

D. Datenschutzrechtliche Bewertung

Zunächst ist der Sachverhalt aus datenschutzrechtlicher Sicht zu bewerten. Zu diesem Zweck sind in den Blick zu nehmen die Anforderungen an die Anonymisierung unter der DSGVO (unter I.), der Vergleich mit den deutschen Anforderungen an die Anonymisierung von Gerichtsentscheidungen (unter II.), der Einsatz des Erlanger Tools zur Entscheidungsanonymisierung (unter III.), die Anwendbarkeit des Datenschutzrechts auf ein mit Gerichtsentscheidungen trainiertes, veröffentlichtes Sprachmodell (unter IV.), die datenschutzrechtlichen Anforderungen im Fall der Anwendbarkeit (unter V.) sowie etwaige Besonderheiten bei der Bearbeitung der Aktenauszüge durch das Erlanger Tool (unter VI.).

I. Vorgang der Anonymisierung

Für die Bestimmung der Anforderungen an eine Anonymisierung auf der Grundlage und am Maßstab des Datenschutzrechts ist der Blick zunächst auf den Begriff der personenbezogenen Daten zu richten (unter 1.). Sodann werden aus diesem Begriff die Anforderungen an Anonymisierungstechniken im Allgemeinen (unter 2.) und an Anonymisierungstechniken mit Blick auf Gerichtsentscheidungen im Besonderen (unter 3.) abgeleitet.

1. Personenbezug und Identifizierbarkeit

Die DSGVO findet nach Art. 2 Abs. 1 DSGVO Anwendung auf die (teil-)automatisierte Verarbeitung personenbezogener Daten und setzt für ihre Anwendbarkeit nach Art. 4 Nr. 1, 2 DSGVO den Umgang mit und die Verarbeitung von Informationen voraus, „die sich auf eine identifizierte oder identifizierbare natürliche Person [...] beziehen“. Auf Informationen, die sich nicht auf eine solche zumindest identifizierbare natürliche Person beziehen (i.e. von vornherein sog. anonyme oder gegebenenfalls nachträgli-

che anonymisierte Daten, ErwGr. 26 S. 5, 6 DSGVO)¹⁴ findet die DSGVO keine Anwendung.

a. „Beziehen“

Über das Merkmal des Beziehens¹⁵ werden (geringe) Anforderungen an den Informationsgehalt mit Blick auf die betroffene Person gestellt.¹⁶ Die Rechtsprechung verlangt, dass die Information „auf Grund ihres Inhalts, ihres Zwecks oder ihrer Auswirkungen mit einer bestimmten Person verknüpft ist“.¹⁷ Die Variante der Verknüpfung über den Inhalt adressiert Informationen unmittelbar über die Person, die Variante des Zwecks bezieht sich auf die Möglichkeit, Informationen zu verwenden, um eine Person zu bewerten, die Person entsprechend der Informationen zu behandeln oder zu beeinflussen, und die Variante der Auswirkungen erfasst subsidiär Informationen, die sich in anderer Weise auf die betroffene Person auswirken (können).¹⁸

Diese Anforderungen sind regelmäßig und auch im vorliegend zu begutachtenden Sachverhalt erfüllt,¹⁹ weshalb das Merkmal des Bezugs im Rahmen der weiteren Untersuchung grundsätzlich nicht weiter problematisiert und lediglich mit Blick auf die mathematische Repräsentation von personenbezogenen Daten in Sprachmodellen nachfolgend unter D.IV näher behandelt wird. Vor diesem Hintergrund ist allerdings die häufig verwendete Formulierung eines absoluten oder relativen Personenbezugs ungenau. Deshalb wird in dieser Untersuchung auf ein – absolutes oder relatives – Konzept der Identifizierbarkeit abgestellt.

14 Zu beiden Begrifflichkeiten *Mantz/Spittka*, in: *Sassenberg/Faber*, § 6 Datenschutz und IT-Sicherheit, Rn. 21.

15 Unter der DSRL über das Merkmal „über eine Person“, s. *Artikel-29-Datenschutzgruppe*, WP 136, S. 10; keine materiellen Änderungen *Winter/Battis/Halvani*, ZD 2019, 489 (489 f.).

16 *EuG*, ZD 2023, 399 (Rn. 69 ff.); *Krügel*, ZD 2017, 455 (459); *Paal*, ZfDR 2024, 129 (133).

17 *EuGH*, ZD 2018, 113 (Rn. 35) – *Nowak*; dies wiederum weit auslegend *Generalanwalt Spielmann*, Schlussanträge C-413/23 P, Rn. 33 f.; *Klabunde/Horváth*, in: *Ehmann/Selmayr*, Art. 4 Rn. 10 f.

18 *Artikel-29-Datenschutzgruppe*, WP 136, S. 11 ff.

19 Die Bedenken des fehlenden Informationsgehalts von in einem Sprachmodell „gespeicherten“ Informationen nach *HmbBfDI*, Diskussionspapier: Large Language Models und personenbezogene Daten, S. 6 f. könnten allerdings dahingehend verstanden werden, dass den in einem Sprachmodell gespeicherten Informationen ein Informationsgehalt fehlt und daher das Vorliegen personenbezogener Daten abzulehnen wäre.

b. Identifizierbarkeit

Eng zusammenhängend mit dem Bezug der Information zu einer natürlichen Person muss diese Person auch identifiziert oder identifizierbar sein.²⁰ Soweit die Person nicht unmittelbar, z.B. über ihren Namen oder andere Kennungen, i.S.d. Art. 4 Nr. 1 DSGVO identifiziert werden kann, kommt es darauf an, ob die Person über weitere Mittel (z.B. Inferenzen und die Zuhilfenahme weiterer Quellen) identifizierbar ist. In diesem Zusammenhang sind insbesondere die Perspektive des jeweiligen Akteurs (relativer Ansatz)²¹ und der aktuelle Stand der Technik zu berücksichtigen.

aa. Relatives Konzept der Identifizierbarkeit

Für die Frage der Identifizierbarkeit sind nach ErwGr. 26 S. 3 DSGVO „alle Mittel [zu berücksichtigen], die von dem Verantwortlichen oder einer anderen Person nach *allgemeinem Ermessen wahrscheinlich* genutzt werden, um die natürliche Person direkt oder indirekt zu identifizieren, wie beispielsweise das Aussondern“ (Hervorhebung d. d. Verf.). In diesem Sinne legt der EuGH einen relativen Ansatz der Identifizierbarkeit zugrunde²² und prüft mit Blick auf einen Verantwortlichen, ob dieser mit dem ihm nach vernünftiger Betrachtung²³ zur Verfügung stehenden Mitteln die natürliche

20 Zur Diff. zwischen den Merkmalen „direkt“ und „indirekt“ *Artikel-29-Datenschutzgruppe*, WP 136, S. 15.

21 Im Überblick zum objektiven bzw. absoluten und dem relativen Ansatz *Keppeler*, CR 2016, 360 (361) m.w.N.; s. auch *Bergt*, ZD 2015, 365.

22 Explizit(er) *Generalanwalt Spielmann*, Schlussanträge C-413/23 P, Rn. 70-74; s. auch *EDPB*, Verbindlicher Beschl. 1/2021, Rn. 147 f.; den EuGH ebenfalls wie hier verstehend *Hüger*, ZfDR 2024, 263 (278 f.); *Paal*, ZfDR 2024, 129 (134 ff.); *Seidel*, DSB 2023, 212; herrschend nach *Karg*, in: *Simitis/Hornung/Spiecker* gen. *Döhmman*, Art. 4 Nr. 1 DSGVO Rn. 60 f.; noch offen nach *Assion*, NJW 2023, 2619-2624 (Rn. 8 ff.); offen, aber mit Tendenz zum relativen Personenbezug nach *Klar/Kühling*, in: *Kühling/Buchner*, Art. 4 Nr. 1 DSGVO Rn. 26; diff. *Franzen*, in: *Franzen/Gallner/Oetker*, Art. 4 DSGVO Rn. 5; relativ mit Nähe zum absoluten Personenbezug nach *Keppeler/Poncza/Wölke*, CR 2024, 18 (Rn. 7); ähnl. *Moos/Rothkegel*, MMR 2016, 845 (845); *Ziegenhorn*, NVwZ 2017, 213 (217); *Eckhardt*, CR 2016, 786; *Kühling/Klar*, ZD 2017, 27 (28 ff.); dazu steht auch nicht *EDPB*, Guidelines 01/2025 on Pseudonymisation, Rn. 22 in Widerspruch, der betont, dass auch im Fall der Löschung von Informationen zur Auflösung einer Pseudonymisierung zu prüfen ist, ob gegebenenfalls andere Mittel zur Identifizierung zur Verfügung stehen.

23 *EuGH*, ZD 2024, 173 (Rn. 49) – FIN; entsprechend dem Wortlaut der DS-RL *Keppeler/Poncza/Wölke*, CR 2024, 18 (Rn. 13).

Person identifizieren kann.²⁴ Hierbei kommt es auf einen Willen zur Identifizierung jedenfalls nicht an für solche Quellen, die sich wie Kennungen innerhalb der Sphäre des Verantwortliche befinden und seinem unmittelbaren Zugriff unterliegen.²⁵ Aus der Möglichkeit zur Identifizierung (und der Anwendbarkeit des Datenschutzrechts) erwächst allerdings nicht die Pflicht, diese Identifizierung tatsächlich durchzuführen, sondern vielmehr bloß eine Pflicht, die Einhaltung der DSGVO zu gewährleisten (Art. 11 Abs. 1 DSGVO).²⁶

Aus der jüngeren Rechtsprechung des EuGH im Zusammenhang mit Art. 9 DSGVO lässt sich ableiten, dass eine Identifizierbarkeit nicht die eindeutige Zuordnung zu einer Person erfordert. Im Fall der Bestellung eines Medikaments auf der Website einer Online-Apotheke kommt es nach dem EuGH nicht darauf an, „ob diese Informationen den Nutzer oder eine andere Person betreffen, für die diese Bestellung getätigt wird“.²⁷ Nach dieser Lesart dürfte regelmäßig auch eine wahrscheinliche Zuordnung zu einer Person aus Sicht des Verantwortlichen genügen.²⁸ Diese Entscheidung des EuGH ist dogmatisch allerdings problematisch, denn eine Wahrscheinlichkeit ist richtigerweise im Rahmen der identifizierenden Mittel zu verorten.

Aus der Sicht eines Verantwortlichen stellt sich nach Maßgabe dieser EuGH-Judikatur die Weitergabe durch ihn pseudonymisierter Daten noch als Übermittlung personenbezogener Daten an einen Empfänger dar,²⁹ während der Empfänger gegebenenfalls³⁰ mangels Zugriffs auf die Zuordnungsschlüssel keine personenbezogenen Daten, sondern anonyme Daten erhebt. Diese kleinschrittige, feingranulare Betrachtung der in Rede stehenden Verarbeitungsschritte lag schon der EuGH-Entscheidung in der Rs. Fashion ID zugrunde.³¹

24 *EuGH*, NJW 2016, 3579 (Rn. 48 f.) – Breyer; ZD 2024, 173 (Rn. 46) – FIN; s. auch *EuG*, ZD 2023, 399 (Rn. 100); zust. *Baumgartner*, ZD 2023, 402 (402); dies nach der Rspr. für weniger eindeutig haltend *Halim/Marosi*, ZD 2024, 333 (334)

25 *Arning/Forgó/Krügel*, DuD 2006, 700 (703); *Brink/Eckhardt*, ZD 2015, 205 (211).

26 *Hornung/B. Wagner*, CR 2019, 565 (Rn. 34).

27 *EuGH*, NJW 2025, 33 (Rn. 88).

28 Zuvor noch die Singularisierung und eindeutige Zuordnung erfordernd *Karg*, in: *Simitis/Hornung/Spiecker* gen. *Döhmann*, Art. 4 Nr. 1 DSGVO Rn. 50; ähnl. auch *Moos*, CR 2024, 442 (Rn. 18); zur Singularisierung *Artikel-29-Datenschutzgruppe*, WP 136, S. 16; *EDPB*, Verbindlicher Beschl. 1/2021, Rn. 146 f.

29 *Generalanwalt Spielmann*, Schlussanträge C-413/23 P, Rn. 74-77; anders zuvor *EuG*, ZD 2023, 399 (Rn. 105).

30 S. insoweit zur Wahrscheinlichkeit nach allgemeinem Ermessen *EDPB*, Guidelines 01/2025 on Pseudonymisation, Rn. 22.

31 *EuGH*, ZD 2019, 455 (Rn. 76 ff.) – Fashion ID.

Das relative Konzept der Identifizierbarkeit lässt Raum für die in Art. 4 Nr. 5 DSGVO definierte und in Art. 6 Abs. 4 lit. e, Art. 25 Abs. 1, Art. 32 Abs. 1 lit. a, Art. 40 Abs. 2 lit. d, Art. 89 Abs. 1 S. 3 DSGVO referenzierte Pseudonymisierung von personenbezogenen Daten (siehe auch ErwGr. 26-29 DSGVO). Im Rahmen der Pseudonymisierung werden die Informationen und Kennungen nebst anderer identifizierender Mittel getrennt.³² Die DSGVO findet zwar grundsätzlich Anwendung (vgl. ErwGr. 26 S. 2 DSGVO),³³ honoriert aber diesen erhöhten Schutz vor der Identifizierung als technische und organisatorische Maßnahme. Für Personen, die nur Zugriff auf die Informationen, nicht aber auf die identifizierenden Mittel haben, stellen sich die Informationen unter Zugrundelegung des relativen Konzepts der Identifizierbarkeit als anonym dar.³⁴ Diese Personen unterfallen für den Umgang mit den Informationen daher grundsätzlich nicht den Vorgaben der DSGVO, können aber im Einzelfall aufgrund ihres Einflusses auf die Verarbeitung und der Zusammenarbeit mit einem Verantwortlichen mit Datenzugriff als gemeinsam Verantwortliche nach Art. 4 Nr. 7, Art. 26 DSGVO anzusehen sein.³⁵

bb. Anforderungen an die heranzuziehenden Mittel

Im Rahmen der Identifizierung sind nur solche Mittel zu berücksichtigen, die nach „allgemeinem Ermessen wahrscheinlich genutzt werden“ (ErwGr. 26 S. 3 DSGVO).³⁶ Hierfür sind alle „objektiven Faktoren, wie die Kosten der Identifizierung und der dafür erforderliche Zeitaufwand [heranzuziehen], wobei die zum Zeitpunkt der Verarbeitung verfügbare Technologie und technologische Entwicklungen zu berücksichtigen sind“ (ErwGr. 26 S. 4 DSGVO). Aus der beispielhaften Nennung („wie“) folgt,

32 Hierzu etwa im Wissenschaftskontext *Volodina et al.*, Grandma Karl is 27 years old.

33 *EuGH*, ZD 2024, 209 (Rn. 58); *Generalanwalt Spielmann*, Schlussanträge C-413/23 P, Rn. 51.

34 Vgl. *EuG*, ZD 2023, 399 (Rn. 94).

35 *EuGH*, ZD 2024, 328 (Rn. 58 f.) – TC-String. S. auch nachfolgend unter DV.3.a.

36 Die geringfügigen Änderungen im Wortlaut ggü. der DSRL sind ohne materielle Auswirkungen, s. *Hubert*, CR 2025, 77 (Rn. 14 f.); *Keppeler/Ponczka/Wölke*, CR 2024, 18 (Rn. 13); die rein hypothetische, nicht wahrscheinliche Identifizierung genügt nicht, s. *Artikel-29-Datenschutzgruppe*, WP 136, S. 17.

dass darüber hinaus auch weitere Faktoren Berücksichtigung finden können.³⁷

Der EuGH leitet aus ErwGr. 26 DSGVO ab, dass der Einsatz der Mittel nicht mit einem unverhältnismäßigen Aufwand einhergehen darf und rechtlich zulässig sein muss.³⁸ Ein etwaig bestehendes Restrisiko einer möglichen, aber unwahrscheinlichen Anonymisierung wird insoweit durch den EuGH hingenommen.³⁹ Die bloß fehlende Verwendungs- bzw. Zusammenführungsabsicht nach einer Pseudonymisierung eines Verantwortlichen soll nach dem EuGH für die Annahme einer Identifizierbarkeit nicht genügen,⁴⁰ da andernfalls die Pseudonymisierung der Anonymisierung faktisch gleichstehen würde. Ein solcher Befund stünde im Widerspruch zu den anfangs aufgezählten Inbezugnahmen auf die Pseudonymisierung innerhalb der DSGVO, die gerade die gesetzgeberische Intention einer eigenständigen Behandlung von pseudonymisierten Daten belegen.

Das Merkmal der rechtlichen Zulässigkeit wird in der Literatur zum Teil in Zweifel gezogen,⁴¹ wobei dessen Ablehnung insoweit mit geringfügigen Änderungen in ErwGr. 26 zwischen DSRL und DSGVO begründet wird (zuvor „vernünftigerweise eingesetzt“, nun „nach allgemeinem Ermessen wahrscheinlich“).⁴² Diese Zweifel dürften allerdings im Ergebnis nicht durchgreifen, denn bei tatsächlicher, rechtswidriger Identifizierung im Einzelfall liegt weiterhin eine Verarbeitung vor; eine solche Verarbeitung ist also bloß abstrakt – d.h. z.B. mit Blick auf ein gesamtes Sprachmodell – zu verneinen.⁴³ Betroffene Personen bleiben daher auch bei Ablehnung eines Personenbezugs eines Sprachmodells geschützt: Dieser Schutz wird zum einen vermittelt durch die weitgehende Verhinderung des Personenbezugs im Rahmen der identifizierenden Mittel, zum anderen durch die Anwendbarkeit der DSGVO, sofern der Personenbezug im Einzelfall nicht verhindert werden kann. Allerdings dürfte der Vorbehalt der rechtlichen Zulässigkeit im Einklang mit der vor allem auf Art. 8 GRCh ausgerichteten Rechtsprechung des EuGH bereits abstrakt nicht eingreifen, wenn von dem

37 Hubert, CR 2025, 77 (Rn. 6); EDPB, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, Rn. 41.

38 Vgl. *EuGH*, BeckRS 2024, 3655 (Rn. 51) – OLAF; s. auch schon *LG Berlin*, ZD 2013, 618.

39 Hubert, CR 2025, 77 (Rn. 10); vgl. auch *Roßnagel*, DuD 2024, 513 (514).

40 *Brink/Eckhardt*, ZD 2015, 205 (211).

41 *Hacker*, Law, Innovation and Technology 13 (2021), 257 (266 ff.); *Pesch/Böhme*, MMR 2023, 917 (920); zust. *Engeler/Rolfes*, ZD 2024, 423 (427).

42 *Krügel*, ZD 2017, 455 (459).

