrücken beruhen, sind LLMs nicht deterministisch, sondern probabilistisch. Sie sind also nicht darauf beschränkt, Wenn-Dann Operationen zu bewältigen und kommen auch mit Aufgaben zurecht, die nicht vollständig definiert sind.¹ Sprachmodelle machen Vorhersagen. Genauer: sie ergänzen angefangene Sätze. Gegeben den Text, den sie bislang erhalten haben: wie lautet die wahrscheinlichste Fortsetzung? Der Anwender steuert Sprachmodelle deshalb mit den Eingaben, die er macht (den sog. Prompts). Ein Prompt braucht dabei nicht aus einem einzigen Satz

¹ Hier und im Folgenden siehe bereits C. Engell/J. Kruse, JZ 2024, S. 997 sowie dies., Professor GPT, Max Planck Institute for Research on Collective Goods, Discussion Paper 2024/14, <https://ssrn.com/abstract=4994131>.

zu bestehen. Die State-of-the-Art LLMs können sehr lange Texte verarbeiten, sogar ein ganzes Buch. Allerdings ist zu berücksichtigen, dass auch unterhalb der Verarbeitungsobergrenze² ab einer gewissen Textlänge die Leistungsfähigkeit abnimmt.³ Bei der Arbeit an dem Kommentar-ohne-Autor zeigte sich, dass GPT-4 Turbo bei sehr langen Entscheidungen einen Teil der für die Auslegung von Art. 8 GG relevanten Passagen übersieht (dem Problem ließ sich mit einer mehrschrittigen Vorselektion begegnen).

Sprachmodelle nutzen maschinelles Lernen. Neue Beobachtungen werden entweder (top down) in Entscheidungsbäume eingeordnet, oder sie werden (bottom up) möglichst verwandten anderen Datenpunkten zugeordnet.⁴ Neuronale Netze sind besonders anspruchsvolle Instrumente für diese Aufgabe. Sie können nicht nur eine hohe Zahl von Dimensionen verarbeiten, sondern diese Dimensionen auch in komplexe Beziehungen zueinander setzen. Transformer übersetzen Inputs (z.B. natürliche Sprache) in lange Ketten von Wahrscheinlichkeiten und nutzen dabei reiche Trainingsdatensätze. Auf diese Weise wird die lokale Klassifikationsaufgabe eingebettet in das „Wissen“, das die Architektur zuvor erworben hat, bearbeitet. Sprachmodelle setzen auf all diesen Elementen auf und erweitern sie um eine generative Komponente. Der Output besteht nicht mehr bloß aus der Zuordnung eines Datenpunkts zu einer Klasse. Vielmehr kann das Modell Texte schreiben (oder Bilder generieren, oder Töne ausgeben). Die Größe eines Sprachmodells bezieht sich sowohl auf die Anzahl seiner Parameter als auch auf den Umfang seines Trainingskorpus. Sog. große Sprachmodelle (Large Language Models oder LLMs) sind Modelle, die Milliarden von Parametern enthalten und mit riesigen Korpora trainiert werden.

Große Sprachmodelle sind bereits für eine Vielzahl juristischer Aufgaben eingesetzt worden:⁵ von der juristischen Ausbildung,⁶ über die empirische Rechtsforschung⁷ bis hin zur juristischen Praxis.⁸ Mit durchaus beachtlichem Erfolg. So ist es jüngst etwa gelungen, unter Verwendung von GPT-4 die Fragen des US-amerikanischen Bar Exam mit einer durchschnittlichen Richtigkeitsquote zu beantworten, die in sämtlichen Bundesstaaten zum Bestehen ausgereicht hätte.⁹ Auch zur Zusammenfassung von juristischen Texten sind LLMs bereits erfolgreich eingesetzt worden.¹⁰

2 Für GPT-4o liegt sie bei 128.000 Tokens, was ungefähr der Zahl der Worte entspricht.

3 Siehe auch N. Liu et al., TACL 12 2024, S. 157 zum sog. Lost in the Middle“-Effekt.

4 Hier und im Folgenden C. Engell/J. Kruse, JZ 2024, S. 997 m.w.N.

5 Zum Ganzen etwa S. Kapoor/P. Henderson/A. Narayanan, Promises and pitfalls of artificial intelligence for legal applications, Preprint v. 10.1.2024, arXiv:2402.01656.

