

Schriften zum Daten-, Informations- und Medienrecht

Boris P. Paal | Tristan Radtke

Training eines Sprachmodells in der Justiz

Urheberrechtliche und datenschutzrechtliche
Anforderungen



Nomos

**Schriften zum Daten-, Informations-
und Medienrecht**

Herausgegeben von
Prof. Dr. Boris P. Paal, M.Jur.

Band 84

Boris P. Paal | Tristan Radtke

Training eines Sprachmodells in der Justiz

Urheberrechtliche und datenschutzrechtliche
Anforderungen



Nomos

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

1. Auflage 2026

© Boris P. Paal | Tristan Radtke

Publiziert von

Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden
www.nomos.de

Gesamtherstellung:

Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden

ISBN (Print): 978-3-7560-2041-6

ISBN (ePDF): 978-3-7489-5581-8

DOI: <https://doi.org/10.5771/9783748955818>



Onlineversion
Nomos eLibrary



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz.

Über das Gutachten

Diesem Werk zugrunde liegt ein Rechtsgutachten, das als selbstständige Gutachterleistung durch Herrn Prof. Dr. Boris P. Paal, M.Jur. (Oxford), im Auftrag des Bayerischen Staatsministeriums der Justiz erstellt wurde.

Über die Autoren des Werks

Prof. Dr. Boris P. Paal, M.Jur. (Oxford), ist Inhaber des Lehrstuhls für Law and Regulation of the Digital Transformation an der School of Social Sciences and Technology, Department of Governance, Technische Universität München.

Dr. Tristan Radtke, LL.M. (NYU), ist Akademischer Rat a.Z. und Habilitand ebenda.

Hinweis zu Abkürzungen

Abkürzungen erfolgen nach *Kirchner*, Abkürzungsverzeichnis der Rechts-sprache, 11. Aufl. 2024.

Inhaltsverzeichnis

A. Sachverhalt	17
I. Projekt Generatives Sprachmodell für die deutsche Justiz (GSJ-Projekt)	17
II. Tatsächliche Annahmen – Thematische Aus- und Eingrenzungen	18
III. Technische Grundzüge des Sprachmodells und späteren KI- Systems	20
B. Gutachtenauftrag und Fragestellungen	23
C. Executive Summary	25
D. Datenschutzrechtliche Bewertung	35
I. Vorgang der Anonymisierung	35
1. Personenbezug und Identifizierbarkeit	35
a. „Beziehen“	36
b. Identifizierbarkeit	37
aa. Relatives Konzept der Identifizierbarkeit	37
bb. Anforderungen an die heranzuziehenden Mittel	39
cc. Mittelbarer Personenbezug oder Identifizierbarkeit	42
dd. Relatives Konzept der Identifizierbarkeit in zeitlicher Hinsicht	44
c. Zwischenergebnis	44
2. Überblick zu den Anonymisierungsverfahren	45
a. Löschung von Merkmalen zur Identifizierung durch Auslassung oder Ersetzung	45
b. Weitere Anonymisierungstechniken und Maßnahmen i.w.S.	46
3. Anforderungen an die Anonymisierung unveröffentlichter Gerichtsentscheidungen	47
a. Typische identifizierende Merkmale	47

Inhaltsverzeichnis

b.	Maßstab für Anonymisierungsverfahren	48
aa.	Berücksichtigung von weiteren Mitteln zur Identifizierung	48
bb.	Identifizierung der zu anonymisierenden Informationen	49
(1)	Untersuchung auf Wortebene bzw. Zeichenebene	49
(2)	Untersuchung auf Entscheidungsebene – Cluster von Gerichtsentscheidungen	50
cc.	Training eines Sprachmodells als mögliche Anonymisierungstechnik	51
c.	Nach allgemeinem Ermessen wahrscheinlich genutzte Mittel	52
d.	Schlussfolgerungen für die Anwendung von Anonymisierungstechniken auf Gerichtsentscheidungen	53
II.	Vergleich mit den Maßstäben zur Veröffentlichung von Gerichtsentscheidungen	54
1.	Anwendbarkeit der Maßstäbe der DSGVO	54
2.	Anonymisierungspraxis in Nordrhein-Westfalen bis zum Jahr 2021	55
3.	Jüngere Anonymisierungspraxis in Nordrhein-Westfalen	55
4.	Abgleich im Hinblick auf die Einhaltung der Anforderungen an die Anonymisierung unter der DSGVO	56
5.	Rechtmäßigkeit der Verarbeitung unter der DSGVO	58
a.	Verwaltungsvorschrift als Rechtsgrundlage	58
b.	Veröffentlichung (teil-)anonymisierter Gerichtsentscheidungen gestützt auf Art. 6, 9 f. DSGVO	58
c.	Keine Haftungsprivilegierung nach Art. 4 ff. Digital Services Act	62
d.	Sonderregelung für die Aufsicht über die justizielle Tätigkeit	63
e.	Verschulden, Art. 82; Aufsichtsmaßnahmen, Art. 58 und 83 DSGVO	63
6.	Zwischenergebnis	64