43 Ähnl. vor der EuGH-Entscheidung *Brink/Eckhardt*, ZD 2015, 205 (211).

rechtlichen Verbot keine effektive Abschreckungswirkung ausgeht.⁴⁴ Aus dem rechtlichen Verbot kann also richtigerweise nur eine widerlegbare Vermutung der Unwahrscheinlichkeit der Identifizierung folgen.

Für die Prüfung der Verhältnismäßigkeit sind aus ex-ante-Perspektive insbesondere das (objektive) Interesse an der Identifizierung mit dem Aufwand abzuwägen.⁴⁵ Denn selbst wenn und soweit die Zuordnung zu einer unzutreffend identifizierten Person der DSGVO unterfällt, kann eine solche Zuordnung durch einen Verantwortlichen nach allgemeinem Ermessen unwahrscheinlich und daher nicht zu berücksichtigen sein (z.B. über die Heranziehung unglaubwürdiger bzw. unglaubhafter Quellen).⁴⁶ Umgekehrt kann die Identifizierung besonders dann wahrscheinlich sein, wenn sich eine Information an ein Publikum richtet, das nach allgemeinem Ermessen weitere Quellen heranziehen kann und typischerweise auch heranziehen wird (z.B. Journalisten).⁴⁷

Es genügt für die Identifizierbarkeit aus Sicht der Verantwortlichen, wenn der Verantwortliche im Rahmen der heranzuziehenden Mittel auf weitere Personen⁴⁸ oder Quellen (z.B. eine – gegebenenfalls KI-gestützte – Internetsuche)⁴⁹ zurückgreifen kann.⁵⁰ Hiervon erfasst sind etwa Konstellationen, in denen die andere Person um identifizierende Mittel ersucht werden kann und (gegebenenfalls vertraglich) zur Beantwortung der ersuchenden Anfrage verpflichtet ist.⁵¹ Insoweit kann es nach dem EuGH sogar ausreichen, wenn rechtliche Mittel zwar zur Verfügung stehen, von diesen

44 Kühling/Klar, ZD 2017, 27 (28 ff.); Purtova, Law, Innovation and Technology 10 (2018), 40 (64 f.); ähnl. auch Karg, in: Simitis/Hornung/Spiecker gen. Döhmann, Art. 4 Nr. 1 DSGVO Rn. 31; Schwartmann et al., Praxisleitfaden zum Anonymisieren personenbezogener Daten, S. 18.

45 Ähnl. Hubert, CR 2025, 77 (Rn. 20); Artikel-29-Datenschutzgruppe, WP 136, S. 18; LG Berlin, ZD 2013, 618 (622 f.); Kroschwald, ZD 2014, 75 (76); Schild, in: BeckOK Datenschutzrecht, Art. 4 DSGVO Rn. 18; Klar/Kühling, in: Kühling/Buchner, Art. 4 Nr. 1 DSGVO Rn. 22; insoweit kann gegebenenfalls auch die Sensibilität von Datenkategorien (und damit auch der potenzielle Schaden für betroffene Personen) berücksichtigt werden, vgl. Roßnagel, DuD 2024, 513 (513).

46 Hubert, CR 2025, 77 (Rn. 25 f.).

47 EuGH, BeckRS 2024, 3655 (Rn. 63) – OLAF.

48 EuGH, NJW 2016, 3579 (Rn. 43) – Breyer; ZD 2024, 328 (Rn. 40, 46 ff.) – TC-String; nach Hubert, CR 2025, 77 (Rn. 18) sollen auch die individuellen Fähigkeiten des Verantwortlichen mit Blick auf die Mittel berücksichtigt werden.

49 Hubert, CR 2025, 77 (Rn. 20 f.).

50 Vgl. EuGH, BeckRS 2024, 3655 (Rn. 55 ff.) – OLAF; s. auch DSK, Orientierungshilfe Künstliche Intelligenz und Datenschutz, Rn. 4.

51 EuGH, ZD 2024, 328 (Rn. 48) – TC-String.

rechtlichen Mitteln aber nur im Ausnahmefall (z.B. für die Auflösung einer IP-Adresse im Zusammenhang mit einer Straftat) Gebrauch gemacht werden kann.⁵² Allerdings unterliegt auch der Zugriff auf Dritte zur Identifizierung wie auch die übrigen Mittel dem Vorbehalt der Verhältnismäßigkeit⁵³ und dürfte entsprechend ErwGr. 26 S.3, 4 DSGVO eine hinreichende Wahrscheinlichkeit voraussetzen.⁵⁴ Die einem Verantwortlichen zugewiesene (öffentlich-rechtliche) Aufgabe ist insoweit zu berücksichtigen (z.B. dürfte eine Identifizierung bei justizinterner Verwendung zur Verbesserung der Arbeitsabläufe unwahrscheinlich sein).⁵⁵

cc. Mittelbarer Personenbezug oder Identifizierbarkeit

In einer Entscheidung betreffend die Einordnung einer Fahrzeugidentifikationsnummer (FIN) gelangt der EuGH zu dem Ergebnis, dass die FIN für den Fahrzeughersteller „mittelbar“ ein personenbezogenes Datum darstellen kann, auch wenn der Fahrzeughersteller weder über die notwendigen Mittel zur Identifizierung verfügt noch auf die betreffenden Mittel über andere Wirtschaftsakteure zugreifen kann.⁵⁶ Diese Aussage könnte als eine Abkehr von einem relativen Konzept der Identifizierbarkeit verstanden werden, gegebenenfalls indem die spätere Identifizierbarkeit durch einen anderen Verantwortlichen auf einen diese Daten übermittelnden Akteur zurückwirkt, sodass dieser Akteur rückwirkend als Verantwortlicher anzusehen ist.⁵⁷

Die englische Sprachfassung des Urteils („indirectly [...] personal data“) legt nahe, dass der EuGH sich hier auf Art. 4 Nr. 1 DSGVO beziehen könnte, wonach die direkte oder indirekte Identifizierbarkeit genügt.⁵⁸ Allerdings ist die in Art. 4 Nr. 1 DSGVO angelegte Identifizierbarkeit nicht pauschal

52 *EuGH*, NJW 2016, 3579 (Rn. 47) – Breyer; *Hubert*, CR 2025, 77 (Rn. 13); wohl eingeschränkt in der jüngeren Rspr., hierzu *Keppeler/Poncza/Wölke*, CR 2024, 18 (Rn. 15); zu dem Hintergrund der Breyer-Entscheidung auch *Keppeler*, CR 2016, 360 (362 f.); *H. Richter*, EuZW 2016, 912 (913).

53 *Hubert*, CR 2025, 77 (Rn. 12).

54 Krit. insoweit zur EuGH-Entscheidung *Kühling/Klar*, ZD 2017, 27 (28 ff.).

55 *Roßnagel*, DuD 2024, 513 (514).

56 *EuGH*, ZD 2024, 173 (Rn. 49) – FIN.

57 *Hanloser*, ZD 2024, 175 (176); krit. auch mit Blick auf die Unklarheit dieses Konzepts *Keppeler/Poncza/Wölke*, CR 2024, 18 (Rn. 21).

58 S. auch *EuGH*, ZD 2024, 173 (Rn. 45) – FIN; ähnl. *Keppeler/Poncza/Wölke*, CR 2024, 18 (Rn. 19).

gleichzusetzen mit dem vom EuGH verwendeten Begriff der mittelbar personenbezogenen Daten.

Der Gerichtshof nimmt in seinen einschlägigen Ausführungen Bezug auf zwei Ausführungen des EuGH-Generalanwalts Sánchez-Bordona (in der Rs. C-319/22), die bei dem Verständnis dieses Merkmals mit Blick auf den Fahrzeughersteller in diesem Zusammenhang weiterhelfen können: Zum einen wird die Aussage in Bezug genommen, dass die FIN nicht stets überhaupt einer Person zugeordnet werden kann (z.B. wenn das Kraftfahrzeug „keiner natürlichen Person gehört“).⁵⁹ Zum anderen stellt der Generalanwalt auf die für die Wirtschaftsakteure verfügbaren Mittel ab und hält fest, dass in dem Fall verfügbarer Mittel zur Identifizierung „die FIN für sie (und mittelbar für den Hersteller, der ihnen die FIN zur Verfügung stellt) personenbezogene Daten [wären], deren Verarbeitung der DSGVO unterliegt“.⁶⁰ Die letztere Aussage erschließt sich nicht ohne Weiteres, dürfte aber mit Blick auf die erste Aussage und den konkret zu beurteilenden Sachverhalt so verstanden werden, dass der Hersteller gegebenenfalls unter Zuhilfenahme der Wirtschaftsakteure selbst natürliche Personen anhand der FIN identifizieren kann. Unter Berücksichtigung der Schlussanträge eines anderen Generalanwalts (i.e. Generalanwalt Spielmann in der Rs. C-413/23 P) könnte diese Aussage jedoch auch dahingehend eingeordnet werden,⁶¹ dass für den Fahrzeughersteller Informationspflichten nach Art. 13, 14 DSGVO eingreifen, wenn der Fahrzeughersteller – für ihn anonyme – Informationen an eine Person oder Stelle weitergibt (sog. Empfänger, Art. 4 Nr. 9 DSGVO), die die Informationen einer natürlichen Person zuordnen kann.⁶²

Trotz verbleibender Ungewissheit ist mithin davon auszugehen, dass der EuGH grundsätzlich an einem relativen Konzept der Identifizierbarkeit festhält.

59 Generalanwalt Sánchez-Bordona, Schlussanträge C-319/22, Rn. 34.

60 Generalanwalt Sánchez-Bordona, Schlussanträge C-319/22, Rn. 41.

61 Generalanwalt Spielmann, Schlussanträge C-413/23 P, Rn. 65-75, der die umgekehrte Konstellation der Übermittlung an eine Person, die nicht über die identifizierenden Mittel verfügt, in den Blick nimmt.

62 Hierzu schon zuvor Gola, in: Gola/Heckmann, Art. 4 DSGVO Rn. 25; Mantz/Spittka, in: Sassenberg/Faber, § 6 Datenschutz und IT-Sicherheit, Rn. 21.

dd. Relatives Konzept der Identifizierbarkeit in zeitlicher Hinsicht

In zeitlicher Hinsicht ist die Anonymisierung, abhängig von dem konkret gewählten Verfahren, ebenfalls relativ und nicht absolut zu bewerten.⁶³ Durch den fortschreitenden Stand der Technik kann eine zuvor anonymisierte Information zu einem personenbezogenen Datum werden,⁶⁴ falls etwa neue Verfahren zur (Wieder-)Herstellung des Personenbezugs zur Verfügung stehen oder sich die Kosten zum Einsatz bestehender Verfahren signifikant reduzieren.⁶⁵ Ferner können (potenzielle) Verantwortliche im Laufe der Zeit den Zugriff erhalten auf ergänzende Datenquellen oder auf Dritte mit Zusatzwissen.⁶⁶ Eine solche ergänzende Datenquelle kann gegebenenfalls auch durch die betroffene Person selbst eröffnet werden, indem Betroffenenrechte ausgeübt und durch Zusatzinformationen eine Zuordnung der bei dem Verantwortlichen bereits gespeicherten Informationen zu der betroffenen Person ermöglicht werden.⁶⁷

Die vorliegende Untersuchung legt ein weites Begriffsverständnis der Anonymisierung zugrunde; dies gilt unabhängig davon, ob einmal anonymisierte Daten zu einem späteren Zeitpunkt erneut einer natürlichen Person zugeordnet werden können.⁶⁸

c. Zwischenergebnis

Für die Frage, ob personenbezogene Daten nach der DSGVO vorliegen, kommt es sowohl darauf an, ob Informationen Aussagen zu einer natürlichen Person zu entnehmen sind, als auch darauf, ob die Informationen aus Sicht eines Verantwortlichen mit den ihm zur Verfügung stehenden Mitteln tatsächlich auch dieser natürlichen Person zuzuordnen sind. Soweit und solange (*sic!*) diese Identifizierbarkeit ausgeschlossen ist, handelt es sich um anonyme Daten. Diese Annahme wird im Folgenden zugrunde gelegt, womit zugleich auch ein Zielzustand definiert ist.

63 Hornung/B. Wagner, CR 2019, 565 (Rn. 3); Roßnagel, DuD 2024, 513 (514).

64 Purtova, Law, Innovation and Technology 10 (2018), 40; vgl. auch Artikel-29-Datenschutzgruppe, WP 216, S. 7; Krügel, ZD 2017, 455 (456).

65 Hornung/B. Wagner, CR 2019, 565 (Rn. 14).

66 Hornung/B. Wagner, CR 2019, 565 (Rn. 10).

67 Hornung/B. Wagner, CR 2019, 565 (Rn. 11).

68 S. zur Unumkehrbarkeit und dem Restrisiko Artikel-29-Datenschutzgruppe, WP 216, S. 6 f.

2. Überblick zu den Anonymisierungsverfahren

Zur Erreichung dieses – vor allem relativen⁶⁹ – Anonymisierungsziels kommen verschiedene Techniken in Betracht, die jeweils eine Verarbeitung personenbezogener Daten erfordern.⁷⁰ Diese Techniken werden im Folgenden überblicksartig und stark vereinfacht dargestellt, soweit ihnen für den dieser Untersuchung zugrunde liegenden Sachverhalt eine Relevanz zukommt.

a. Löschung von Merkmalen zur Identifizierung durch Auslassung oder Ersetzung

In Betracht kommen zunächst die Auslassung oder Ersetzung⁷¹ als Löschung von Kennungen und anderen identifizierenden Merkmalen⁷² (oder auch: Neutralisierung),⁷³ wie sie beispielsweise im Rahmen der Anonymisierung von Gerichtsentscheidungen im Wege der Veröffentlichung praktiziert werden (siehe dazu nachfolgend unter D.II). Hierbei werden beispielsweise der Name einer natürlichen Person, E-Mail-Adressen, Internetdomain-Bezeichnungen, Telefonnummern, Kontodaten und sonstige Identifikationsnummern entfernt. Ohne diese Kennungen und sonstigen Merkmale können grundsätzlich auch die übrigen Informationen (z.B. eine getätigte Aussage) nicht mehr einer natürlichen Person zugeordnet werden.

Der Anonymisierungserfolg dieses Verfahrens hängt somit maßgeblich davon ab, ob tatsächlich alle Merkmale identifiziert und entfernt werden, die für sich genommen oder gegebenenfalls mit anderen Merkmalen mit hinreichender Wahrscheinlichkeit⁷⁴ eine Identifizierung der natürlichen Person ermöglichen (z.B. ein nicht entfernter Straßename samt Hausnummer, der aber eindeutig einer Stadt und sodann einem Immobilieneigentümer zugeordnet werden kann). Insoweit kommt es auch darauf an, welche zusätzlichen öffentlichen oder nicht-öffentlichen Quellen zum Zwecke der Identifizierung herangezogen werden können. Insgesamt wird vielfach nicht ausgeschlossen werden können, dass in der Zukunft weitere Quellen

69 *Schwartmann et al.*, Praxisleitfaden zum Anonymisieren personenbezogener Daten, S. 3.

70 S. zur Rechtsgrundlage unter D.V.4.c.aa.

71 *Deuber/Keuchen*, MMR 2023, 338 (342); vgl. auch die Schwärzung *Determann/Paal*, KI-Recht international, S. 57.

72 *Klar/Kühling*, in: *Kühling/Buchner*, Art. 4 Nr. 1 DSGVO Rn. 34.

73 *BVerwG*, NJW 1997, 2694 (2695).

74 S. zur Inferenz *Artikel-29-Datenschutzgruppe*, WP 216, S. 11.

zur Verfügung stehen, die z.B. eine Anschrift leichter auf eine Person zurückführen lassen.

b. Weitere Anonymisierungstechniken und Maßnahmen i.w.S.

Die Artikel-29-Datenschutzgruppe differenziert im Weiteren zwischen Techniken der Randomisierung als Entfernen der Verbindung zwischen Person und Datum (z.B. Vertauschen von Datenzuordnungen, stochastische Überlagerung, Differential Privacy) und der Generalisierung als Abstrahierung einzelner Worte (z.B. Aggregation durch Entfernung von Straßennamen, k-Anonymität, l-Diversität und t-Closeness).⁷⁵ Diese Techniken eignen sich besonders für die Anwendung auf strukturierte Datensätze. Für die im GSJ-Projekt bearbeiteten Gerichtsentscheidungen und Aktenauszüge liegt die Herausforderung demgegenüber aber in der potenziellen Identifizierung der Merkmale. Die bloße Löschung oder Pseudonymisierung dieser Merkmale schadet der Verständlichkeit nicht, sodass es regelmäßig keiner besonderen Anonymisierungsverfahren über eine Identifizierung und Löschung von Kennungen und sonstigen identifizierenden Merkmalen hinaus bedarf.

Neben den eigentlichen Anonymisierungstechniken kommen zudem auch Maßnahmen in Betracht, die primär an den identifizierenden Mitteln ansetzen und hierdurch zum gleichen Ergebnis gelangen. Der Rückgriff auf Quellen kann gesetzlich eingeschränkt oder einzelne Personen können von diesen Quellen abgeschottet werden.⁷⁶ Zu denken ist ferner an eine Selbstverpflichtung des Verantwortlichen, gegebenenfalls als Verhaltensregel i.S.d. Art. 40 DSGVO, sich des Rückgriffs auf verfügbare Mittel zur Identifizierung zu enthalten.⁷⁷

75 Artikel-29-Datenschutzgruppe, WP 216, S.14 ff.; Dewitte, in: AI Meets the GDPR, S. 140 f.; Schwartmann et al., Praxisleitfaden zum Anonymisieren personenbezogener Daten, S. 6; Winter/Battis/Halvani, ZD 2019, 489 (490 f.); International Working Group on Data Protection in Technology, Working Paper on Large Language Models (LLMs), S. 51 ff.; s. auch zur Pseudonymisierung EDPB, Guidelines 01/2025 on Pseudonymisation.

76 Zu Letzterem Hornung/B. Wagner, CR 2019, 565 (Rn. 42).

77 Hornung/B. Wagner, CR 2019, 565 (Rn. 49).

3. Anforderungen an die Anonymisierung unveröffentlichter Gerichtsentscheidungen

Die ermittelten Maßstäbe sind nunmehr auf die Anonymisierung unveröffentlichter Gerichtsentscheidungen und die darin enthaltenen Kennungen sowie indirekten Identifikationsmerkmale zu übertragen.⁷⁸

a. Typische identifizierende Merkmale

Hierzu bedarf es zunächst der Betrachtung einer typischen Gerichtsentscheidung.