6 J. Choi/K. Hickman et al., J. Legal Educ. 71 2021, S. 387.

7 J. Choi, JITE 180, 2024, S. 214; M. Livermore/F. Herron/D. Rockmore, JITE 180, 2024, S. 244.

8 I. Rodgers/J. Armour/M. Sako, Annual Review of Law and Social Science 19, 2023, S. 299.

9 D. Katz/M. Bommarito et al., Phil. Trans. R. Soc. A. 382, 2024, 20230254.

10 J. Gesnoui/Y. Tannier et al., LLaMandement: Large Language Models for Summarization of French Legislative Proposals, Preprint v. 29.1.2024, arXiv:2401.16182.

de Faria/Xie/Steffek konnten jüngst eindrucksvoll zeigen, wie sich GPT-4 zur Informationsextraktion aus Gerichtsentscheidungen einsetzen lässt.¹¹

II. Entstehungsprozess

Entstanden ist der Kommentar aus einer Kombination von Python (für alle deterministischen Elemente, d.h. für solche, die sich in exakte „Wenn-Dann“ Ketten auflösen lassen) und GPT-4. Das Schreiben eines ganzen Kommentars ist ein komplexerer (technischer) Prozess, der sich mit dem populären „ChatGPT“ nicht bewältigen ließe. Ferner wäre der Prozess der Kommentarherstellung auf diese Weise für Dritte nicht mehr nachvollziehbar gewesen. Gerade vor diesem Hintergrund wurde für die Interaktion mit dem Sprachmodell die API (Application Programming Interface) genutzt, wobei Anfragen aus einem in der Programmiersprache Python geschriebenen Programm gestellt wurden. Im Einzelnen ist der Kommentar in den nachfolgenden Schritten entstanden (hier für die Art. 8 GG Kommentierung).

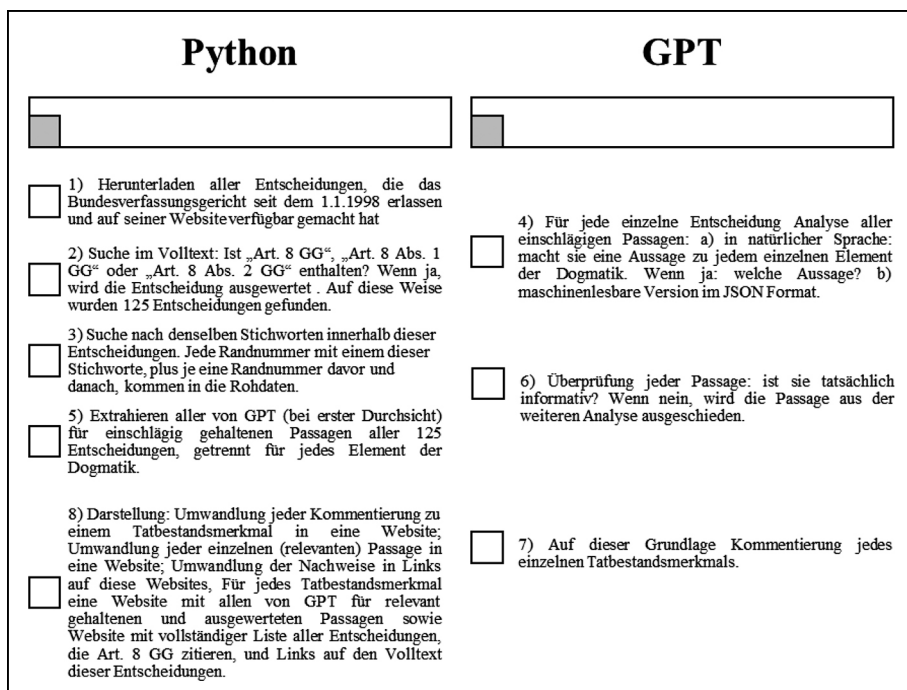


Abbildung 4: Entstehungsprozess im Überblick

11 J. de Faria/H. Xie/F. Steffek, Automatic Information Extraction From Employment Tribunal Judgments Using Large Language Models, Preprint v. 19.5.2024, arXiv:2403.12936.