III. (Teil-)Anonymisierung unveröffentlichter Gerichtsentscheidungen durch das Erlanger Tool	64
1. Einhaltung der Anonymisierungsanforderungen im Allgemeinen	65
2. Wahrscheinlichkeit der Identifizierung nach allgemeinem Ermessen anhand bearbeiteter Gerichtsentscheidungen	66
a. Veröffentlichung der bearbeiteten Gerichtsentscheidung	66
b. Einsatz zum KI-Training im GSJ-Projekt	67
aa. KI-Training als Anonymisierungsmaßnahme	67
bb. Versehentliche Ausgabe von Trainingsdaten	68
cc. Gezielte Extraktion von Trainingsdaten	69
IV. Anwendbarkeit des Datenschutzrechts auf ein (veröffentlichtes) Sprachmodells	69
1. Sprachmodell einschließlich seiner Gewichte als Speicherung personenbezogener Daten	69
a. Training i.w.S.	70
b. Speicherung des Sprachmodells	71
aa. Differenzierung zwischen Kennungen und sonstigen identifizierenden Merkmalen	73
bb. Reichweite des Personenbezugs – Mischdatensätze	73
cc. (Wahrscheinliche) Identifizierbarkeit	74
(1) Dogmatische Bedenken mit Blick auf das Zusammenfallen einer Verarbeitung und der Heranziehung von Identifizierungsmitteln	75
(2) Begünstigende Faktoren für eine Identifizierung	76
(3) Vorbehalt der wahrscheinlichen Identifizierung nach allgemeinem Ermessen	76
dd. Informationsgehalt	78
(1) Abstraktheit	78
(2) Unsicherheit unter Berücksichtigung von Halluzinationen	79
ee. Speicherung	79
(1) Identifizierung im Zusammenhang mit der Speicherung	80
(2) Aufbewahrung als Ausgangspunkt für eine Reproduktion	81

Inhaltsverzeichnis

ff. Zwischenergebnis	81
2. Verarbeitung durch Einsatz des Sprachmodells im Einzelfall	82
3. Anforderungen an die Veröffentlichung des Sprachmodells mit Blick auf die Anonymisierung	83
a. Zurverfügungstellung des KI-Systems	84
b. Veröffentlichung des trainierten Sprachmodells (Open Source)	84
c. Veröffentlichung des Trainingskorpus (Open Source)	85
V. Datenschutzrechtliche Anforderungen im Fall des Vorliegens personenbezogener Daten	85
1. Relevante Verarbeitungen	85
2. Verhältnis zum mitgliedstaatlichen Recht	86
3. Verantwortlichkeit	87
a. Anforderungen	87
b. Einordnung der Rollen der Ministerien, der ausführenden Stellen sowie der anwendenden Justizbediensteten und Richter	88
aa. (Teilweise) Anonymisierung	89
bb. KI-Training	90
cc. Speicherung des KI-Modells	91
dd. Einsatz des KI-Systems	91
ee. Veröffentlichung des KI-Modells oder KI-Systems	92
4. Rechtsgrundlage nach Art. 6 DSGVO	93
a. Einschränkungen aufgrund des Vorbehalts des Gesetzes (Art. 6 Abs. 1 UAbs. 2 DSGVO)	93
b. Überblick über die Rechtsgrundlagen aus Art. 6 Abs. 1 UAbs. 1 DSGVO	95
c. Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO (öffentliches Interesse)	95
aa. (Teil-)Anonymisierung der Gerichtsentscheidungen	96
bb. Weiterverarbeitung teilanonymisierter Entscheidungen: Training i.w.S., Speicherung und Einsatz des Sprachmodells	98
cc. Veröffentlichung des Sprachmodells	101
d. Art. 6 Abs. 1 UAbs. 1 lit. c DSGVO (rechtliche Pflicht)	102