Eine Gerichtsentscheidung der Zivilgerichtsbarkeit enthält mit Blick auf Kläger, Beklagte, Prozessbevollmächtigte, sonst im Rubrum genannte Personen (z.B. Nebenintervenienten), Richter und sonstige Beteiligte (z.B. Zeugen) verschiedene Informationen, die sich auf diese Personen beziehen können.

Maßgeblich für die Identifizierbarkeit dieser natürlichen Personen werden insbesondere die folgenden Merkmale sein:

Rubrum	<ul style="list-style-type: none"> • Vor- und Nachname einer (natürlichen) Person, gegebenenfalls als Vertreter einer juristischen Person • Anrede der Person • Private Wohnanschrift oder gegebenenfalls Anschrift des Arbeitgebers bzw. der Partei • Aktenzeichen (im Zusammenhang mit gegebenenfalls öffentlich einsehbaren Sitzungsplänen) • Nachname des Richters oder der Richterin sowie des Urkundsbeamten oder der Urkundsbeamtin • Ort des Gerichts
Tatbestand und Entscheidungsgründe	<ul style="list-style-type: none"> • Nach- und gegebenenfalls Vorname einer Person • Ortsangaben • Datumsangaben (im Zusammenhang mit weiteren Informationen, wie etwa dem Ort) • Angaben zu Kommunikationsmitteln (z.B. Telefonnummer oder E-Mail-Adresse) • Angaben zu Zahlungsmitteln (z.B. Kreditkartennummern, Bankverbindungen) • Eigentums- oder vergleichbare Positionen (z.B. Marken, Domains) • Angaben in der Sache (z.B. getätigte Aussagen, Vereins- oder Gesellschaftszugehörigkeit, Geschehen des öffentli-

⁷⁸ Zum Begriff *Schwartzmann et al.*, Praxisleitfaden zum Anonymisieren personenbezogener Daten, S. 10.

	<p>chen Lebens im Zusammenhang mit öffentlich zugänglicher Berichterstattung)</p> <ul style="list-style-type: none">• Identifikationsnummern (z.B. Fahrzeugidentifikationsnummer (FIN), persönliche Steueridentifikationsnummer, Sozialversicherungsnummer)
Unterschriften	<ul style="list-style-type: none">• Nach- und gegebenenfalls Vorname einer Person

Für Entscheidungen anderer Gerichtsbarkeiten gelten die vorstehenden Ausführungen und Überlegungen entsprechend.

b. Maßstab für Anonymisierungsverfahren

Die zentrale Herausforderung der Anonymisierung von Gerichtsentscheidungen liegt in der (fehlenden) Struktur⁷⁹ der Dokumente, wodurch eine Identifizierung der zu anonymisierenden Informationen erheblich erschwert wird. Während Datenbanken nach einzelnen Attributen aufgliedert sind, enthalten die Gerichtsentscheidungen jeweils einen Text. Die in dem Text enthaltenen Wörter können allein oder in Verbindung mit anderen Wörtern ein solches Attribut darstellen. Besondere Aufmerksamkeit ist daher zu richten auf die Ermittlung eines identifizierenden Personenbezugs unter Berücksichtigung der voraussichtlich verfügbaren (öffentlichen) Quellen im Zusammenhang mit der Anonymisierung von Gerichtsentscheidungen.

aa. Berücksichtigung von weiteren Mitteln zur Identifizierung

Das Gelingen der Anonymisierung hängt maßgeblich davon ab, welche Mittel die Verantwortlichen in rechtmäßiger Weise und mit verhältnismäßigem Aufwand heranziehen könnten, um eine natürliche Person zu identifizieren.

Solche Mittel zur Identifizierung umfassen insbesondere Listen der Orts- und Straßenbezeichnungen,⁸⁰ Kalender, identifizierende und nichtidentifizierende mediale Berichterstattung, Geschäftsverteilungspläne sowie weitere Informationsquellen. Dies gilt unabhängig davon, ob diese Informationsquellen über das Internet, über den Handel oder auf individuelle Nachfrage verfügbar sind.

79 Vgl. den Begriff der strukturierten Daten als Schlüssel-Wert-Paar *Schwartmann et al.*, Praxisleitfaden zum Anonymisieren personenbezogener Daten, S. 3.

80 *Deuber/Keuchen*, MMR 2023, 338 (340).

Für einzelne Personen können zudem amtliche Auskünfte (z.B. des Einwohnermeldeamts) und Registerauszüge (z.B. aus dem Handelsregister) heranzuziehen sein. Darüber hinaus kommt in Betracht, dass einzelne Personen, die über ein Insiderwissen mit Blick auf den Prozess verfügen, durch dieses Wissen für weitere Informationen aus der Gerichtsentscheidung einen Personenbezug herstellen können (z.B. Personen, denen Parteien von einem Gerichtsprozess berichtet haben). Vor diesem Hintergrund ist eine (absolute)⁸¹ Anonymisierung allen Personen gegenüber zwar nicht möglich, datenschutzrechtlich aber gerade auch nicht zwingend erforderlich.⁸² Nachfolgend wird daher die Anonymisierung mit Ausnahme der Parteien mit Insiderwissen zu der jeweiligen Gerichtsentscheidung im Vordergrund stehen.

bb. Identifizierung der zu anonymisierenden Informationen

Für die Identifizierung der personenbezogenen Daten sind über die typischen identifizierenden Merkmale in Form von Kennungen hinaus sämtliche weiteren Wörter aus dem Text der Gerichtsentscheidung (unter (1)) bzw. typisierend die Entscheidung selbst (unter (2)) darauf zu untersuchen, ob sie mit hinreichend hoher Wahrscheinlichkeit die Identifizierung unter Berücksichtigung der öffentlich verfügbaren Informationsquellen ermöglichen.

(1) Untersuchung auf Wortebene bzw. Zeichenebene

Folgt man einem Ansatz der vollständigen Risikoeliminierung, so sind in einem ersten Schritt sämtliche Informationen als identifizierend zu betrachten und sodann in einem zweiten Schritt typischerweise nicht-identifizierende Wörter auszusondern (z.B. unbestimmte oder bestimmte Artikel als sog. Stoppwörter).⁸³ Ein solcher Ansatz führt allerdings zur Aussonderung zahlreicher Informationen (z.B. durch Aussonderung verschiedener Substantive und Prädikate) in einem Umfang, dass der Text für Zwecke des Trainings eines Sprachmodells nach einer Veränderung dieser Informationen regelmäßig nicht mehr hinreichend geeignet sein dürfte.

81 Schwartmann et al., Praxisleitfaden zum Anonymisieren personenbezogener Daten, S. 17.

82 S. hierzu insb. unter D.II.5.b sowie unter DV.4.c.

83 Brockmeyer, Text und Data Mining, S. 21.

Es bietet sich daher ein Rekurs auf das der DSGVO zugrundeliegende Konzept eines risikobasierten Ansatzes an, um eine Einordnung der Zeichenfolgen, Wörter, Wortkombinationen in verschiedene Risikokategorien vorzunehmen. Ein solcher Ansatz unter angemessener Berücksichtigung des Risikos einer Identifizierbarkeit differenziert beispielsweise zwischen:

1. mit hoher Wahrscheinlichkeit identifizierenden Wörtern (z.B. Ortsbezeichnungen, Telefonnummern, E-Mail-Adressen und spezifischen Berufsbezeichnungen);
2. potenziell identifizierenden Wörtern oder Zeichenfolgen (z.B. Datumsangaben, besonderen Verhaltensweisen und Eigenschaften einer Person, Angaben in Anführungszeichen, jeweils unter Berücksichtigung länderspezifischer Besonderheiten);⁸⁴
3. ohne vernünftigen Zweifel nicht mehr identifizierenden Wörtern (z.B. grundsätzlich identifizierungsneutrale (Stopp-)Wörter wie „zwischen“, „durch“, „das“, „ein“, „oder“, „und“).

(2) Untersuchung auf Entscheidungsebene – Cluster von Gerichtsentscheidungen

Gerichtsentscheidungen mit allenfalls potenziell identifizierenden Wörtern können durch eine Clusterung weitgehend ausgeschlossen werden bzw. für eine händische Anonymisierung vorsortiert werden. So empfiehlt sich in diesem Sinne beispielsweise die Filterung von Gerichtsentscheidungen mit einem besonders umfangreichen Tatbestand sowie anhand ausgewählter Spezialzuständigkeiten (z.B. einer Kammer für Arzthaftungssachen, vgl. § 348 Abs. 1 Nr. 2 lit. e ZPO) und der Anzahl der beteiligten Personen.

In ähnlicher Weise differenziert die Dokumentationsstelle des OLG Frankfurt a. M. und ordnet Gerichtsentscheidungen in vier Anonymisierungsstufen ein:⁸⁵

1. Stufe A – geringer Anonymisierungsaufwand: Ersetzung von Eigennamen durch Kürzel oder Auslassung⁸⁶;
2. Stufe B – erhöhter Anonymisierungsaufwand: insbesondere in Familien- und Arzthaftungssachen: Ersetzung bzw. Auslassung auch von „Straßen-

⁸⁴ S. das anschauliche Beispiel in einem Arzthaftungsfall bei *Nöhre*, MDR 2019, 136 (137); länderspezifische Besonderheiten zu berücksichtigen nach *BayLDA*, Datenschutzkonforme Künstliche Intelligenz, S. 5.

⁸⁵ *Nöhre*, MDR 2019, 136 (138 ff.).

⁸⁶ Hierzu etwa *Winter/Battis/Halvani*, ZD 2019, 489 (491) m.w.N.

und Ortsnamen, Hausnummern, Kfz-Kennzeichen, Telefonnummern, Konto- oder Kundennummern, E-Mail- und Internetadressen sowie Identifikationsnummern von Behörden“ sowie gegebenenfalls auch von „Staatsangehörigkeiten, Alters- und Datumsangaben sowie Aktenzeichen anderer Verfahren“;

3. Stufe C – brisant: zusätzliche Auslassung von Textpassagen, z.B. Untersuchungsbefunden;
4. Stufe D – keine Veröffentlichung: z.B. in Staatsschutzprozessen.

Die hierdurch erfolgende Differenzierung kann sich grundsätzlich für eine Vorfilterung eignen. Die Effektivität dieser Vorgehensweise hängt maßgeblich ab von der richtigen Stufenzuordnung und der sorgfältigen Anwendung von Anonymisierungstechniken im Einzelfall.

cc. Training eines Sprachmodells als mögliche Anonymisierungstechnik

Durch das Training eines Sprachmodells (insbesondere eines Large Language Models, LLM) kann die Identifizierbarkeit einer natürlichen Person womöglich ausgeschlossen oder zumindest erschwert werden.⁸⁷

Im Rahmen des Trainings gehen durch die Tokenisierung und Neuberechnung von Informationen auf Grundlage der Trainingsdaten Informationen aus den Ausgangstexten verloren bzw. werden neu angeordnet. Einige zuvor eindeutige Verknüpfungen (z.B. „Herr Müller wohnt in der Rathausstr.“) gehen zugunsten einer wahrscheinlichen Beziehung der Buchstabenketten verloren (z.B. auf die Token „Herr“, „Müller“, „woh“, „nt“, „in“, „der“ folgt mit einer gewissen Wahrscheinlichkeit das Token „Rathaus“). Diese Wahrscheinlichkeit hängt von den übrigen Trainingsdaten ab, z.B. davon, ob das Token „Müller“ auch in anderen Zusammenhängen in den Trainingsdaten verwendet wird. Ferner beeinflusst die Eingabeaufforderung an ein KI-System, ob die im Sprachmodell repräsentierten Token in der Ausgabe in Beziehung zueinander gesetzt werden. Denn diese Eingabeaufforderung wird durch das Sprachmodell als eine Sammlung von Token behandelt, deren wahrscheinliche Fortführung durch das Modell ausgegeben wird.

87 S. auch zu Anonymisierungstechniken im Zusammenhang mit KI-Modellen (Privacy Preserving Machine Learning) im rechtswissenschaftlichen Kontext Winter/Battis/Halvani, ZD 2019, 489 (492 f.); allg. Al-Rubaie/J. Morris Chang, Privacy Preserving Machine Learning: Threats and Solutions.

Nimmt man in einem Gedankenexperiment an, dass die zuvor genannten Token sich nur in exakt dieser Reihenfolge in den Trainingsdaten wiederfinden und in Auszügen im Rahmen der Eingabeaufforderung übergeben werden, wird theoretisch⁸⁸ stets der Straßensname ausgegeben. In Ansehung des Umfangs der Trainingsdaten eines LLM ist die genaue Ausgabe dieses identifizierenden Auszugs aus den Trainingsdaten zwar unwahrscheinlich, aber möglich.

Gerade aufgrund dieser Ungewissheit als eine Form der sog. Vertauschung⁸⁹ der Merkmale „Person“ und „Information“ bzw. Verwässerung der Verknüpfung kann das Training eines LLM als eine Anonymisierungstechnik im weiteren Sinne angesehen werden.⁹⁰ Das führt zwar nicht dazu, dass das Sprachmodell generell und stets keine personenbezogenen Daten enthält, die in dem Modell gespeichert sind.⁹¹ Die Zuordnung von Informationen aus den Trainingsdaten zu der jeweiligen natürlichen Person wird aber jedenfalls deutlich unwahrscheinlicher. In Abhängigkeit von den in Rede stehenden Trainingsdaten, typischen Eingabeaufforderungen und den dem potenziell Verantwortlichen zur Verfügung stehenden Mitteln kann mit Blick auf einen potenziellen Verantwortlichen die Schwelle der vernünftigerweise bestehenden Identifizierungsmöglichkeit unterschritten werden und es können ihm gegenüber nicht-personenbezogene Daten vorliegen.

c. Nach allgemeinem Ermessen wahrscheinlich genutzte Mittel

Soweit der bearbeitete Entscheidungstext zwar keine identifizierenden Kennungen mehr enthält, aber natürliche Personen durch Hinzuziehung weiterer Quellen identifizierbar bleiben, kommt es darauf an, auf welche Mittel der potenziell Verantwortliche nach vernünftiger Betrachtung⁹² wahrscheinlich zurückgreifen wird (vgl. auch ErwGr. 26 S. 3 DSGVO).⁹³

88 Soweit hier Faktoren, wie etwa die Temperatur für eine gewisse Zufälligkeit als eine nachgeahmte Kreativität des KI-Systems, ausgeblendet werden.

89 Artikel-29-Datenschutzgruppe, WP 216, S. 16 f.

90 Ähnl. i.E. Hüger, ZfDR 2024, 263 (278).

91 So aber HmbBfDI, Diskussionspapier: Large Language Models und personenbezogene Daten.

92 EuGH, ZD 2024, 173 (Rn. 49) – FIN.

93 S. schon unter D.I.1.b.bb.

Das spätere Anwendungsszenario des Sprachmodells, der Verbleib und die Zugänglichkeit des Trainingsdatensatzes können insoweit von Bedeutung für die Anforderungen an die Anonymisierung der Gerichtsentscheidungen sein. Legt man beispielsweise ein Anwendungsszenario zugrunde, bei dem ausschließlich Personen aus der über Art. 1 Abs. 3, Art. 20 Abs. 3 GG besonders verpflichteten Richterschaft und Justizbedienstete einen Zugriff auf das Sprachmodell erhalten, erscheint es unwahrscheinlich(er), dass diese weiteren Mittel tatsächlich verwendet werden, um zufällige Bruchstücke aus den Trainingsdaten einer natürlichen Person zuzuordnen. Bei einer öffentlichen Zugänglichkeit des KI-Systems (z.B. in Form eines öffentlich verfügbaren Chatbots), gegebenenfalls auch des Sprachmodells einschließlich der Gewichte (z.B. als Open Source) sowie des Trainingskorpus werden eine Heranziehung weiterer Quellen und ein Missbrauchspotenzial demgegenüber deutlich wahrscheinlicher sein.⁹⁴

d. Schlussfolgerungen für die Anwendung von Anonymisierungstechniken auf Gerichtsentscheidungen

Für die möglichst weitreichende Entfernung eines Personenbezugs sind ausgewählte Wortkategorien stets zu entfernen, auszutauschen oder mit anderen geeigneten Anonymisierungstechniken zu behandeln (z.B. Straßen- und Ortsbezeichnungen). Gerichtsentscheidungen mit Wortkategorien, die abhängig vom Einzelfall identifizierend wirken können, sind anhand von äußeren Merkmalen (z.B. dem Tatbestandsumfang, der Anzahl mitwirkender Personen oder dem Rechtsgebiet) gegebenenfalls gesondert zu behandeln. Insoweit kann aber bereits das Training eines Sprachmodells die Wahrscheinlichkeit einer Identifizierbarkeit erheblich reduzieren und hierdurch zu einer Anonymisierung beitragen.

Sämtliche Anonymisierungstechniken sind in regelmäßigen Abständen nach dem Stand der Technik zu überprüfen und gegebenenfalls anzupassen.

Für die Anforderungen an die Anonymisierung ist zudem darauf abzustellen, welcher Personenkreis einen Zugriff erhält auf die anonymisierten Gerichtsentscheidungen und gegebenenfalls das hierdurch trainierte Sprachmodell. Bei einer Verwendung innerhalb der Justiz werden insoweit zugunsten einer vereinfachten Annahme der Anonymisierung geringere

⁹⁴ Ausf. nachfolgend unter D.IV.3.

Anforderungen an die Heranziehung zusätzlicher Quellen zu stellen sein als im Fall einer darüber hinausgehenden Veröffentlichung.

Falls eine Anonymisierung im Sinne der vorbenannten Maßgaben nicht erfolgreich ist und auch nicht aufgrund des konkreten Einsatzes des GSJ-Sprachmodells anzunehmen ist (hierzu nachfolgend unter D.IV.), findet die DSGVO Anwendung und es bedarf für die Datenverarbeitung insbesondere des Vorliegens einer Rechtsgrundlage nach Art. 6 Abs. 1 DSGVO sowie gegebenenfalls nach Art. 9, 10 DSGVO für die Verarbeitung besonderer Kategorien personenbezogener Daten.

II. Vergleich mit den Maßstäben zur Veröffentlichung von Gerichtsentscheidungen

Für die etablierte Anonymisierungspraxis der Rechtspflege betreffend die Veröffentlichung von Gerichtsentscheidungen ist zu untersuchen, ob diese Praxis den vorstehend herausgearbeiteten Vorgaben der DSGVO gerecht werden muss und wird.

1. Anwendbarkeit der Maßstäbe der DSGVO

Die DSGVO als Verordnung i.S.d. Art. 288 Abs. 2 AEUV genießt allgemeine Geltung unmittelbar in jedem Mitgliedstaat und Anwendungsvorrang vor konfligierenden nationalen Rechtsvorschriften (vgl. Art. 23 Abs. 1 GG).