Der dabei verwendete Prompt war recht umfangreich und kann hier lediglich auszugsweise wiedergegeben werden.¹²

Aufgabe: Zusammenfassung der Rechtsprechung zu einem ganz bestimmten Tatbestandsmerkmal

Mit diesem Prompt erhalten Sie Auszüge aus Entscheidungen des Bundesverfassungsgerichts, die sich mit einem Element der Dogmatik des Grundrechts der Versammlungsfreiheit auseinandersetzen. Auf welches Element sich diese Auszüge beziehen, steht am Beginn des Dokuments, das die Auszüge enthält. Es ist allerdings möglich, dass der Auszug das fragliche Element der Dogmatik nur am Rande berührt. Bitte fassen Sie wirklich nur die Interpretation dieses einen Elements der Dogmatik zusammen, so wie das Gericht dieses Element interpretiert.

Struktur der Grundrechtsdogmatik

Im Folgenden erläutere ich zunächst noch einmal die Struktur der Grundrechtsdogmatik. Diese Erläuterungen sollen Ihnen das Verständnis der Entscheidung erleichtern.

Grundrechte prüft das Bundesverfassungsgericht in drei Schritten: 1. dem Schutzbereich, 2. der Frage nach einem Eingriff, und 3. der Möglichkeit seiner Rechtfertigung.

Vorgaben für die zusammenfassende Darstellung

Bitte schreiben Sie einen Text, der die Auszüge aus den Entscheidungen zusammenfasst. Wenn mehrere Entscheidungen die gleiche oder eine ähnliche Aussage treffen, machen Sie das bitte deutlich. Bitte belegen Sie alle Aussagen mit der Ordnungsnummer, nach der Art von "12_25". (Zum Verständnis: die Ordnungsnummer setzt sich aus zwei Elementen zusammen <EntscheidungsNummer>_<RandNummer>.) Nennen Sie ggf. auch mehrere solcher Belege, wenn sich die gleiche Aussage mehrfach findet. Ordnen Sie die Zusammenfassung nach sachlichen Zusammenhängen, nicht nach Ordnungsnummern.

Bitte gliedern Sie die Zusammenfassung wie folgt:

[Ende Auszug Prompt]

Abbildung 5: Der Prompt zur Art. 8 GG Kommentierung (Auszug)

GPT benötigte also nicht nur sehr umfangreiche und präzise Anweisungen; vielmehr musste ihm zunächst einmal die Grundrechtsdogmatik vermittelt werden.

C. Bewertung

Dies führt uns zu der zentralen Frage: Wie sind die Ergebnisse zu bewerten? Wie schlägt sich der Kommentar ohne Autor im Vergleich zu den traditionellen (d.h. menschengemachten) Kommentaren? Insoweit lassen sich zwei Gesichtspunkte unterscheiden: die Verlässlichkeit LLM-generierter juristischer Inhalte im Allgemeinen und die Vor- und Nachteile des Kommentars ohne Autor im Besonderen.

Mit Blick auf die generelle Verlässlichkeit derartiger Tools ist zu sagen, dass LLMs zwar mit jedem Entwicklungsschritt besser werden, aber nach wie vor Fehler ma-

12 Für den vollständigen Prompt zur Art. 8 GG Kommentierung siehe C. Engel/J. Kruse, JZ 2024, S. 997 (S. 1003 ff.).

chen (sog. Halluzinationen).¹³ Die Fehler- bzw. Halluzinationsrate ließe sich aber auf ein Minimum reduzieren, indem man Prompt Engineering für eine effektive Steuerung nutzt und dem Modell das nötige Hintergrundwissen zur Verfügung stellt.¹⁴ Gänzlich vermeiden lassen, werden sich Halluzinationen aber wohl nie.

Dementsprechend enthält auch der Kommentar ohne Autor (zumindest zu Art. 8 GG) vereinzelt Fehler. So wird dem BVerfG etwa eine Rechtsansicht zugeschrieben, die allein vom Beschwerdeführer vertreten wurde und der sich das BVerfG gerade nicht angeschlossen hat. Derartige Fehler ließen sich aber ohne Weiteres mittels einer zusätzlichen Kontroll-Schleife beseitigen. Deutlich positiver fällt das Bild hingegen mit Blick auf die Art. 11 EMRK Kommentierung aus. Zumindest ließen sich dort keine entsprechenden Fehler feststellen. Die höhere Qualität der Inhalt dürfte vor allem darauf zurückzuführen sein, dass mit GPT-4o (gegenüber GPT-4-1106-preview für die Art. 8 GG Kommentierung) ein deutlich leistungsfähigeres Sprachmodell zum Einsatz kam.