e. Zweckänderung (Art. 5 Abs. 1 lit. b, Art. 6 Abs. 4 DSGVO)	103
aa. Verhältnis zu Art. 6 Abs. 1 DSGVO	103
(1) Art. 6 Abs. 4 DSGVO als ergänzende Regelung	104
(2) Art. 6 Abs. 4 DSGVO als eigenständige Rechtsgrundlage	104
(3) Berücksichtigung der EuGH-Rechtsprechung	106
(4) Zwischenergebnis	107
bb. Kompatibilität im Einzelfall	107
(1) Rechtsvorschrift (Hs. 1)	108
(2) Kompatibilitätstest (Hs. 2)	108
cc. Schlussfolgerungen für das untersuchungsgegenständliche Projekt	110
5. Rechtsgrundlagen für besondere Datenkategorien nach Art. 9, 10 DSGVO	110
a. Relevanz für das untersuchungsgegenständliche Projekt	111
b. Personenbezogene Daten über politische Meinungen, Gesundheit und sexuelle Orientierung (Art. 9 DSGVO)	112
aa. Art. 9 Abs. 2 lit. e DSGVO (offensichtliche Öffentlichmachung)	112
bb. Art. 9 Abs. 2 lit. f DSGVO (Rechtsverteidigung und justizielle Tätigkeit der Gerichte)	113
cc. Art. 9 Abs. 2 lit. g DSGVO (Öffnungsklausel für ein erhebliches öffentliches Interesse)	114
dd. Art. 9 Abs. 2 lit. j DSGVO (wissenschaftliche Forschungszwecke)	115
c. Personenbezogene Daten über strafrechtliche Verurteilungen und Straftaten (Art. 10 DSGVO)	116
d. Verhältnis zu Anforderungen an die Zweckänderung (Art. 5 Abs. 1 lit. b, Art. 6 Abs. 4 DSGVO)	117
e. Schlussfolgerungen für das untersuchungsgegenständliche Projekt	118
6. Weitere Pflichten aus der DSGVO	118
a. Informationspflichten (Art. 13, 14 DSGVO)	119
b. (Weitere) Betroffenenrechte und Datenschutzgrundsätze	120
aa. Auskunft und Datenkopie (Art. 15 DSGVO)	121

Inhaltsverzeichnis

bb. Berichtigung und Mitteilungspflicht (Art. 16 und 19 DSGVO)	122
cc. Löschung und Mitteilungspflicht (Art. 17 und 19 DSGVO), auch im Zusammenhang mit einem Widerspruch (Art. 21 DSGVO)	123
dd. Einschränkung der Verarbeitung und Mitteilungspflicht (Art. 18 und 19 DSGVO)	125
c. Datenschutzfolgenabschätzung und Pflicht zur Implementierung technischer und organisatorischer Maßnahmen	126
d. Grundsatz der Speicherbegrenzung	127
7. Pflichten unter der JI-RL	128
8. Ergänzende Anforderungen mit Blick auf Eingabedaten	128
VI. Besonderheiten bei der Einbeziehung von Aktenauszügen	129
1. Inhalte von gerichtlichen Akten	129
2. Rechtliche Leitplanken für die (Teil-)Anonymisierung der Aktenauszüge	130
a. Kein Veröffentlichungsstandard	130
b. Quantitativ: Recall	130
c. Qualitativ: Clusterung	132
VII. Exkurs: Eigene Forschungszwecke der ausführenden Stellen	133
1. (Gemeinsame) Verantwortlichkeit der ausführenden Stellen	134
2. Rechtsgrundlage nach Art. 6, 9 DSGVO für die Verarbeitung	135
a. Übermittlung an die ausführenden Stellen	136
aa. Art. 6 Abs. 1 UAbs. 1 lit. e, Abs. 2, 3 DSGVO i.V.m. Art. 5 Abs. 1 S. 1 Nr. 2 BayDSG bzw. § 8 Abs. 2 Nr. 2 DSG NRW	137
bb. Art. 6 Abs. 1 UAbs. 1 lit. e, Abs. 3 DSGVO i.V.m. Art. 4 Abs. 1 BayDSG bzw. § 3 Abs. 1 DSG NRW	137
cc. Art. 6 Abs. 1 UAbs. 1 lit. e, Abs. 3 DSGVO i.V.m. Hochschulgesetzen der Länder	138
dd. Art. 6 Abs. 1 UAbs. 1 lit. e, Abs. 3 DSGVO i.V.m. § 17 Abs. 1 DSG NRW bzw. Art. 4 Abs. 1 BayDSG i.V.m. BayHIG oder Erst-Recht-Schluss aus Art. 8 Abs. 1 S. 1 Nr. 5 BayDSG i.V.m. Art. 6 Abs. 2 Nr. 3 lit. c BayDSG	139