In Ansehung der grundsätzlich vollharmonisierenden⁹⁵ Natur der DSGVO findet das DSGVO-Regelungsregime Anwendung auf sämtliche (teil-)automatisierten Vorgänge mit personenbezogenen Daten nach Art. 2, 4 Nr. 1, 2 DSGVO vorbehaltlich der territorialen Anforderungen nach Art. 3 DSGVO. Die DSGVO ist mithin zur Anwendung berufen betreffend die Verarbeitung von in Gerichtsentscheidungen enthaltenen personenbezogenen Daten. Dieser Befund gilt unabhängig davon, ob eine Verwaltungsvorschrift zur Anonymisierung einer Gerichtsentscheidung eingehalten und die Information veröffentlicht wurden⁹⁶ oder ob es sich um die Veröffentlichung von Dokumenten durch Behörden handelt. Denn die DSGVO

95 S. insb. ErwGr. 7 ff. DSGVO und die zahlreichen Öffnungsklauseln innerhalb der DSGVO.

96 Vgl. auch *International Working Group on Data Protection in Technology*, Working Paper on Large Language Models (LLMs), S. 43 f.

enthält gerade keine dem § 5 UrhG entsprechende Vorschrift zu amtlichen Werken, die Informationsanordnungen nach Durchlaufen eines vorgesehenen Verfahrens den Schutz unter dem Rechtsakt verwehrt.

2. Anonymisierungspraxis in Nordrhein-Westfalen bis zum Jahr 2021

Die Anonymisierungspraxis der Gerichte, hier anhand des Beispiels Nordrhein-Westfalens hat sich im Laufe der Jahrzehnte maßgeblich verändert. So stützte sich die Sicherheit der Anonymisierungsrichtlinien der NRW-Justiz aus dem Jahr 2002 – und wohl auch mit Blick auf die Richtlinie aus dem Jahr 2018 – vornehmlich auf die Geheimhaltung der Anonymisierungstechnik.⁹⁷ Sicherheit allein durch Geheimhaltung (z.B. im Hinblick auf eine Substitutionstabelle für Anfangsbuchstaben) ist allerdings kein anerkanntes Konzept nach dem Stand der Technik und wird in der Kryptanalyse abgelehnt.⁹⁸ So konnte nachgewiesen werden, dass sich aufgrund öffentlicher Datenquellen Rückschlüsse auf Straßennamen ziehen lassen, die mit monoalphabetischer Substitution (vermeintlich) anonymisiert worden waren.⁹⁹

3. Jüngere Anonymisierungspraxis in Nordrhein-Westfalen

Jedenfalls die öffentlich zugängliche Verwaltungsvorschrift der Justiz Nordrhein-Westfalen aus dem Jahr 2021 (nachfolgend: Verwaltungsvorschrift)¹⁰⁰ legt geeignetere Techniken der Anonymisierung zugrunde und wird im Folgenden untersucht. Allgemein gilt, dass die effektive Umsetzung mit zunehmendem Fortschritt von KI-basierten Anonymisierungsprojekten leichter fallen dürfte.¹⁰¹

Die Verwaltungsvorschrift geht zunächst aus von einer Abwägungsentscheidung gestützt auf nationale Rechtsgrundlagen i.S.d. Art. 6 Abs. 1 lit. c

97 *VG Düsseldorf*, BeckRS 2020, 34887 (Rn. 34); in Abgrenzung hierzu die Fassung von 2021, vgl. *OVG Düsseldorf*, BeckRS 2023, 2497 (Rn. 35).

98 Nur *Deuber/Keuchen*, MMR 2023, 338 (338 f.).

99 *Deuber/Keuchen*, MMR 2023, 338; s. auch zur möglichen Deanonymisierung *Deuber/Keuchen/Christin*, Assessing Anonymity Techniques Employed in German Court Decisions: A De-Anonymization Experiment, 2023.

100 Übermittlung von Entscheidungsabschriften an Dritte und Veröffentlichung in Datenbanken, RV d. JM vom 10. Mai 2021 (1552 - I. 12).

101 S. hierzu z.B. *Biallaß*, MMR 2024, 646 (649 f.) unter Verweis auf JANO und ALKS.

DSGVO (so konkret etwa § 299 Abs. 2 ZPO, §§ 475, 480 Abs. 1 StPO (Ziff. 1.2.2. Verwaltungsvorschrift)). Zudem legt die Verwaltungsvorschrift in Ziff. 2.3 zugrunde, dass ein „Rückschluss auf die dahinterstehende Person auszuschließen“ und gegebenenfalls die Möglichkeit der Identifizierung durch weitere Angaben im Entscheidungstext zu berücksichtigen ist (z.B. Lebensumstände der betroffenen Person und sensible ärztliche Untersuchungsbefunde).

Somit bleiben insbesondere die Namen von absoluten Personen der Zeitgeschichte, literarische Quellenangaben und Aktenzeichen gerichtlicher und staatsanwaltlicher Entscheidungen unberührt (Ziff. 2.4 lit. b, c, d Verwaltungsvorschrift).

Die folgenden Merkmale werden jeweils durch einen für den gesamten Entscheidungstext individuellen Buchstaben oder eine Kombination aus mehreren Buchstaben ersetzt:

- Namen von natürlichen oder juristischen Personen, Städten und Gemeinden, Straßen und Wegen, Grundstücksbezeichnungen (Ziff. 2.5, 2.6, 2.9, 2.10 Verwaltungsvorschrift);
- Telefonnummern, Konto- oder Kundennummern, E-Mail- und (persönliche) Internetadressen, Autokennzeichen sowie Identifikationsnummern von Behörden (z.B. E-Mail-Adresse01; Ziff. 2.8 Verwaltungsvorschrift).

Alters-, Berufs- und Datumsangaben werden entfernt bzw. durch die Ziffer „0“ oder Ziffernkombinationen aus „0“ ersetzt (Ziff. 2.7, 2.11, 2.12 Verwaltungsvorschrift). Gleiches gilt grundsätzlich für die Benennung der Personen, die an der Entscheidung mitgewirkt haben (z.B. Richter; Ziff. 2.14 Verwaltungsvorschrift).

4. Abgleich im Hinblick auf die Einhaltung der Anforderungen an die Anonymisierung unter der DSGVO

Die unter 3. dargestellte Anonymisierungspraxis der Gerichte ist grundsätzlich geeignet, (zunächst) eine Anonymisierung i.S.d. DSGVO herbeizuführen.

Nicht den Anforderungen an eine Anonymisierung unter der DSGVO im eigentlichen Sinne (unbeschadet der Rechtmäßigkeit¹⁰²) entsprechen dürfte allerdings die Beibehaltung der Namen von absoluten Personen

102 S. unter DV.4.

der Zeitgeschichte. Denn die DSGVO findet ausnahmslos Anwendung auf jede natürliche Person (Art. 4 Nr. 1 DSGVO) und schließt insoweit auch absolute Personen der Zeitgeschichte ein.

Im Übrigen hängt der Erfolg einer Anonymisierung ab von den im Einzelfall ergriffenen Maßnahmen. Ziff. 2.3 Verwaltungsvorschrift erkennt insoweit ausdrücklich an, dass die dort genannten Anonymisierungsmaßnahmen „in der Regel“ und nicht in allen Fällen ausreichend sind. Ob und inwieweit eine Anonymisierung i.S.d. DSGVO erfolgt, hängt daher maßgeblich ab von der Anonymisierung der jeweiligen Entscheidung im Einzelfall.

Eine (erfolgreiche) Anonymisierung i.S.d. DSGVO liegt nahe, wenn es sich um eine einfach gelagerte Entscheidung aus der ordentlichen Gerichtsbarkeit ohne Sonderzuständigkeiten handelt (i.e. keine Arzthaftungstatbestände; keine komplexen Sachverhalte, die Umstände über eine mit einem Pseudonym gekennzeichnete Person offenbaren), die Entscheidung nicht von besonderem öffentlichem Interesse war (i.e. in der Berichterstattung keine Umstände thematisiert wurden, die die Identifizierung einzelner natürlicher Personen ermöglichen) und (nur) wenige Personen beteiligt waren.

Demgegenüber besteht ein nicht nur unerhebliches Risiko der Anwendbarkeit der DSGVO, wenn es sich um komplexe und ungewöhnliche Sachverhalte handelt. Eine dahingehende Vermutung kann sich beispielsweise ergeben aus dem Umfang des Tatbestands, aus einer Spezialzuständigkeit (z.B. einer Kammer für Arzthaftungssachen, vgl. § 348 Abs. 1 Nr. 2 lit. e ZPO) und aus der Anzahl der beteiligten und zu pseudonymisierenden Personen. Ferner können mehrere verknüpfbare Tatsacheninstanzen oder Dokumente mit sich ergänzenden Ausführungen zum Sachverhalt eine Identifizierbarkeit natürlicher Personen erleichtern. Das betrifft beispielsweise zwei aufgrund des Aktenzeichens und Verfahrensverlaufs miteinander verknüpfbare Entscheidungen (z.B. die landgerichtliche Entscheidung in der Berufung gegen eine Entscheidung des Strafrichters, Schöffengerichts oder des Amtsgerichts in Zivilsachen), und auch die jeweilige Entscheidung i.V.m. staatsanwaltlichen Entscheidungen, deren Aktenzeichen nach Ziff. 2.4 lit. d Verwaltungsvorschrift nicht anonymisiert wird. In diesen Fällen kann aufgrund eingeschränkter Verfügbarkeit des Textes der staatsanwaltschaftlichen Entscheidung allerdings mit Blick auf die breite Öffentlichkeit eine Anonymisierung i.S.d. DSGVO vorliegen.

5. Rechtmäßigkeit der Verarbeitung unter der DSGVO

Soweit die Einhaltung der Anonymisierungsanforderungen aus einer Verwaltungsvorschrift nicht zu einer Anonymisierung i.S.d. DSGVO führt, kann unter Beachtung der vorstehend dargestellten Anforderungen die Verarbeitung gegebenenfalls gleichwohl rechtmäßig sein und den DSGVO-Vorgaben genügen.

a. Verwaltungsvorschrift als Rechtsgrundlage

Zunächst ist zu fragen, ob die Verwaltungsvorschrift als eine Rechtsgrundlage unter die Öffnungsklausel des Art. 6 Abs. 1 lit. c, e, Abs. 2, 3 DSGVO fällt und damit grundsätzlich die Verarbeitung erlaubt. Für Art. 6 Abs. 1 lit. c, e, Abs. 2, 3 DSGVO erforderlich ist ein materielles Gesetz,¹⁰³ sodass nicht allgemein, sondern nur innerhalb der Verwaltung wirkende (Verwaltungs-)Vorschriften den Anforderungen nicht genügen.¹⁰⁴ Überdies zielt die Verwaltungsvorschrift gerade ab auf die Anonymisierung als eine Verarbeitung, nicht aber auf Verarbeitungen in Folge einer unzureichenden Anonymisierung.

b. Veröffentlichung (teil-)anonymisierter Gerichtsentscheidungen gestützt auf Art. 6, 9 f. DSGVO

Allerdings kommt unter Rekurs auf andere Vorschriften eine Rechtsgrundlage für die Verarbeitung gelegentlich verbliebener personenbezogener Daten¹⁰⁵ in zu veröffentlichenden Gerichtsentscheidungen in Betracht, namentlich betreffend Art. 6 Abs. 1 UAbs. 1 lit. c, e, f DSGVO¹⁰⁶ i.V.m. Art. 4 Abs. 1 BayDSG, § 3 Abs. 1 DSG NRW sowie gegebenenfalls Art. 9 Abs. 2 lit. f, Art. 10 S. 1 Var. 2 DSGVO. Eine derartige Rechtsgrundlage senkt faktisch den Anonymisierungsstandard ab, indem zwar die Gerichtsentscheidungen gegebenenfalls nicht vollständig anonymisiert sind, diese aber aufgrund der einschlägigen Rechtsgrundlage dennoch in der (teil-)anonymisierten Form zulässigerweise veröffentlicht und gegebenenfalls auch weiterverarbeitet werden können i.S.d. Art. 5 Abs. 1 lit. b, Art. 6 Abs. 4 DSGVO.

103 Frenzel, in: Paal/Pauly, Art. 6 DSGVO Rn. 16.

104 Schulz, in: Gola/Heckmann, Art. 6 DSGVO Rn. 56.

105 Zur Rechtsgrundlage für die Anonymisierung s. unter DV.4.c.aa.

106 Zur Differenzierung zwischen lit. c, e s. nachfolgend unter DV.4.b.

Die Veröffentlichung von Gerichtsentscheidungen ist eine öffentliche Aufgabe bzw. gegebenenfalls im Einzelfall sogar Ausfluss einer Pflicht der Gerichte.¹⁰⁷ Zum einen rechtfertigt dieser Befund die Anonymisierung als schonende(re) Maßnahme zum Ausgleich der Interessen der Allgemeinheit an der Veröffentlichung und der Persönlichkeits- und Datenschutzinteressen der betroffenen Personen.¹⁰⁸ Denn die Anonymisierung verwirklicht gerade die Datenschutzgrundsätze der Datenminimierung und Speicherbegrenzung i.S.d. Art. 5 Abs. 1 lit. c, e DSGVO. Zum anderen kann auch die Verarbeitung von personenbezogenen Daten in nicht vollständig anonymisierten Gerichtsentscheidungen der Erfüllung von öffentlichen Aufgaben dienen.¹⁰⁹ Diese Verarbeitung könnte in diesem Sinne (noch) als Kerntätigkeit der Justiz angesehen werden und daher – mit gleichem Ergebnis – auch auf Art. 6 Abs. 1 UAbs. 1 lit. f DSGVO gestützt werden.¹¹⁰

Die Brücke zu einer rechtlich niedergelegten Aufgabe i.S.d. Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO schlagen insoweit die Generalklauseln in Art. 4 Abs. 1 BayDSG¹¹¹ bzw. § 3 Abs. 1 DSG NRW, die als gesetzliche Regelung grundsätzlich die Zulässigkeit einer erforderlichen Verarbeitung zur Erfüllung einer der jeweiligen öffentlichen Stelle obliegenden Aufgabe vorsehen. Diese Generalklauseln können im Hinblick auf die Anforderungen aus Art. 6 Abs. 2, 3 DSGVO und die grundrechtlich unterlegten Interessen betroffener Personen allerdings keine eingriffsintensiven Verarbeitungen rechtfertigen¹¹² und setzen insbesondere die Erforderlichkeit, also nicht nur eine bloße Nützlichkeit, für die jeweilige öffentliche Aufgabe voraus.¹¹³

107 VG Stuttgart, ZD 2022, 583 (Rn. 33).

108 VG Stuttgart, ZD 2022, 583; allgemein die Zulässigkeit einer Maskierung unterstellend EDPB, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, Rn. 101; s. auch EDPB, Guidelines 01/2025 on Pseudonymisation, Rn. 23.

109 VG Stuttgart, ZD 2022, 583.

110 S. nämlich unten zu Art. 6 Abs. 1 UAbs. 2 DSGVO unter DV.4.a.

111 Art. 4 Abs. 2 BayDSG zum Grundsatz der Direkterhebung ist für die Weiterverarbeitung durch die Ministerien gestützt auf Art. 4 Abs. 2 S. 2 Nr. 2, 4, S. 3 BayDSG gewährt. Insbesondere werden überwiegende schutzwürdige Interessen der betroffenen Personen durch die justizinterne Weiterverarbeitung nicht beeinträchtigt.

112 Hornung/B. Wagner, ZD 2020, 223 (225); vgl. LfDIBW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, S. 31.

113 Zur vergleichbaren Regelung in Baden-Württemberg LfDIBW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, S. 31.

Die Notwendigkeit der Veröffentlichung von jedenfalls publikationswürdigen,¹¹⁴ anonymisierten Gerichtsentscheidungen ist anerkannt,¹¹⁵ gleich ob sie auf verfahrensrechtliche Vorschriften, gegebenenfalls in analoger Anwendung (z.B. § 299 Abs. 2 ZPO zur Akteneinsicht, §§ 474 ff. StPO in strafrechtlichen Verfahren, § 78 Abs. 1 S. 4, 6 SGB X für Sozialdaten, § 30 AO für dem Steuerverhältnis unterfallende Daten; s. Ziff. 1.2.2 Verwaltungsvorschrift), Richterrecht oder sonstige – grundrechtliche – Erwägungen gestützt wird.¹¹⁶ Hieraus folgt im Ergebnis eine hinreichende Zuweisung der öffentlichen Aufgabe, die i.V.m. der jeweiligen landesrechtlichen Generalklausel für den Bereich der Rechtsprechung den Anforderungen aus der DSGVO genügen kann.

Es ist anerkannt, dass es für einen schonenden Ausgleich zwischen dem öffentlichen und rechtsstaatlich unterlegten Interesse an der Urteilsveröffentlichung und dem Interesse betroffener Personen einer weitreichenden Anonymisierung bedarf.¹¹⁷ Dieser Ausgleich ist auf Ebene der Erforderlichkeit für die öffentliche Aufgabenerfüllung zu verwirklichen.

Mit Blick auf die für die Anonymisierung zu berücksichtigenden Faktoren, der Öffentlichkeit einer Gerichtsverhandlung und der möglichen Presseberichterstattung, kann eine vollständige Anonymisierung allenfalls selten erreicht werden.¹¹⁸ Insbesondere erfordert die öffentliche Aufgabe der Urteilsveröffentlichung einen nachvollziehbaren Entscheidungstext, selbst wenn hierdurch ein Personenbezug erhalten bleibt (z.B. durch die Wahl von einheitlichen Kürzeln für jede auftauchende Person anstelle eines einheitlichen Kürzels für alle Personen oder die Beibehaltung von Aktenzeichen¹¹⁹). Weiterhin sind der Bekanntheitsgrad der betroffenen Person,¹²⁰ die

114 Für dieses in der Rechtsprechung verwendete Merkmal fehlen klare Maßstäbe. Das *BVerwG*, NJW 1997, 2694 (2695) grenzt etwa zur grundsätzlichen Bedeutung i.S.d. Revisionsrechts ab und geht davon aus, dass sich die Veröffentlichungswürdigkeit „aus der Sicht derjenigen, die mit der Publikation erreicht werden sollen“, beurteilt.

115 *BVerfG*, ZD 2016, 77; *BVerwG*, NJW 1997, 2694; *BGH*, NJW 2018, 3123 (Rn. 14); *OLG München*, ZD 2021, 379; *Schild*, in: BeckOK Datenschutzrecht, Syst. E. Rn. 56; *OLG Celle*, NStZ 1990, 553; *OVG Lüneburg*, NJW 1996, 1489; *VG Hannover*, NJW 1993, 3282; *Walker*, JurPC 1998, 34; *Hirte*, NJW 1988, 1698; *OVG Bremen*, NJW 1989, 926; *Hoffman-Riem*, JZ 1989, 637; *Kockler*, JurPC 1996, 46.