Vergleicht man den Kommentar ohne Autors zu Art. 8 GG mit bestehenden Grundgesetzkommentaren, so ist zunächst darauf hinzuweisen, dass es sich bei Lichte besehen lediglich um einen Kommentar „light“ handelt.¹⁵ GPT fasst zwar die einschlägige Rechtsprechung zusammen, analysiert sie und ordnet sie entlang der grundrechtlichen Dogmatik ein. Traditionelle Kommentare leisten darüber hinaus aber auch noch eine kritische Bewertung der Entscheidungen und unterbreiten Vorschläge zu deren Weiterentwicklung.¹⁶ Überdies sind in den Kommentar ohne Autor lediglich Entscheidungen des Bundesverfassungsgerichts eingegangen. Die Literatur blieb hingegen vollständig unberücksichtigt. Dies hat aber keine technischen, sondern rein urheberrechtliche Gründe.

Der Kommentar ohne Autor hat gegenüber seinen menschengemachten Vorgängern aber auch handfeste Vorteile. Ein wesentlicher Vorteil liegt darin, dass die Rechtsprechung deutlich umfassender sowie präziser verarbeitet wird. Blicken wir zunächst auf Umfang bzw. Tiefe der Rezeption der Rechtsprechung. Anhand von zwei Fragen: Wie viele verschiedene BVerfG-Entscheidungen werden mindestens einmal zitiert? Wie häufig werden die zitierten Entscheidungen jeweils zitiert? Die

13 Vgl. etwa M. Dahl/V. Magesh et al., JLA 16, 2024, S. 64 sowie A. Deroy/K. Ghosh/S. Ghosh, How Ready are Pre-trained Abstractive Models and LLMs for Legal Case Judgement Summarization?, Preprint v. 14.6.2023, arXiv:2306.01248, S. 2.

14 M. Dahl/V. Magesh et al., JLA 16, 2024, 64, 76 bzw. S. Bsharat/A. Myrzakhan/Z. Shen, Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4, Preprint v. 18.1.2024, arXiv:2312.16171, S. 7; J. Wei/X. Wang et al., Advances in Neural Information Processing Systems 35 (2022), S. 24824; T. Kojima/S. Shane et al., Advances in Neural Information Processing Systems 35 (2022), S. 22199.

15 C. Engel/J. Kruse, JZ 2024, S. 997 (1006 f.).

16 Vgl. etwa D. Käßle-Lamparter, Welt der Kommentare, Tübingen 2016, S. 329 ff.

erste Zahl ist ein Anhalt für die Zitierbreite, die zweite Zahl für die Zitiertiefe.¹⁷ Dieser Vergleich führt zu dem folgenden Ergebnis.¹⁸

	<i>Dreier</i>	<i>Dürig/ Herzog/ Scholz</i>	<i>Huber/ Voßkuhle</i>	<i>Sachs</i>	<i>BeckOK</i>	<i>GPT</i>
Anzahl Entscheidungszitate	175	84	155	196	142	553
Anzahl zitatierter Entscheidungen	51	28	45	43	55	126 (88) ¹⁹

Das Ergebnis ist recht eindeutig: GPT berücksichtigt viel häufiger Judikate des BVerfG, arbeitet also rechtsprechungsbezogener; zumal, wenn man bedenkt, dass der Umfang der Kommentierung (d.h. die Textmenge) bei den verfügbaren (Vergleichs-)Kommentaren (deutlich) größer ist. Überdies bildet GPT die Judikatur auch umfassender ab, berücksichtigt also mehr Entscheidungen. Die Art. 11 EMRK Version kommt sogar auf 12.254 Entscheidungsitate und 572 zitierte Entscheidungen.²⁰ Der Kommentar ohne Autor ist auch deutlich präziser als alle verfügbaren Kommentare, weil er stets auf die Randnummern der einschlägigen Entscheidungen verweist.²¹ Ferner ist er auch deutlich aktueller. Die Funktion „living document“ ist im Prototyp des Kommentars ohne Autor bereits implementiert. Damit ist es möglich, den gesamten Kommentar sofort zu aktualisieren, sobald eine neue Entscheidung des BVerfG zu Art. 8 GG veröffentlicht wird (bzw. eine Entscheidung des EGMR zu Art. 14 EMRK). Überdies ergeben sich auch weitgehende Personalisierungsmöglichkeiten: Ein Kommentar zur Versammlungsfreiheit in NRW (regionale Versionierung) ist ebenso möglich wie zum Recht der wehrhaften Demokratie (Querschnittskommentar). Schließlich soll noch auf eine Eigenschaft hingewiesen werden, durch welche der Kommentar ohne Autor zu Art. 11 EMRK gegenüber seiner Vorgängerversion (zu Art. 8 GG) hervorsticht: Sein dogmatisches Innovationspotential. Er schreibt nicht bloß die bisherige Dogmatik des EGMR fort, sondern fügt neuartige dogmatische (Sub-)Kategorien ein. Das dabei zum Ausdruck kommende Innovationspotential deckt sich mit der neueren For-