ee. Besonderheiten mit Blick auf besondere Kategorien personenbezogener Daten nach Art. 9 DSGVO	141
b. Weiterverarbeitung durch die ausführenden Stellen	141
3. Datenschutzrechtliche Anforderungen im Übrigen	142
4. Zwischenergebnis	142
VIII. Ableitung von datenschutzrechtlichen Handlungsempfehlungen	143
1. Datenschutzrechtliche Ausgangslage in drei Stufen: Anonymisierung – Rechtfertigung nach Art. 6 DSGVO – Rechtfertigung nach Art. 9, 10 DSGVO	143
2. Anforderungen an die Zusammenstellung des Trainingskorpus aus Gerichtsentscheidungen und Aktenauszügen (insbesondere Clusterung)	144
3. Anforderungen an das Anonymisierungstool (insbesondere Recall-Wert)	146
4. Anforderungen an das KI-System und dessen Einsatz (insbesondere System Prompt und Ausgabefilter)	146
5. Anforderungen an die Veröffentlichung des KI-Modells oder KI-Systems	147
6. Beachtung weiterer datenschutzrechtlicher Pflichten	148
E. Urheberrechtliche Anforderungen an die Verwendung von Aktenauszügen zu Trainingszwecken	151
I. Urheberrechtlicher Schutz und Zuordnung	151
1. Schriftsätze	151
a. Schutz als Werk nach § 2 UrhG	151
b. Kein Ausschluss nach § 5 Abs. 1 UrhG	154
c. Verwandte Schutzrechte	155
d. Zwischenergebnis	155
2. Aktenauszüge im Übrigen	155
II. Nutzung innerhalb der Justiz	156
1. Betroffene Verwertungsrechte	157
a. Zusammenstellung des Trainingskorpus und Vervielfältigungen im Trainingsprozess	157
aa. Vervielfältigungen im Einzelnen	158

Inhaltsverzeichnis

bb. Vervielfältigung bei möglicher Rekonstruierbarkeit aus abgeleiteten Textformaten oder anhand eines Sprachmodells	159
b. Einsatz und Ausgabe des KI-Systems	163
c. Bearbeitungen des Werkes im Rahmen des Trainingsprozesses und des Einsatzes des KI-Systems	164
d. Zwischenergebnis	164
2. Zulässigkeit aufgrund von Nutzungsrechtseinräumungen oder Schrankenbestimmungen	165
a. Nutzungsrechtseinräumungen	165
b. Rechtspflege (§ 45 UrhG)	166
c. Text und Data Mining (§§ 44b UrhG und 60d UrhG)	167
aa. KI-Training als Text und Data Mining	168
bb. KI-Einsatz und Ausgabe als Text und Data Mining	171
cc. Anwendbarkeit des § 44b Abs. 2 S. 1 UrhG	172
(1) Intertemporale Anwendbarkeit	172
(2) Vorbehalt der Vervielfältigungen zu Zwecken des Text und Data Mining	172
(a) Form des Vorbehalts	173
(b) Rechtsinhaber	174
(c) Rechtmäßig zugängliche Werke	175
(d) Weitere Anforderungen	175
(3) Zwischenergebnis	176
dd. Anwendbarkeit des § 60d UrhG	177
d. Vorübergehende Vervielfältigungshandlungen (§ 44a UrhG)	179
e. Zitate (§ 51 UrhG)	181
f. Wissenschaftliche Forschung (§ 60c UrhG)	181
3. Einbindung weiterer Stellen	182
4. Ableitung von Einsatzbedingungen	183
III. Besonderheiten bei der Veröffentlichung des Sprachmodells	185
IV. Exkurs: Eigene Forschungszwecke der ausführenden Stellen	186
1. Vervielfältigungen im Rahmen des GSJ-Projekts	186
2. Vervielfältigungen über das GSJ-Projekt hinaus	187
V. Schlussbetrachtung aus urheberrechtlicher Perspektive	187
1. Anforderungen an die Zusammenstellung der Trainingsdaten	187
2. Anforderungen an das KI-System und dessen Einsatz	188

Inhaltsverzeichnis

3. Anforderungen an die Veröffentlichung des Sprachmodells	189
F. Literaturverzeichnis	191