116 Überblick etwa bei *OLG Celle*, NStZ 1990, 553 (553); *BGH*, NJW 2018, 3123 (Rn. 14); *Ludyga*, ZUM 2021, 887 (887 f.).

117 *BVerwG*, NJW 1997, 2694; *Nöhre*, MDR 2019, 136 (136) m.w.N.

118 Dementsprechend spricht bspw. das *BVerwG*, NJW 1997, 2694 (2695) auch von einer „neutralisierten Fassung“.

119 S. hierzu Ziff. 2.4 Verwaltungsvorschrift.

120 S. hierzu Ziff. 2.4 Verwaltungsvorschrift.

(zulässigen) übrigen Quellen sowie der Sensibilitätsgrad der Daten auch unterhalb der Schwellen von Art. 9, 10 DSGVO zu berücksichtigen.

Vor diesem Hintergrund fällt die Abwägung regelmäßig zugunsten der Veröffentlichung aus, wenn:

- Kennungen durch geeignete Pseudonyme ersetzt werden, z.B.:
 - o Namen werden nicht durch Initialen, sondern durch andere, eindeutige Kürzel ersetzt (z.B. „Person 1“);
 - o Ortsangaben werden nicht bloß gekürzt, sondern – soweit nicht für das Verständnis erforderlich – ebenfalls durch andere, eindeutige Kürzel ersetzt (z.B. „Anschrift 1“);
 - o E-Mail-Adressen, Telefonnummern werden ebenfalls in geeigneter Weise ersetzt;
- sensible personenbezogene Daten auch unter Inkaufnahme von Verständlichkeitseinbußen möglichst nicht mehr einer natürlichen Person zugeordnet werden können;
- gegebenenfalls einzelne Datumsangaben entfernt werden, die sich besonders zur Identifizierung einer natürlichen Person eignen, jedoch nicht verständnisfördernd sind;
- sonstige identifizierende Merkmale überprüft und gegebenenfalls bei hoher Identifizierungswahrscheinlichkeit und sehr geringem Beitrag zur Verständlichkeit der Gerichtsentscheidung entfernt bzw. neutralisiert werden.

Die unterschiedliche Anonymisierungspraxis der Gerichte dürfte, wie es das Beispiel Nordrhein-Westfalen zeigt, regelmäßig diesen Anforderungen genügen. Das gilt insbesondere für die jüngeren, strengeren Anonymisierungsrichtlinien.

Die Verarbeitung eines Restbestands an besonderen Kategorien personenbezogener Daten i.S.d. Art. 9, 10 DSGVO kann verfassungs- und grundrechtlich geboten sein, da ansonsten Gerichtsentscheidungen aus beispielsweise dem Strafrecht nie veröffentlicht werden könnten, was eine rechtstaatlich bedenkliche Lücke in die Veröffentlichungspraxis reißen würde. Insbesondere Art. 9 DSGVO hält nur wenige Ausnahmetatbestände bereit. Die Veröffentlichung einer weitgehend anonymisierten Gerichtsentscheidung kann – mit einem Restrisiko einer abweichenden Auffassung des EuGH – allerdings noch als Handlung des Gerichts im Rahmen der justiziellen Tätigkeit im konkreten Verfahren angesehen werden und daher gestützt auf Art. 9 Abs. 2 lit. f DSGVO zulässig sein. Für personenbezogene Daten über Straftaten mag das deutsche Verfassungsrecht i.V.m. den ein-

schlägigen Anknüpfungs- und Konkretisierungstatbeständen in den Verfahrens- und Prozessordnungen¹²¹ als mitgliedstaatliches Recht herangezogen werden.

Die Veröffentlichung weitgehend anonymisierter Gerichtsentscheidungen ist somit datenschutzrechtlich grundsätzlich zulässig. Voraussetzung hierfür ist regelmäßig ein hoher Anonymisierungsgrad (mind. 99 %) mit Blick auf Kennungen, wobei ein deutlich geringerer Anonymisierungsgrad sonstiger identifizierender Merkmale (u.U. sogar deutlich unter 50 %) unter Berücksichtigung der Verständlichkeitsanforderungen genügen kann. Veröffentlichte Gerichtsentscheidungen, die – wie es für Gerichtsentscheidungen aus der jüngeren Vergangenheit oft der Fall sein dürfte – diese Voraussetzungen erfüllen, oder unveröffentlichte, aber nach diesem Standard anonymisierte Urteile können daher in Ansehung einer zulässigen Veröffentlichung auch risikoarm für justizinterne Zwecke weiterverarbeitet werden. Es besteht zwar insoweit keine echte Zulässigkeitsfiktion, wohl aber die Möglichkeit der zulässigen Weiterverarbeitung für justizinterne Zwecke.

c. Keine Haftungsprivilegierung nach Art. 4 ff. Digital Services Act

Der Digital Services Act (DSA) sieht in Art. 4 ff. für die Übermittlung, die Zwischenspeicherung („Caching“) und die Speicherung („Hosting“) Haftungsprivilegien für Diensteanbieter im Internet im Hinblick auf fremde Inhalte vor. Zu fragen ist, ob diese Haftungsprivilegien für den Rückgriff eines Drittanbieters auf veröffentlichte Gerichtsentscheidungen Anwendung finden und damit die weitere Verwertbarkeit gegebenenfalls nicht vollständig anonymisierter Gerichtsentscheidungen erleichtern können. An dieser Stelle ist darauf zu verweisen, dass Art. 2 Abs. 4 lit. g DSA einen (eingeschränkten) Vorrang der DSGVO vorsieht. Die Haftungsprivilegierungen entfalten insoweit gegenüber einer Haftung nach der DSGVO für die Verarbeitung personenbezogener Daten grundsätzlich keine Wirkung.¹²² Überdies lägen ohnehin mangels durch einen Nutzer bereitgestellter Informationen die Tatbestandsmerkmale der Art. 4 ff. DSA nicht vor.

121 S. z.B. unter D.II.3.

122 T. Radtke, in: Gersdorf/Paal, Art. 6 DSA Rn. 15.

d. Sonderregelung für die Aufsicht über die justizielle Tätigkeit

Die Aufsichtsbehörden i.S.d. DSGVO sind nach Art. 55 Abs. 3 DSGVO nicht zuständig¹²³ für „die Aufsicht über die von Gerichten im Rahmen ihrer justiziellen Tätigkeit vorgenommenen Verarbeitungen“. Das betrifft die Tätigkeit der Rechtsprechung, die gerade Ausdruck der verfassungsrechtlich garantierten richterlichen Unabhängigkeit ist.¹²⁴ Nur die einzelne Veröffentlichung durch ein Gericht kann dieser Ausnahme unterfallen, nicht dagegen die Anonymisierung zahlreicher Urteile durch eine andere Stelle zur Optimierung von Arbeitsabläufen der Justiz.

e. Verschulden, Art. 82; Aufsichtsmaßnahmen, Art. 58 und 83 DSGVO

Sofern es im Einzelfall zu einem Verstoß kommt, sind gegebenenfalls etwaig getroffene Maßnahmen im Rahmen einer Schadensersatzhaftung nach Art. 82 DSGVO sowie im Rahmen von Aufsichtsmaßnahmen nach Art. 58, 83 DSGVO zu berücksichtigen.

Für die Begründung eines Schadensersatzanspruchs nach Art. 82 DSGVO genügt ein weitgefasstes Vertretenmüssen, das nur dann abzulehnen sein wird, wenn der Verantwortliche „nachweist, dass er in keinerlei Hinsicht für den Umstand, durch den der Schaden eingetreten ist, verantwortlich ist“ (Art. 82 Abs. 3 DSGVO). In Ansehung der hohen Anforderungen, die an den Entlastungsbeweis zu stellen sind,¹²⁵ dürfte die Berufung auf eine gegebenenfalls unzureichende Anonymisierungsrichtlinie regelmäßig nicht für eine Haftungsbefreiung i.S.d. Art. 82 Abs. 3 DSGVO genügen. Dabei ist insbesondere zu berücksichtigen, dass dem Verantwortlichen selbst (z.B. einem Oberlandesgericht) regelmäßig die Anonymisierungsrichtlinien zugerechnet werden können und der Verantwortliche sich insoweit nicht für Unzulänglichkeiten in der von ihm zugrunde gelegten Anonymisierungsrichtlinie entlasten kann.

Ein datenschutzrechtlich induziertes Bußgeld kann deutsche öffentliche Stellen nicht treffen (vgl. § 43 Abs. 3 BDSG, Art. 22 BayDSG, § 32 DSG NRW, Art. 83 Abs. 7 DSGVO). In das Auswahlermessen bzgl. etwaiger Aufsichtsmaßnahmen dürfte (unter anderem) einzustellen sein die Schwere des Verstoßes (Art. 83 Abs. 2 S. 2 lit. a DSGVO), die „Vorsätzlichkeit oder

123 S. zur Zuständigkeit ErwGr. 20 DSGVO.

124 Vgl. *Nguyen/Stroh*, in: Gola/Heckmann, Art. 55 DSGVO Rn. 10.

125 *EuGH*, NJW 2024, 1561 (Rn. 44 ff.).

Fahrlässigkeit des Verstoßes“ (Art. 83 Abs. 2 S. 2 lit. b DSGVO) sowie der „Umfang der Zusammenarbeit mit der Aufsichtsbehörde, um dem Verstoß abzuhelpen und seine möglichen nachteiligen Auswirkungen zu mindern“ (Art. 83 Abs. 2 S. 2 lit. f DSGVO).

6. Zwischenergebnis

Die DSGVO wird auf bereits (vermeintlich) anonymisierte und veröffentlichte Gerichtsentscheidungen anzuwenden sein. Insbesondere ältere Gerichtsentscheidungen sind gegebenenfalls nicht nach dem Stand der Technik anonymisiert und erfordern Nachbesserungen.

Am Beispiel der Anonymisierungsrichtlinien der Justiz in Nordrhein-Westfalen zeigt sich, dass der Begriff der Anonymisierung nicht stets im technischen Sinne verstanden wird. So unterfallen beispielsweise Informationen über absolute Personen der Zeitgeschichte ebenfalls der DSGVO. Selbst wenn eine Gerichtsentscheidung unter Berücksichtigung dieser Aspekte gegebenenfalls als nicht vollständig anonymisiert anzusehen ist, wird insoweit allerdings vielfach eine Rechtfertigung nach Art. 6, 9 f. DSGVO in Betracht kommen können. Neuere Anonymisierungsrichtlinien wie die der Justiz in Nordrhein-Westfalen können vor diesem Hintergrund die Grundlage für eine datenschutzkonforme Teilanonymisierung und zulässige Veröffentlichung der Gerichtsentscheidungen legen.

Im Einklang mit den Anonymisierungsrichtlinien ist ein besonderes Augenmerk auf die Anonymisierung von Entscheidungen zu legen, die nach äußeren Merkmalen (z.B. Umfang des Tatbestands, Anzahl der beteiligten Personen oder gerichtliche Spezialzuständigkeit) verschiedene Informationen enthalten können, weshalb zusammengekommen und unter Berücksichtigung weiterer Umstände natürliche Personen identifiziert können.

III. (Teil-)Anonymisierung unveröffentlichter Gerichtsentscheidungen durch das Erlanger Tool

Zur automatisierten Durchführung der Anonymisierung kommt im vorliegend zu begutachtenden Sachverhalt das sog. Erlanger Tool zum Einsatz. Das Erlanger Tool wird nachfolgend daraufhin untersucht, ob es die zuvor dargelegten datenschutzrechtlichen Anforderungen an eine Anonymisierung umsetzt und welche datenschutzrechtlichen Restrisiken gegebenenfalls bestehen.

1. Einhaltung der Anonymisierungsanforderungen im Allgemeinen

Das Erlanger Tool ist grundsätzlich geeignet, eine wirksame Anonymisierung von Gerichtsentscheidungen durchzuführen, kann aber zugleich keine pauschale Anonymisierung sämtlicher Gerichtsentscheidungen gewährleisten.

Das Erlanger Tool erzielt nach den dieser Untersuchung zugrunde gelegten Annahmen insbesondere einen Recall, d.h. einen Anteil erfolgreicher anonymisierter Merkmale an allen zu anonymisierenden Merkmalen eines Text innerhalb einer Kategorie, von 94-96 % für eindeutig identifizierende Merkmale bzw. Kennungen, wie etwa Namen und Anschriften natürlicher Personen.¹²⁶ Bei den übrigen Informationskategorien (z.B. Daten zum Prozessablauf, Aktenzeichen und Gerichtsort) wird demgegenüber nur ein Recall in einem (weiten) Spektrum von 68-96 % erzielt.¹²⁷ Daten aus diesen übrigen Informationskategorien können die Identifizierung natürlicher Personen durch die Heranziehung weiterer Quellen ermöglichen.

In Ansehung der Recall-Prozentwerte, die für eindeutig identifizierende Informationskategorien (i.e. Kennungen) hoch erscheinen, aber noch einen Spielraum in Richtung auf einen Recall-Wert von mind. 99 %¹²⁸ erkennen lassen, sind nicht alle bearbeiteten Gerichtsentscheidungen allen in Betracht kommenden Akteuren gegenüber anonym. Mit diesem Befund für sich genommen ist allerdings gerade noch keine abschließende Bewertung der datenschutzrechtlichen Zulässigkeit verbunden, denn die DSGVO kann auch die Veröffentlichung und Weiterverarbeitung von (teil-)anonymisierten Entscheidungen zulassen.¹²⁹

126 *Adrian et al.*, in: *Adrian/Kohlhase/Evert/Zwickel*, Manuelle und automatische Anonymisierung von Urteilen, S. 188–189, 194, 196 ff.

127 *Adrian et al.*, in: *Adrian/Kohlhase/Evert/Zwickel*, Manuelle und automatische Anonymisierung von Urteilen, S. 188–189, 194, 196 ff.

128 Eine 100%ige Recall-Wahrscheinlichkeit dürfte technisch bei mehreren Millionen Urteilen in Ansehung der Vielgestaltigkeit des Aufbaus der Gerichtsentscheidungen sowie möglicher unterschiedlicher Formulierung und etwaiger Tippfehler nicht erreichbar sein. Einer solchen Sicherheit bedarf es mit Blick auf die Weiterverarbeitung im Sprachmodell auch nicht.

129 S. zuvor ausf. unter D.II.5.b.

2. Wahrscheinlichkeit der Identifizierung nach allgemeinem Ermessen anhand bearbeiteter Gerichtsentscheidungen

Für die datenschutzrechtskonforme Umsetzung des GSJ-Projekts ist daher zunächst im Rahmen der Frage nach dem Vorliegen personenbezogener Daten mit Blick auf die geringen Recall-Werte für sonstige identifizierende Merkmale die Wahrscheinlichkeit der Identifizierung nach allgemeinem Ermessen zu berücksichtigen.¹³⁰ Wenn und soweit eine solche Wahrscheinlichkeit im Einzelfall bejaht wird, kann gegebenenfalls eine Rechtfertigung nach Art. 6 Abs. 1 DSGVO in Betracht kommen.¹³¹

Es ist daher im Hinblick auf die Anonymisierung zu fragen, ob die Bearbeitung der Gerichtsentscheidungen durch das Erlanger Tool eine Identifizierung natürlicher Personen mittels der Entfernung von Kennungen verhindern kann oder eine solche Identifizierung jedenfalls nach allgemeinem Ermessen hinreichend unwahrscheinlich ist im Fall einer Veröffentlichung der bearbeiteten Gerichtsentscheidung (unter a.) und im Fall einer Verwendung der bearbeiteten Gerichtsentscheidung für ein KI-Training und einen Einsatz eines trainierten KI-Systems für die Use-Cases im GSJ-Projekt (unter b.).

a. Veröffentlichung der bearbeiteten Gerichtsentscheidung

Im Fall der Veröffentlichung steht die bearbeitete Gerichtsentscheidung in vollem Umfang einem unbeschränkten Adressatenkreis zur Verfügung. Diese Adressaten werden mindestens den Zugriff auf sämtliche öffentlich verfügbaren Quellen haben (z.B. Kartendienste, Presseberichterstattung und Registerauskünfte). Das Recalldefizit des Erlanger Tools i.H.v. 4-6 % für eindeutig identifizierende Merkmale (z.B. die Anschrift einer natürlichen Person) lässt eine Identifizierung hierbei möglich erscheinen¹³² bzw. führt gegebenenfalls bei eindeutigen Kennungen (z.B. dem Namen einer natürlichen Person) unmittelbar zur Identifizierung und damit jeweils zur Annahme eines Personenbezugs i.S.d. DSGVO. Die (teilweise deutlich) schlechteren Recall-Prozentwerte bei den übrigen Merkmalskategorien in

130 S. oben unter D.I.1.b.

131 S. zur allgemeinen Veröffentlichung teilanonymer Gerichtsentscheidungen unter D.II.5.b sowie zur Weiterverarbeitung teilanonymer Gerichtsentscheidungen unter D.V.4.

132 S. beispielhaft einen solchen Versuch von *Deuber/Keuchen*, MMR 2023, 338.

Verbindung mit dem unbeschränkten Adressatenkreis erhöhen die Wahrscheinlichkeit eines Personenbezugs weiter.

Unter Verwendung des Erlanger Tools bearbeitete Gerichtsentscheidungen sind daher in dieser Konstellation regelmäßig nicht als anonym anzusehen.

Wegen der Unterschreitung eines Recall-Werts i.H.v. mind. 99 % mit Blick auf Kennungen genügt eine solche Anonymisierung regelmäßig zudem auch nicht den Anforderungen an eine Anonymisierung der Gerichtsentscheidung vor Veröffentlichung.¹³³

b. Einsatz zum KI-Training im GSJ-Projekt

Das konkrete Verwendungsszenario der bearbeiteten Gerichtsentscheidungen im GSJ-Projekt hat erheblichen Einfluss auf die Wahrscheinlichkeit einer Identifizierung.

aa. KI-Training als Anonymisierungsmaßnahme

Das KI-Training wirkt bereits wie eine (Teil-)Anonymisierungsmaßnahme, indem die bearbeitete Gerichtsentscheidung nicht einfach gespeichert wird, sondern – trotz möglicher Memorising-Effekte (d.h. der Möglichkeit zur Ausgabe von Artefakten der Trainingsdaten durch ein KI-System) – lediglich die Ausgangsbasis für eine Berechnung darstellt. In Ansehung der zu großen Teilen herausgefilterten Kennungen und identifizierenden Merkmalen sinkt die Wahrscheinlichkeit, dass sich die noch verbleibenden Trainingsdaten eindeutig¹³⁴ in dem Sprachmodell oder in den Ausgaben des KI-Systems wiederfinden. Gerade diese Ausgabe des KI-Systems ist aber maßgeblich,¹³⁵ wenn man eine mögliche Veröffentlichung auch des Sprachmodells an dieser Stelle außer Betracht lässt.