17 J. Kruse, AcP 224, 2024, S. 38 (60); ders. NZKart 2023, 138 (141); C. Coupette, Juristische Netzwerkforschung, Tübingen 2019, S. 267.

18 In den nachstehenden Vergleich gehen folgende Kommentare ein (jeweils die Kommentierung zu Art. 8 GG): Dreier, GG, 4. Auflage, Tübingen 2023 (Bearbeiter: B. Kaiser); Dürig/Herzog/Scholz, GG, 102. EL, München 2023 (Bearbeiter: O. Depenheuer); Huber/Voßkuhle, GG 8. Aufl., München 2024 (Bearbeiter: C. Gusy); Sachs, GG 9. Aufl., München 2021 (Bearbeiter: W. Höfling) sowie Epping/Hillgruber, BeckOK GG, 57. Ed. 1/2024 (Bearbeiter: J. Schneider).

19 126 umfasst neben den im Kommentierungstext enthaltenen Verweisen (88) auch die lediglich in einer der Zusammenstellungen aller relevanten Passagen enthaltenen Entscheidungen.

20 C. Engel/J. Kruse, Max Planck Institute for Research on Collective Goods, Discussion Paper 2024/14, <https://ssrn.com/abstract=4994131>, S. 12.

21 Vgl. C. Engel/J. Kruse, JZ 2024, S. 997 (1006) für die Randnummernverweisquote der bisherigen Kommentare.

schung zur Kreativität großer Sprachmodelle.²² Diese Modelle sind zwar nicht vollständig kreativ im menschlichen Sinne, können aber neuartige Inhalte, d.h. Innovationen hervorbringen. Dieses Innovationspotential lässt sich (innerhalb gewisser Grenzen) sogar steuern. Eine Studie konnte zeigen, dass eine Erhöhung der temperature-Einstellung zu mehr neuartigen Ergebnissen führen kann.²³

D. Ausblick

LLM-generierte Kommentare können und sollen menschengemachte Kommentare zwar (noch) nicht ersetzen, die Arbeit des Kommentierens aber deutlich erleichtern. Die (so mühsame) Sammlung und Sichtung einschlägiger Entscheidungen sowie deren zusammenfassende Darstellung kann man nämlich bereits heute einem Sprachmodell überlassen. Auf diese Weise werden Einstiegshürden gesenkt und man gelangt zu einem offeneren Kommentarwesen.

Mit dem GPT-Kommentar-Tool²⁴ können sich Rechtswissenschaftler*innen künftig schon mit Python-Grundlagenkenntnissen ihren eigenen Kommentar ohne Autor erstellen. Insoweit trägt das Projekt auch zu einer Open Science Infrastruktur in den Rechtswissenschaften bei. Wird die Kommentarerstellung deutlich einfacher bzw. kommt sie mit geringeren (finanziellen sowie personellen) Ressourcen aus, so stellt dies gerade für OA-Kommentar-Projekte eine substantielle Erleichterung dar.

22 Vgl. nur W. Orwig/E. Edenbaum et al., The Language of Creativity: Evidence from Humans and Large Language Models, *The Journal of Creative Behavior* 58, 2024, S. 128.

23 M. Peeperkorn/T. Kouwenhoven et al., Is Temperature the Creativity Parameter of Large Language Models?, Preprint v. 1.5.2024, arXiv:2405.00492.

24 Demnächst über GitHub verfügbar: <https://github.com/J0hKruse/GPT-Online-Kommentar>.