Sofern sich – stark vereinfachend – der Name „Petra Musterfrau“ aufgrund des Einsatzes des Erlanger Tools in dem gesamten Trainingskorpus nur noch in 40-60 Fällen statt zuvor in 1.000 Fällen wiederfindet, sinkt der Anteil dieser Information an den übrigen Trainingsdaten und damit grundsätzlich auch die Wahrscheinlichkeit, dass sich diese Information in

133 S. oben unter D.II.5.b.

134 S. zu Kennungen im Sprachmodell nachfolgend unter D.IV.1.

135 Vgl. oben unter D.I.1.b.bb.

einer Ausgabe wiederfindet. Durch die Tokenisierung, sprich die Aufteilung eines Wortes in zusammenhängende Zeichenfolgen, sinkt ferner die Wahrscheinlichkeit einer eindeutigen Personenzuordnung beim „Zusammenführen“ der Trainingsdaten, indem nun auf das Token „Pet“ sowohl „ra“ (für „Petra“) als auch „er“ (für „Peter“) und verschiedene andere mögliche Zeichen aus einem ähnlichen Kontext folgen könnten.

Die Wahrscheinlichkeit der Identifizierung hängt für das KI-Training im GSJ-Projekt somit maßgeblich ab von der konkreten Eingabeaufforderung an das trainierte KI-System. Ohne eine solche Eingabeaufforderung erhalten – ungeachtet einer Veröffentlichung der Gerichtsentscheidung – keine neuen Personengruppen den Zugriff auf möglicherweise personenbezogene Informationen aus den Gerichtsentscheidungen.

bb. Versehentliche Ausgabe von Trainingsdaten

In den vorbenannten Use-Cases 1-3 sind Eingabeaufforderungen zu erwarten, die auf einen Entwurf von Sachverhalts- und chronologischen Darstellungen und Gegenüberstellungen abzielen. Hierzu werden als Eingabekontext allfällige Auszüge aus der jeweils zu bearbeitenden Akte übergeben. Diese Eingabeaufforderungen lassen vorbehaltlich einer empirischen Untersuchung vereinzelte Ausgaben aus den Trainingsdaten möglich erscheinen. Eine belastbare prozentuale Wahrscheinlichkeitsprognose kann in dieser rechtlichen Untersuchung insoweit allerdings nicht vorgenommen werden.

Unter Berücksichtigung der Eingabeaufforderungen sowie der Bearbeitung der Gerichtsentscheidungen durch das Erlanger Tool dürfte eine solche beiläufige Ausgabe von – dann gegenüber den Nutzern aufgedrängten Trainingsdaten¹³⁶ – jedenfalls sehr unwahrscheinlich sein und zudem nicht signifikant abweichen von dem stets verbleibenden Risiko eines Datenlecks trotz effektiver technischer und organisatorischer Maßnahmen nach dem Stand der Technik. Die Wahrscheinlichkeit einer solchen versehentlichen Ausgabe von personenbezogenen Daten kann ferner durch die Verwendung eines entsprechenden System Prompts und durch Ausgabefilter weiter reduziert werden.

136 S. unter DV.3.b.dd.

cc. Gezielte Extraktion von Trainingsdaten

Neben der versehentlichen ist auch die gezielte Extraktion identifizierender Merkmale und damit letztlich personenbezogener Daten im Rahmen der Wahrscheinlichkeit der Anonymisierung zu betrachten. Die gezielte Extraktion aus den Trainingsdaten ist in den Use-Cases zwar nicht unmittelbar vorgesehen, könnte aber in Abweichung von den Use-Cases durch eine missbräuchliche Benutzung im Einzelfall erfolgen. Eine solche missbräuchliche Nutzung könnte dadurch erleichtert werden, dass die Struktur der Trainingsdaten bekannt ist und gegebenenfalls sogar die Trainingsdaten in Form von veröffentlichten Urteilen öffentlich einsehbar sind. Die Wahrscheinlichkeit einer solchen missbräuchlichen (erfolgreichen) Eingabeaufforderung wird nach allgemeinem Ermessen bei Richtern, Beamten und sonstigen Justizangestellten aufgrund der institutionellen, organisatorischen und gegebenenfalls vertraglichen Bindung an die Rechtmäßigkeit der Nutzung unter Berücksichtigung von Nutzungsanweisungen in einem äußerst niedrigen Bereich anzusiedeln sein.

In dem GSJ-Projekt mit den konkreten Use-Cases führt eine Bearbeitung der Gerichtsentscheidungstexte durch das Erlanger Tool somit grundsätzlich zu einer Anonymisierung der Gerichtsentscheidungen. Die hiermit bereits angedeuteten datenschutzrechtlichen Implikationen eines Sprachmodells sind sogleich unter D.IV näher zu untersuchen. Das verbleibende Restrisiko der Identifizierung natürlicher Personen ist im Hinblick auf eine datenschutzrechtliche Rechtfertigung nach Art. 6, 9 f. DSGVO nachfolgend unter D.V.4 und D.V.5 ausführlich zu prüfen.

IV. Anwendbarkeit des Datenschutzrechts auf ein (veröffentlichtes) Sprachmodells

Von dem Training eines Sprachmodells mit Gerichtsentscheidungen, die durch das Erlanger Tool behandelt wurden, können zusätzliche datenschutzrechtliche Implikationen ausgehen.

1. Sprachmodell einschließlich seiner Gewichte als Speicherung personenbezogener Daten

Ein Sprachmodell erfordert stark vereinfacht – jedenfalls – drei (Verarbeitungs-)Schritte mit Blick auf die Verarbeitung personenbezogener Trai-

ningsdaten: (a) das Training i.w.S. des Sprachmodells, (b) die Speicherung des Sprachmodells und (c) den Einsatz des Sprachmodells als Teil eines KI-Systems durch eine Eingabeaufforderung (sog. Prompt).

a. Training i.w.S.

Das Training erfolgt in mehreren Verarbeitungsschritten (so insbesondere Pre-Training für ein Basismodell, Fine-Tuning für die Anpassung z.B. an einen Einsatz als Chatbot),¹³⁷ in denen die Texte verschiedene Formatänderungen durchlaufen und die Texte nur die Grundlage für eine in dem Modell präsentierte Wahrscheinlichkeit der Aufeinanderfolge von Buchstabenketten unter Berücksichtigung von wünschenswerten, hilfreichen Ausgaben bilden.¹³⁸ Soweit die Trainingsdaten nicht vollständig anonymisiert sind, werden diese personenbezogenen Daten im Laufe des Trainingsprozesses ausgelesen, verändert, gespeichert und sonst verwendet i.S.d. Art. 4 Nr. 2 DSGVO. Diese Prozesse stellen mithin eine Verarbeitung personenbezogener Daten dar.¹³⁹ Hierbei schließt der Verarbeitungsbegriff der DSGVO insbesondere nicht die wahrscheinlichkeitsbasierte Zusammenführung von personenbezogenen Informationen aus.¹⁴⁰

Zugleich kann diese Verarbeitung als eine Anonymisierungstechnik wirken (s.o. unter D.I.3.b.cc). Es ist allerdings zu berücksichtigen, dass das Training abhängig von der Zusammensetzung des Trainingsdatensatzes einzelne, oft in ähnlicher Weise vorkommende Textabschnitte verstärkt berücksichtigen und daher gegebenenfalls eher in die Ausgabe einfließen lassen dürfte. Textbausteine, die aus einem Verfahren stammen und in andere Verfahren übernommen wurden, sind daher besonders sorgfältig darauf zu prüfen, ob sich hierin Kennungen oder sonstige identifizierende Merkmale finden.

137 S. etwa Dornis/Stober, Urheberrecht und Training generativer KI-Modelle, S. 48.

138 Zu den einzelnen Trainingsphasen *International Working Group on Data Protection in Technology*, Working Paper on Large Language Models (LLMs), S. 11 ff.

139 Hüger, ZfDR 2024, 263 (277); Schäfer, ZD 2025, 12 (14); Golland, EuZW 2024, 846 (847); Werry, MMR 2023, 911 (912); wohl auch Dewitte, in: AI Meets the GDPR, S. 140; a.A. aufgrund der Tokenisierung A. Spies, MMR 2024, 289 (290).

140 Schwartmann/Köhler, RDV 2024, 316 (316).

b. Speicherung des Sprachmodells

Mit Abschluss des Trainings könnte ein etwaiger Personenbezug (zunächst) vollständig verloren gegangen sein und erst wieder (neu) im Rahmen des Einsatzes des Sprachmodells hergestellt werden. In diesem Fall wäre die fortlaufende Speicherung des Sprachmodells keine Speicherung personenbezogener Daten i.S.d. Art. 4 Nr. 2 DSGVO und die DSGVO wäre insoweit nicht zur Anwendung berufen.

Dieser Befund betrifft insbesondere die Betroffenenrechte auf Berichtigung (Art. 16 DSGVO) und Löschung (Art. 17 DSGVO), die nur Verantwortliche aufgrund der Konzeption eines Sprachmodells vor Herausforderungen stellen, wenn in dem Sprachmodell gegebenenfalls zu berichtigende oder löschende Informationen gespeichert sind. Die weniger umstrittene Einstufung der Ausgabe personenbezogener Daten als eine Verarbeitung¹⁴¹ bliebe demgegenüber grundsätzlich ohne Auswirkungen auf das Sprachmodell insgesamt. In diesem Sinne wäre für die Berichtigung oder Löschung einer Ausgabe nur die konkrete Ausgabe, nicht aber das gesamte Sprachmodell in den Blick zu nehmen.

Die Diskussion um die Speicherung personenbezogener Daten in einem Sprachmodell wurde maßgeblich durch ein Diskussionspapier des HmbBfDI angestoßen.¹⁴² Hiernach seien die in einem Sprachmodell gespeicherten Parameter als bloße Wahrscheinlichkeitswerte nicht mit kodierten oder verschlüsselten Informationen vergleichbar.¹⁴³ Die Parameter wiesen insbesondere keinen „individuellen Informationsgehalt [...] ‚über‘ natürliche Personen“ auf.¹⁴⁴ Verfahren zur gezielten Extraktion von personenbezogenen Daten durch eine Eingabeaufforderung seien aufgrund des unverhältnismäßigen und gegebenenfalls rechtswidrigen Einsatzes nicht als Mittel eines Verantwortlichen zur Identifizierung natürlicher Personen zu berücksichtigen.¹⁴⁵

141 Hierzu nachfolgend unter D.IV.2.

142 *HmbBfDI*, Diskussionspapier: Large Language Models und personenbezogene Daten; s. auch *Fuchs/Wünschelbaum*, RDV 2024, 314.

143 *HmbBfDI*, Diskussionspapier: Large Language Models und personenbezogene Daten, S. 4; *Fuchs/Wünschelbaum*, RDV 2024, 314 (314).

144 *HmbBfDI*, Diskussionspapier: Large Language Models und personenbezogene Daten, S. 6 f.

145 *HmbBfDI*, Diskussionspapier: Large Language Models und personenbezogene Daten, S. 8.

Die übrigen Aufsichtsbehörden¹⁴⁶ und das Schrifttum zeigen sich in dieser Frage gespalten:¹⁴⁷ Es zeichnet sich allerdings eine Tendenz gegen die pauschale Annahme eines Personenbezugs ab. Das EDPB scheint maßgeblich und vorzugswürdig darauf abzustellen, ob nach allgemeinem Ermessen (bzw. im Rahmen von „reasonable means“) personenbezogene Trainingsdaten aus den Parametern abgeleitet oder über die Ausgabe im Einzelfall aus dem Modell extrahiert werden können.¹⁴⁸ Es verbleibt aber das (Rest-)Risiko, dass der EuGH gegebenenfalls einen strengen Maßstab im Rahmen des allgemeinen Ermessens anlegt.

Ohnehin gelangt das Datenschutzrecht jedenfalls dann isoliert zur Anwendung, wenn und soweit konkrete personenbezogene Daten aus dem Trainingskorpus wiedergegeben werden.

Zur weiteren Auseinandersetzung mit diesem Meinungsstreit gilt es, zunächst die Frage nach einer notwendigen Differenzierung zwischen Kennungen und sonstigen identifizierenden Merkmalen (unter aa.), der Reichweite des Personenbezugs im Lichte von sog. Mischdatensätzen (unter bb.), die Frage der Identifizierbarkeit (unter cc.), des Informationsgehalts (unter dd.) und der Anforderungen an die Speicherung (unter ee.) zu adressieren. Hierbei wird herausgearbeitet werden, dass richtigerweise zwischen Kennungen und sonstigen identifizierenden Merkmalen zu differenzieren ist,

146 Personenbezug des Sprachmodells entgegen HmbBfDI möglich nach *LfDIBW*, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, S. 9; wohl auch *BayLDA*, Datenschutzkonforme Künstliche Intelligenz, S. 5; womöglich auch *EDPB*, Report of the work undertaken by the ChatGPT Taskforce, Rn. 25; ähnl. dem HmbBfDI *Dänische Datatilsynet*, Offentlige myndigheders brug af kunstig intelligens, S. 7.

147 Dem HmbBfDI folgend *Hüger*, ZfDR 2024, 263 (277 f.); *Moos*, CR 2024, 442 (Rn. 15); *Wrobel/Pentzien*, DSB 2024, 200 (203); in der Regel kein Personenbezug nach *Golland*, EuZW 2024, 846 (847); ähnl. *Mühlhoff/Rusche-meier*, ZfDR 2024, 337 (341); gegen HmbBfDI *Hansen/Walczak*, KIR 2024, 82 (85 f.); Personenbezug ebenfalls denkbar nach *Pesch/Böhme*, MMR 2023, 917 (920); *Engeler/Rolfes*, ZD 2024, 423 (426 f.); *Kaulartz*, in: *Kaulartz/Braegelman*, Kap. 8.9, Rn. 13, 16; abhängig von den Ausgaben *Schwartmann/Köhler*, in: *Schwartmann/Keber/Zenner*, 3. Kapitel, Rn. 10; schon vor einigen Jahren einen Personenbezug des Modells in Abhängigkeit vom möglichen Output annehmend *Veale/Binns/Edwards*, Phil. Trans. R. Soc. A 376, 20180083 (6 ff.).

148 *EDPB*, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, Rn. 38, 43: „that, with reasonable means: (i) personal data, related to the training data, cannot be extracted out of the model; and (ii) any output produced when querying the model does not relate to the data subjects whose personal data was used to train the model“; *EDPB*, Report of the work undertaken by the ChatGPT Taskforce, Rn. 25.

wobei nur erstere im untersuchungsgegenständlichen Projekt vereinzelt zur Annahme eines Personenbezugs des Modells führen können. Für sonstige identifizierende Merkmale wird ein Personenbezug in Ansehung der im GSJ-Projekt wahrscheinlichen KI-Ausgaben grundsätzlich ausgeschlossen sein.

aa. Differenzierung zwischen Kennungen und sonstigen identifizierenden Merkmalen

Kennungen, wie etwa der Name einer natürlichen Person (vgl. Art. 4 Nr. 1 DSGVO), ermöglichen die direkte Identifizierbarkeit einer natürlichen Person und unterliegen nicht dem Vorbehalt der wahrscheinlichen Zuordnung. Soweit also Kennungen im KI-Modell gespeichert sind, ist ein Personenbezug anzunehmen. Zwar schadet insoweit nicht die Darstellung der Kennung in einem anderen Format, es dürften aber die Kennungen regelmäßig nur in zerlegte Bestandteile als wahrscheinliche Verbindung gespeichert sein. Eine direkte Identifizierung ist in diesen Fällen nicht möglich, sodass das Sprachmodell mit Blick auf Kennungen grundsätzlich nicht per se personenbezogen ist bzw. sich der Personenbezug auch für die Kennungen nur aus der indirekten Identifizierung auf Grundlage einer Wahrscheinlichkeitsprognose ergeben kann.

Daher ist nachfolgend für Kennungen als auch für sonstige identifizierende Merkmale vor allem nach verfügbaren Mitteln für die Annahme einer Identifizierbarkeit und nach den übrigen Anforderungen an eine Speicherung zu fragen.

bb. Reichweite des Personenbezugs – Mischdatensätze

Es könnte erwogen werden, dass ein Personenbezug und die Anwendbarkeit der DSGVO bereits deshalb ausscheiden, weil nur ganz vereinzelt personenbezogene Daten enthalten sein sollen.

Für sog. Mischdatensätze aus besonderen Kategorien personenbezogener Daten (i.e. sensiblen Daten) und sonstigen Kategorien personenbezogener Daten hat der EuGH allerdings bereits ausgeurteilt, dass „wenn ein Datensatz, der sowohl sensible als auch nicht sensible Daten enthält, Gegenstand solcher Vorgänge ist und insbesondere als Ganzes erhoben wird, ohne dass die Daten zum Zeitpunkt dieser Erhebung voneinander getrennt werden

können, ist die Verarbeitung dieses Datensatzes aber [grundsätzlich] als im Sinne von Art. 9 Abs. 1 DS-GVO untersagt anzusehen“.¹⁴⁹ Bei untrennbaren Datenverbindungen ist die Datenkategorie maßgeblich, an die die strengsten Anforderungen zu stellen sind.

Entsprechendes gilt auch für die (Vor-)Frage, ob ein Datensatz überhaupt personenbezogen ist. Sofern die Trainingsdaten sich nach dem Training des Sprachmodells in Form der Parameter nicht mehr (sinnvoll) trennen lassen, führt der Personenbezug von einzelnen Bestandteilen gegebenenfalls insgesamt zu einer Verarbeitung personenbezogener Daten im Sinne der DSGVO-Vorgaben. Denn die DSGVO sieht gerade keine Schwelle oder das Erfordernis eines überwiegenden Anteils personenbezogener Daten vor.

cc. (Wahrscheinliche) Identifizierbarkeit

Die in dem Sprachmodell gespeicherten Parameter könnten nicht eindeutig und hinreichend wahrscheinlich die Identifizierung natürlicher Personen zulassen und daher als anonym im Sinne der DSGVO zu bewerten sein.

Unterstellt man eine weitgehende Anonymisierung der Trainingsdaten und das Training als zufällig effektive Anonymisierungstechnik (z.B. aufgrund der Zusammensetzung der verwendeten Trainingsdaten),¹⁵⁰ kann eine Person gegebenenfalls nicht (mehr) anhand der Parameter unter Zuhilfenahme des Einsatzes des Sprachmodells durch eine Eingabeaufforderung identifizierbar sein. Für die Frage der wahrscheinlichen Identifizierung nach allgemeinem Ermessen kommt im GSJ-Projekt grundsätzlich¹⁵¹ nur die Ausgabe des KI-Systems in Betracht (vgl. zur Wiedergabe von Trainingsdaten in der Ausgabe die sog. Memorisierung¹⁵² und diese aus-

149 *EuGH*, ZD 2023, 664 (Rn. 89) – Meta Platforms u.a.; hierzu auch *Nabulsi*, ZD 2025, 3.

150 *Carlini et al.*, The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks, S. 2 f., 10, 12 ff.

151 S. aber nachfolgend zur möglichen Veröffentlichung des Sprachmodells unter D.I.3.

152 Hierzu etwa *Carlini et al.*, Extracting Training Data from Large Language Models; *Karamolegkou et al.*, Copyright Violations and Large Language Models; *Nasr et al.*, Scalable Extraction of Training Data from (Production) Language Models; *Kaulartz*, in: *Kaulartz/Braegelmann*, Kap. 8.9, Rn. 12; *Nolte/Finck/Meding*, Machine Learners Should Acknowledge the Legal Implications of Large Language Models as Personal Data, S. 2 m.w.N.; *Mueller et al.*, LLMs and Memorization:

nutzend sog. Model Inversion Attacks¹⁵³). Denn die justizinternen Nutzer können im Ausgangsfall nur auf das KI-System mit seiner Benutzeroberfläche zugreifen, nicht aber selbst das Sprachmodell mitsamt der enthaltenen Gewichte einsehen. Extraktionen außerhalb der Ausgabe¹⁵⁴ sind daher vorliegend außer Betracht zu lassen.

Nachfolgend sind nach einer grundlegenden dogmatischen Betrachtung (unter (1)) Faktoren für die Wahrscheinlichkeit der Identifizierung im GSJ-Projekt herauszuarbeiten (unter (2) und (3)).

(1) Dogmatische Bedenken mit Blick auf das Zusammenfallen einer Verarbeitung und der Heranziehung von Identifizierungsmitteln

Der Schritt von der Speicherung des Modells zum Einsatz im Einzelfall kann womöglich auch nur als ein dem Verantwortlichen zur Verfügung stehendes und vernünftigerweise eingesetztes Mittel zur Identifizierung natürlicher Personen angesehen werden.¹⁵⁵ In der Folge wären die in einem Sprachmodell gespeicherten Informationen auf eine natürliche Person zurückzuführen (und daher gegebenenfalls personenbezogen), womit bereits die Speicherung des Sprachmodells als eine Speicherung personenbezogener Daten anzusehen wäre.

Dem lässt sich entgegenhalten, dass die Mittel nicht gleichbedeutend sein dürfen mit einer Verarbeitung personenbezogener Daten. Mit anderen Worten gilt, dass erst der Einsatz des Sprachmodells (als Verarbeitung) mittels einer entsprechenden Eingabeaufforderung (als ein „Mittel“) die Verarbeitung personenbezogener Daten darstellt. Diese Verarbeitung wirkt nicht zurück auf vorherige Vorgänge im Zusammenhang mit den Daten, über die sich noch keine natürliche Person identifizieren ließ.

On Quality and Specificity of Copyright Compliance; hierzu auch im datenschutzrechtlichen Kontext *Maltzan/Käde*, DSRITB 2020, 505 (516).

153 *Fredrikson/Jha/Ristenpart*, in: Ray/N. Li/Kruegel, 1322; *Veale/Binns/Edwards*, Phil. Trans. R. Soc. A 376, 20180083; speziell zu sog. Membership Inference Attacks, die im vorliegenden Fall Rückschlüsse darauf zulassen könnten, ob eine natürliche Person an einem Gerichtsprozess beteiligt war, *Shokri et al.*, Membership Inference Attacks against Machine Learning Models; *Kaulartz*, in: Kaulartz/Braegelmann, Kap. 8.9, Rn. II.

154 *EDPB*, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, Rn. 38.

155 Vgl. oben unter D.I.1.b.bb.

Im Ergebnis lässt sich jedoch der einschlägigen EuGH-Judikatur die Grundtendenz entnehmen, dass die spätere Zusammenführung der identifizierenden Merkmale und übriger Informationen einen Personenbezug der zuvor getrennt gespeicherten Merkmale und Informationen begründet.¹⁵⁶

(2) Begünstigende Faktoren für eine Identifizierung

Die Möglichkeit zur Identifizierung einzelner natürlicher Personen besteht insbesondere dann fort, wenn und wo eine Vorstellung vom Inhalt und der Struktur der Trainingsdaten gegeben ist.¹⁵⁷ Für ein in der Justiz eingesetztes Sprachmodell dürfte eine solche Vorstellung regelmäßig anzunehmen sein, in der (anonymisierte) Gerichtsentscheidungen zumeist frei zugänglich und die Grundzüge der Struktur der Trainingsdaten grundsätzlich bekannt sind.

(3) Vorbehalt der wahrscheinlichen Identifizierung nach allgemeinem Ermessen

Dieser Befund der grundsätzlich möglichen Identifizierung steht allerdings unter dem Vorbehalt, dass verfügbare Mittel (wie z.B. entsprechende Eingabeaufforderungen) nach allgemeinem Ermessen wahrscheinlich zur Identifizierung genutzt werden.¹⁵⁸ Insoweit ist zwischen der zufälligen Ausgabe von personenbezogenen Trainingsdaten und der gezielten Extraktion solcher Trainingsdaten zu unterscheiden.

Die zufällige Ausgabe von personenbezogenen Trainingsdaten ist unter Berücksichtigung des Einsatzes des Erlanger Tools, einer erweiterten Anonymisierung einzelner Gerichtsentscheidungen nach Clusterung sowie des Einsatzes von System Prompts und Ausgabefiltern sehr unwahrscheinlich. Diese Prognose kann durch Beschränkungen der Eingabemöglichkeiten weiter abgesichert werden, indem z.B. die Eingabeaufforderung nicht selbst

156 S. z.B. *EuGH*, NJW 2016, 3579 – Breyer.

157 S. zur erhöhten Memorisierungswahrscheinlichkeit durch längeren Kontext *Carlini et al.*, Quantifying Memorization Across Neural Language Models, S. 5 f.; vgl. *Moos*, CR 2024, 442 (Rn. 48 ff.); *Kaulartz*, in: *Kaulartz/Braegelman*, Kap. 8.9, Rn. 13.

158 Ebenfalls die Ausgabe in diesem Zusammenhang hervorhebend *EDPB*, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, Rn. 38.

durch den Nutzer in einem Freitextfeld festgelegt werden, sondern demgegenüber nur unter festgelegten Möglichkeiten gewählt werden kann. Selbst wenn es zu einer Ausgabe kommen sollte, die eine Identifizierung ermöglichen könnte, ist es bei Verwendung der Ausgabe innerhalb der Justiz nach allgemeinem Ermessen ebenfalls als sehr unwahrscheinlich zu bewerten, dass weitere Mittel genutzt werden, um die Identifizierung tatsächlich durchzuführen.¹⁵⁹ Das Interesse an einer Identifizierung wird innerhalb der justiziellen Tätigkeit und Befassung mit einem anderen Verfahren gering sein, während zugleich dürfte der Aufwand für diese Identifizierung regelmäßig als hoch bis sehr hoch einzustufen sein dürfte.¹⁶⁰

Entsprechende Überlegungen gelten für die Möglichkeit, durch gezielte Eingabeaufforderungen Informationen zu extrahieren, die gegebenenfalls unter Heranziehung weiterer Quellen auf eine natürliche Person zurückgeführt werden. Denn auch insoweit besteht nach allgemeinem Ermessen justizintern kein Interesse an einer solchen Identifizierung.¹⁶¹ Dieser Befund wird durch zwei weitere Argumente gestützt: Zum einen kann innerhalb der Justiz gegebenenfalls schon anderweitig die – rechtlich zulässige – Möglichkeit bestehen, an eine Gerichtsentscheidung zu gelangen. Das Interesse an der aufwändigen Extrahierung von personenbezogenen Daten ist dementsprechend als noch geringer einzuschätzen. Zum anderen sind nach der Rechtsprechung des EuGH grundsätzlich¹⁶² nur rechtlich zulässige Mittel einer Identifizierung heranzuziehen. Die gezielte Extraktion durch die Richterschaft und die Justiz im Übrigen kann eine Pflichtverletzung darstellen oder sogar dem § 202a StGB oder sonstigen Strafvorschriften unterfallen.¹⁶³ Entsprechende Dienstanweisungen oder separate vertragliche Vereinbarungen für den Umgang mit einem KI-System können gegebenenfalls (soweit zulässig) diesen Maßstab weiter ausgestalten und folglich die Schwelle zu einer wirksamen Anonymisierung weiter herabsetzen. Flankiert werden sollten diese Vorgaben auch und gerade durch Log-Mechanismen, die eine missbräuchliche Nutzung im Nachhinein nachvollziehbar machen und insoweit auch präventiv wirken können.

159 S. allgemein zu diesem Interesse am Beispiel von Unternehmen *Moos*, CR 2024, 442 (Rn. 44).

160 *Moos*, CR 2024, 442 (Rn. 54 ff.).

161 Vgl. auch *Moos*, CR 2024, 442 (Rn. 52 f.).

162 S. oben unter D.I.1.b.bb.

163 Dies hervorhebend *HmbBfDI*, Diskussionspapier: Large Language Models und personenbezogene Daten, S. 8.

Eine andere Bewertung wird angezeigt sein (und erfolgt nachfolgend unter D.IV.3), sofern das KI-System oder sogar die Gewichte des Sprachmodells und gegebenenfalls der Trainingskorpus öffentlich verfügbar sind. Denn in diesem Fall sind vor allem auch sog. White-Box Attacks durch Dritte möglich, sprich Angriffe unter Ausnutzung der bekannten Struktur des Modells.¹⁶⁴

dd. Informationsgehalt

Allerdings könnte über das Merkmal des „Beziehens“ den im Sprachmodell gespeicherten Parametern kein hinreichend sicherer Informationsgehalt über eine natürliche Person zuzuordnen sein. Die Rechtsprechung verlangt, dass die Information „auf Grund ihres Inhalts, ihres Zwecks oder ihrer Auswirkungen mit einer bestimmten Person verknüpft ist“.¹⁶⁵ Die im Modell gespeicherten Parameter als Repräsentation einer Wahrscheinlichkeit des Zusammenhangs von Buchstaben(schnipseln) könnten als zu abstrakt oder zu prognostisch angesehen werden, um ihnen einen solchen Informationsgehalt über eine natürliche Person zu entnehmen.

(1) Abstraktheit

Diese Ansicht scheint der HmbBfDI zu teilen und verweist auf einen mangelnden Informationsgehalt der Parameter betreffend eine Person.¹⁶⁶ Mit Blick auf das weite Verständnis des Merkmals „beziehen“ dürfte allerdings die Rekonstruierbarkeit aussagekräftiger Informationen für ein „Beziehen“ der Informationen auf eine natürliche Person genügen.¹⁶⁷ In Abhängigkeit von den jeweiligen Informationen ist daher die abstrakte Ausgestaltung der Token bzw. ihrer wahrscheinlichkeitsbasierten Repräsentation im Modell regelmäßig unschädlich.

164 Fredrikson/Jha/Ristenpart, in: Ray/N. Li/Kruegel, 1322 (1322 f.); C. Song/Ristenpart/Shmatikov, in: Thuraisingham/Evans/Malkin/Xu, 587 (587 f.).

165 EuGH, ZD 2018, 113 (Rn. 35) – Nowak; dies wiederum weit auslegend *Generalanwalt Spielmann*, Schlussanträge C-413/23 P, Rn. 33 f.; *Klabunde/Horváth*, in: Ehmann/Selmayr, Art. 4 Rn. 10 f. S. hierzu auch schon unter D.I.1.a.

166 HmbBfDI, Diskussionspapier: Large Language Models und personenbezogene Daten, S. 6 f.

167 Veale/Binns/Edwards, Phil. Trans. R. Soc. A 376, 20180083 (7).

(2) Unsicherheit unter Berücksichtigung von Halluzinationen

Die Unsicherheit über den Wahrheitsgehalt einer Information aufgrund möglicher Halluzinationen schadet nicht per se.¹⁶⁸ Die EuGH-Judikatur zu besonderen Kategorien personenbezogener Daten lässt sich entsprechend auf sämtliche Kategorien personenbezogener Daten übertragen. So soll es bereits genügen, wenn aus einem Medikament auf mögliche Diagnosen geschlossen werden kann, auch wenn etwa bei einem rezeptfreien Medikament die Möglichkeit einer Bestellung einer natürlichen Person für eine andere Person besteht.¹⁶⁹ Auf eine Richtigkeit der Information wird es gerade nicht ankommen (arg. ex Art. 16 DSGVO).¹⁷⁰ Bei einer Verdachtsberichterstattung dürfte ebenfalls kein Zweifel an der Annahme einer Verarbeitung personenbezogener Daten bestehen. Offen bleibt, ob bei einer gänzlich unwahrscheinlichen und als solcher erkennbaren Annahme einer Verarbeitung ein „Beziehen“ auf eine natürliche Person abzulehnen ist. In Ansehung der weiten EuGH-Rechtsprechung wäre eine solche Ausnahme jedenfalls sehr restriktiv zu handhaben und auf Extremfälle zu beschränken.

ee. Speicherung

Die notwendige Heranziehung weiterer Mittel in Form der Eingabeaufforderung an das KI-System kann sich nicht nur auf Ebene der Identifizierbarkeit auswirken, sondern betrifft auch die Frage des Vorliegens einer Speicherung i.S.d. Art. 4 Nr. 2 DSGVO.

Die Speicherung wird definiert als die „Aufbewahrung personenbezogener Daten in verkörperter Form auf einem Datenträger [...] mit dem Ziel, die Daten zu einem späteren Zeitpunkt weiterverarbeiten zu können“.¹⁷¹

168 A.A. und u.a. hierauf die Ablehnung des Personenbezugs eines Sprachmodells stützend Moos, CR 2024, 442 (Rn. 18 ff.); s. hierzu auch *Artikel-29-Datenschutzgruppe*, WP 136, S. 17.

169 EuGH, NJW 2025, 33 (Rn. 83, 88).

170 EuGH, NJW 2025, 33 (Rn. 87); *Artikel-29-Datenschutzgruppe*, WP 136, S. 7; Hubert, CR 2025, 77 (Rn. 24).

171 Arning/Rothkegel, in: Taeger/Gabel, Art. 4 DSGVO Rn. 76.

(1) Identifizierung im Zusammenhang mit der Speicherung

Die Verantwortlichkeit für eine Speicherung setzt grundsätzlich insbesondere auch eine Zugriffsmöglichkeit voraus.¹⁷² Eine solche Zugriffsmöglichkeit besteht auch und gerade durch die für das Sprachmodell verantwortlichen Stellen.

Das OVG Hamburg verlangt für das Speichern insoweit einen menschlichen, aktiven Umgang mit den entsprechenden personenbezogenen Daten und verneint das Vorliegen eines solchen Umgangs im Fall des Eintritts in das Eigentum an einem Gebäude mit Blick auf darin gelagerte Akten.¹⁷³ Diese Überlegung führt in der vorliegend zu begutachtenden Konstellation allerdings nicht zu einer Ablehnung einer Speicherung. Denn das Sprachmodell wurde gerade auch durch die verantwortliche Stelle trainiert.

Dennoch bestehen Zweifel, ob für die Speicherung als rein passiver Vorgang (vgl. Art. 18 Abs. 2 DSGVO)¹⁷⁴ die gleichen Anforderungen an die Einbeziehung von identifizierenden Mitteln gelten wie für andere, aktive(re) Verarbeitungsvorgänge (wie z.B. die Verwendung). Denn das Risiko der Identifizierbarkeit realisiert sich erst dann, wenn über die Speicherung hinaus weitere Daten erhoben und gegebenenfalls abgeglichen werden. Insoweit ist auf Art. 4 Nr. 2, Art. 18 DSGVO zu verweisen, die zeigen, dass grundsätzlich auch die Speicherung auf denselben Begriff der personenbezogenen Daten abstellt.¹⁷⁵

Es kann für den Personenbezug nach den Schlussanträgen des EuGH-Generalanwalts Spielmann in der Rs. C-41/23 P also das jeweilige Verhältnis und die Verpflichtung zu berücksichtigen sein,¹⁷⁶ sodass die Annahme personenbezogener Daten in dem konkreten Verhältnis mit Blick auf die Speicherung gegebenenfalls einer anderen Bewertung unterfallen mag.

Weiterhin kommt in Betracht, dass die „Speicherung“ des Sprachmodells aufgrund des Begriffs der Speicherung i.S.d. Art. 4 Nr. 2 DSGVO keine Verarbeitung personenbezogener Daten darstellt. Die für Ende 2025 zu

172 Herbst, in: Kühling/Buchner, Art. 4 Nr. 2 DSGVO Rn. 24; eine Herrschaft über die Aufbewahrung erfordernd Roßnagel, in: Simitis/Hornung/Spiecker gen. Döhmman, Art. 4 Nr. 2 DSGVO Rn. 19.

173 OVG Hamburg, BeckRS 2020, 30248 (Rn. 21).

174 Danach hat die betroffene Person bspw. bei Unklarheiten betreffend die Richtigkeit der personenbezogenen Daten das Recht auf Einschränkung der Verarbeitung. In diesem Fall bleibt die Speicherung als eingriffsarme Verarbeitung nach Art. 18 Abs. 2 DSGVO zulässig.

175 Hubert, CR 2025, 77 (Rn. 31).

176 Generalanwalt Spielmann, Schlussanträge C-413/23 P, Rn. 72.

erwartende EuGH-Entscheidung in der Rs. C-413/23 P könnte weitere Anhaltspunkte hierfür bieten.

(2) Aufbewahrung als Ausgangspunkt für eine Reproduktion

Gegen die Annahme einer Speicherung lässt sich ferner anführen, dass die Trainingsdaten in einem Sprachmodell nicht aufbewahrt werden, sondern lediglich die Ausgangsbasis für die Produktion von gänzlich neuen auszugebenden Informationen basierend auf Wahrscheinlichkeiten darstellen. Allerdings sind in dem Sprachmodell gerade wahrscheinliche Ableitungen aus den Trainingsdaten gespeichert, die eine Zuordnung von in den Trainingsdaten enthaltenen Informationen zu einer Person deutlich wahrscheinlicher machen als dies beispielsweise bei einem von Grund auf neu geschriebenen Text der Fall ist. Mit anderen Worten gilt: Wenn die Trainingsdaten einen engen Zusammenhang zwischen „Herrn Müller“ und der „Rathausstraße“ nahelegen, besteht eine hohe Wahrscheinlichkeit, dass sich durch eine entsprechende Eingabeaufforderung dieser Zusammenhang wiederherstellen lässt. Dieser Zusammenhang wurde aber nicht neu erdacht, sondern ist gerade aufgrund der aus den Trainingsdaten abgeleiteten Zusammenhänge wahrscheinlich für eine Ausgabe. Ebendiese Zusammenhänge sind in dem Sprachmodell repräsentiert und insoweit aufbewahrt.

Mit Blick auf die bisherige Judikatur des EuGH¹⁷⁷ sprechen gewichtige Argumente dafür, dass der EuGH unter diesen Umständen von einer Speicherung ausgehen wird; gleichwohl ist auch eine abweichende Entscheidung mit Blick auf das Merkmal der Aufbewahrung denkbar.

ff. Zwischenergebnis

Die Speicherung eines Sprachmodells kann als die Verarbeitung personenbezogener Daten zu bewerten sein, wenn in den Trainingsdaten (noch) personenbezogene Daten enthalten waren.¹⁷⁸ Hierfür genügt bereits das

177 S. auch die Verarbeitung personenbezogener Daten im Zusammenhang mit einem Wahrscheinlichkeitswert (Score) in *EuGH*, MMR 2024, 153 – Schufa; s. aber früher zur Differenzierung zwischen Wahrscheinlichkeitswert und personenbezogenen Daten *BGH*, ZD 2014, 306.

178 *EDPB*, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, Rn. 27-34.

vereinzelte Vorkommen personenbezogener Daten, die sich nicht von nicht-personenbezogenen Daten trennen lassen. Die wahrscheinkeitsorientierte Architektur genügt den Anforderungen an einen Informationsgehalt über eine natürliche Person. Zudem dürfte eine Speicherung trotz Reproduktion anzunehmen sein, wobei diese Frage allerdings als (noch) weitgehend ungeklärt einzustufen ist.

Die Bewertung des Vorliegens von Anonymität hängt – im Einklang mit der Auffassung des EDPB – maßgeblich davon ab, ob entweder Kennungen eindeutig gespeichert sind oder die Nutzer des Sprachmodells und eines hierauf aufbauenden KI-Systems über die tatsächlichen und rechtlichen Möglichkeiten verfügen, über Eingabeaufforderungen personenbezogene Daten zu extrahieren. Eindeutige Kennungen werden regelmäßig nicht nach herkömmlichem Verständnis in dem Modell „gespeichert“. Vielmehr wird es auch und gerade darauf ankommen, ob sich Kennungen und sonstige identifizierende Merkmale über die Ausgabe des KI-Systems extrahieren lassen.

Ein justizintern eingesetztes KI-System (i.e. ohne Möglichkeit, die Gewichte des Sprachmodells einzusehen), dessen Einsatz zur Extrahierung personenbezogener Daten rechtswidrig ist und für das durch technische Mittel (z.B. Ausgabefilter und System Prompt) die Extrahierung verhindert wird, kann daher als anonym zu bewerten sein. Eine solche Bewertung setzt die Beachtung weiterer technischer und organisatorischer Maßgaben voraus. So sollten etwa behörden- bzw. justizinterne Leitlinien zum Umgang mit dem KI-System den Einsatz des KI-Systems absichern, um das Restrisiko einer abweichenden Bewertung durch die Rechtsprechung zu minimieren.

2. Verarbeitung durch Einsatz des Sprachmodells im Einzelfall

Durch den Einsatz des Sprachmodells als KI-System mit einer Eingabeaufforderung im Einzelfall kann es zu einer Verarbeitung personenbezogener Daten kommen. Eine solche Verarbeitung wirkt über die Identifizierungsmittel nicht nur zurück auf die zuvor untersuchte Fragestellung des Personenbezugs eines Sprachmodells, sondern ist darüber hinaus auch eigenständig an der DSGVO zu messen.

Zur Wahrung der Anonymität des Sprachmodells kann zunächst auf die oben unter D.IV.1.b.cc(3) erörterten Sicherungsmechanismen verwiesen werden. Soweit diese Sicherungsmechanismen einen Personenbezug effek-

tiv verhindern, stellen weder interne Vorgänge innerhalb des Sprachmodells noch die Ausgabe eine Verarbeitung personenbezogener Daten im Sinne der DSGVO dar; im Übrigen kommt eine rechtfertigende Rechtsgrundlage nach Art. 6 DSGVO in Betracht.¹⁷⁹

Für den unwahrscheinlichen Fall einer Offenlegung personenbezogener Daten in der Ausgabe handelt es sich um eine Verletzung des Schutzes personenbezogener Daten (Art. 4 Nr. 12 DSGVO), die gegebenenfalls Meldepflichten nach Art. 33 f. DSGVO auslösen kann. Hieraus folgt keinesfalls zwangsläufig rückwirkend ein Personenbezug des gesamten Sprachmodells, sofern eine Identifizierung im Einzelfall nach allgemeinem Ermessen sehr unwahrscheinlich war und bleibt.

Die sog. Halluzination, d.h. die Ausgabe unrichtiger personenbezogener Daten, stellt ebenfalls eine Verarbeitung dar.¹⁸⁰ Eine Halluzination, die maßgeblich auf den Eingabedaten beruht (z.B. einem eingegebenen Namen und einer hierzu „halluzinierten“ Information), wirkt allerdings nicht auf die Identifizierbarkeit anhand des trainierten Sprachmodells zurück, sondern ist als Verarbeitung der Eingabedaten eigenständig zu betrachten.

3. Anforderungen an die Veröffentlichung des Sprachmodells mit Blick auf die Anonymisierung

Die Veröffentlichung des Sprachmodells kann sich mit Blick auf die Mittel zur Identifizierung natürlicher Personen auf die Frage nach deren Anonymität und der Anwendbarkeit der DSGVO auswirken. Dieser Befund gilt unabhängig davon, ob eine solche Veröffentlichung z.B. mit Blick auf den fair trial-Grundsatz¹⁸¹ oder aus Gründen einer transparenten Verwaltung gewünscht ist.

Für die Veröffentlichung ist zwischen – jedenfalls – drei Stufen zu unterscheiden: Zurverfügungstellung eines KI-Systems (unter a.), Veröffentlichung des Sprachmodells (unter b.) und des Trainingskorpus (unter c.). Mit fortschreitender Stufe nimmt die Wahrscheinlichkeit einer tatsächlich erfolgreichen Anonymisierung jeweils ab.¹⁸²

179 S. unter DV.4.

180 Arg. Art. 16 DSGVO und s. schon oben unter D.IV.1.b.dd(2) sowie *noyb*, Complaint against OpenAI.

181 *Biallaß*, in: Ory/Weth, Kap. 8, Rn. 372.

182 Vgl. *Roßnagel*, DuD 2024, 513 (519); *Kaulartz*, in: Kaulartz/Braegelman, Kap. 8.9, Rn. 7.

a. Zurverfügungstellung des KI-Systems

Bei einer Zurverfügungstellung des KI-Systems in Form eines Chatbots erhält zwar ein gegebenenfalls unbeschränkter Personenkreis den Zugriff, dies beschränkt sich aber auf die Eingabeaufforderung und im System implementierte Filter. Anders als bei einem Einsatz innerhalb der Justiz bestehen vielgestaltige Interessenlagen, die gezielte Attacks zur Extrahierung personenbezogener Daten wahrscheinlicher werden lassen. Durch die zuvor durchgeführte Anonymisierung der Trainingsdaten sowie System Prompts und Ausgabefilter besteht weiterhin ein erheblicher Aufwand für Extrahierungsangriffe. In dieser Konstellation kommt daher zwar grundsätzlich eine Nichtanwendbarkeit des Datenschutzrechts in Betracht, ist aber von Faktoren wie der Urteilsanonymisierung, den öffentlich zur Verfügung stehenden Quellen und der Effektivität von Filtern im KI-System abhängig.

Sofern das Datenschutzrecht zur Anwendung berufen ist, kann gegebenenfalls eine Rechtsgrundlage aus Art. 6 DSGVO eingreifen (siehe nachfolgend unter DV.4). Ferner sind im Hinblick auf die weltweite Abrufbarkeit des KI-Systems etwaige Drittstaaten-Übermittlungen zu beachten und gegebenenfalls zu beschränken.¹⁸³

b. Veröffentlichung des trainierten Sprachmodells (Open Source)

Mit einer Veröffentlichung des Sprachmodells und seiner Parameter entfallen sämtliche Filter- und sonstigen Maßnahmen des KI-Systems. In diesem Sinne werden sog. White Box Attacks,¹⁸⁴ abzugrenzen von den Black Box Attacks im vorgenannten Szenario,¹⁸⁵ durch die Kenntnis der Modellparameter durch jedermann möglich. Vor diesem Hintergrund steigt die Wahrscheinlichkeit einer Anwendbarkeit der DSGVO signifikant und ermittelt sich (fast) nur noch in Abhängigkeit davon, wie erfolgreich die

183 Ashkar, ZD 2023, 523 (530); DSK, Orientierungshilfe Künstliche Intelligenz und Datenschutz, Rn. 17.

184 Fredrikson/Jha/Ristenpart, in: Ray/N. Li/Kruegel, 1322 (1322 f.); C. Song/Ristenpart/Shmatikov, in: Thuraishingham/Evans/Malkin/Xu, 587 (587 f.).

185 Ico, <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-should-we-assess-security-and-data-minimisation-in-ai/>; Kaulartz, in: Kaulartz/Braegelman, Kap. 8.9, Rn. 7.

Trainingsdaten anonymisiert wurden. Eine datenschutzrechtliche Rechtsgrundlage für die Verarbeitung aus Art. 6 DSGVO kommt vor allem dann in Betracht,¹⁸⁶ wenn enthaltene Gerichtsentscheidungen in dieser Form ohnehin hätten veröffentlicht werden dürfen.¹⁸⁷

c. Veröffentlichung des Trainingskorpus (Open Source)

Sofern darüber hinaus auch der gesamte Trainingskorpus veröffentlicht wird, hängt die Anwendbarkeit der DSGVO (nur) noch davon ab, ob die Trainingsdaten selbst als anonym zu betrachten sind. Da das Erlanger Tool keine 100%ige Anonymisierung gewährleisten kann und zugleich die gesamte Öffentlichkeit den Zugriff auf die gerichtlichen Entscheidungen hat, sprechen gewichtige Gründe für die Annahme einer Anwendbarkeit der DSGVO. Gleichwohl ist auch in diesem Fall eine Rechtfertigung der betreffenden Datenverarbeitung nach Art. 6 DSGVO (theoretisch) denkbar.¹⁸⁸ Das betrifft insbesondere Gerichtsentscheidungen, die den allgemeinen (hohen) Anforderungen an die Veröffentlichung entsprechen.¹⁸⁹

V. Datenschutzrechtliche Anforderungen im Fall des Vorliegens personenbezogener Daten

Wenn und soweit personenbezogene Daten verarbeitet werden, was im GSJ-Projekt mit Blick auf das trainierte Sprachmodell bei Umsetzung hinreichender technischer und organisatorischer Maßnahmen nicht der Fall ist (s. hierzu unter D.IV.1.b.ff.), ergeben sich aus der DSGVO entsprechende Anforderungen an diese Verarbeitung.

1. Relevante Verarbeitungen

Die datenschutzrechtlichen Anforderungen knüpfen an die Verarbeitung als „jeden mit oder ohne Hilfe automatisierter Verfahren ausgeführten Vorgang oder jede solche Vorgangsreihe im Zusammenhang mit personen-

186 S. nachfolgend unter D.V.4.c.cc.

187 S. hierzu unter D.II.5.b.

188 S. nachfolgend unter D.V.4.c.cc.

189 S. hierzu unter D.II.5.b.

bezogenen Daten“ nach Art. 4 Nr. 2 DSGVO. Der nicht-abschließende Katalog der Vorschrift nennt beispielhaft Vorgänge wie das Erheben und Erfassen personenbezogener Daten, wobei die konkrete Einordnung vielfach offenbleiben kann.

Für das untersuchungsgegenständliche Projekt in Betracht kommen insbesondere Verarbeitungen im Zusammenhang mit der (teilweisen) Anonymisierung, dem Training, einer etwaigen Speicherung im Sprachmodell und dem Einsatz des Sprachmodells und der Ausgabe.

2. Verhältnis zum mitgliedstaatlichen Recht

Von der DSGVO abweichende Anforderungen nach mitgliedstaatlichem Recht können Bedeutung erlangen, wenn und soweit sich das mitgliedstaatliche Recht auf eine Öffnungsklausel innerhalb der DSGVO stützen kann.

Hierzu kommen neben den nachfolgend unter D.V.4 zu erörternden Art. 6 Abs. 1 lit. c, e, Abs. 2, 3, 4 DSGVO insbesondere in Betracht: Art. 23 Abs. 1 lit. f DSGVO mit Blick auf Einschränkungen von Betroffenenrechten zum Schutz der Unabhängigkeit der Justiz und des Schutzes der Gerichtsverfahren, Art. 85 Abs. 1, 2 DSGVO mit Blick auf weitgehende Abweichungen verschiedener DSGVO-Kapitel zur Wahrung der Meinungs- und Wissenschaftsfreiheit, Art. 86 DSGVO mit Blick auf Einschränkungen der DSGVO unter Wahrung des Wesenskerns¹⁹⁰ für den Zugang zu amtlichen Dokumenten¹⁹¹ sowie Art. 89 DSGVO mit Blick auf weitgehende Einschränkungen der DSGVO-Vorschriften für wissenschaftliche Forschungszwecke.

Allerdings setzen diese Vorschriften jeweils Unions- oder mitgliedstaatliches Recht¹⁹² voraus, das von einer dieser Öffnungsklauseln für die Verarbeitung personenbezogener Daten im Zusammenhang mit dem Training eines Sprachmodells Gebrauch macht. Eine solche (einschlägige) unionsrechtliche oder mitgliedstaatliche spezifische Regelung ist vorliegend aber gerade nicht ersichtlich.

Die KI-VO lässt die DSGVO ebenfalls nach Art. 2 Abs. 7 KI-VO unberührt (siehe auch ErwGr. 10 KI-VO).

190 Schnabel, in: Simitis/Hornung/Spiecker gen. Döhmman, Art. 86 DSGVO Rn. 30 ff.

191 Zu Gerichten als Behörden etwa Schnabel, in: Simitis/Hornung/Spiecker gen. Döhmman, Art. 86 DSGVO Rn. 20.

192 Buchner/Petri, in: Kühling/Buchner, Art. 6 DSGVO Rn. 84.

3. Verantwortlichkeit

Die Adressaten des Pflichtenkanons der DSGVO richten sich maßgeblich nach der datenschutzrechtlichen Verantwortlichkeit.

a. Anforderungen

Soweit die DSGVO zur Anwendung berufen ist, trifft der umfangreiche Pflichtenkatalog grundsätzlich den für die Verarbeitung – allein oder gemeinsam – Verantwortlichen. Der Verantwortliche ist nach Art. 4 Nr. 7 Hs. 1 DSGVO „die natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, die allein oder gemeinsam mit anderen über die Zwecke und Mittel der Verarbeitung von personenbezogenen Daten entscheidet“.

Die Entscheidung oder Festlegung setzt regelmäßig die tatsächliche¹⁹³ Einflussnahme und Steuerung der Verarbeitung aus Eigeninteresse voraus.¹⁹⁴ Die Festlegung erfordert hierbei ausgeübte Entscheidungsgewalt mit Blick auf die wesentlichen Umstände der Verarbeitung,¹⁹⁵ wobei diese Festlegung zu einem gewissen Grad abstrakt bleiben kann.¹⁹⁶ Bezugspunkt für die Entscheidung müssen die Zwecke und die wesentlichen Mittel der Verarbeitung sein, sprich die Kategorien personenbezogener Daten und Kategorien betroffener Personen sowie die Speicherdauer und etwaige Empfänger.¹⁹⁷ Eines sicheren Wissens um das Vorliegen personenbezogener Daten bedarf es hierbei nicht,¹⁹⁸ ebenso wenig wie eines tatsächlichen Zugriffs auf personenbezogene Daten durch einen Verantwortlichen.¹⁹⁹

Wenn und soweit mehrere Verantwortliche gemeinsam entscheiden, liegt eine gemeinsame Verantwortlichkeit (nur) für den konkreten Verarbeitungsvorgang vor,²⁰⁰ wobei diese gemeinsame Verantwortlichkeit zusätz-

193 Abzugrenzen von der bloßen formalen Stellung *EDPB*, Leitlinien 07/2020 zu den Begriffen „Verantwortlicher“ und „Auftragsverarbeiter“ in der DSGVO, Rn. 12.

194 *EuGH*, ZD 2018, 469 (Rn. 68) – Zeugen Jehovas.

195 *EDPB*, Leitlinien 07/2020 zu den Begriffen „Verantwortlicher“ und „Auftragsverarbeiter“ in der DSGVO, Rn. 40.

196 *T. Radtke*, Gemeinsame Verantwortlichkeit unter der DSGVO, S. 133 ff.

197 *EDPB*, Leitlinien 07/2020 zu den Begriffen „Verantwortlicher“ und „Auftragsverarbeiter“ in der DSGVO, Rn. 40.

198 Vielmehr genügt die abstrakte Kenntnis um stattfindende Verarbeitungen, s. *T. Radtke*, Gemeinsame Verantwortlichkeit unter der DSGVO, S. 135 ff.

199 *EuGH*, NJW 2018, 2537 (Rn. 38) – Wirtschaftsakademie.

200 *EuGH*, ZD 2019, 455 (Rn. 74 ff.) – Fashion ID.

liche Pflichten und eine gesamtschuldnerische Haftung nach Art. 26 Abs. 3, Art. 82 Abs. 4 DSGVO auslöst. Indizien für die Annahme einer gemeinsamen Verantwortlichkeit sind nach der weiten Auslegung der maßgeblichen Bestimmungen durch den EuGH für betroffene Personen nicht erwartbare Übermittlungen sowie sich ergänzende²⁰¹ Verarbeitungsbeiträge und Zwecke bezogen auf gemeinsame Datensätze bei gleichzeitigen Elementen der Kooperation.²⁰²

Die Annahme einer Verantwortlichkeit ist insbesondere dann zu rechtfertigen, wenn sich aufgrund der Beteiligung des betreffenden Akteurs das bestehende datenschutzrechtliche Risiko für betroffene Personen erhöht oder ein neues Risiko entsteht (z.B. durch Zweckfestlegungen oder die Änderung wesentlicher Mittel, wie etwa der Speicherdauer oder der Datenempfänger). Vor diesem Hintergrund stuft das EDPB etwa die Verbesserung eines bereits existierenden Dienstes aus einem Nutzungsverhältnis und die Betrugsprävention als eigenständige Zwecke ein,²⁰³ deren Festlegung eine (gemeinsame) Verantwortlichkeit indizieren kann.

In den Fällen des Art. 6 Abs. 1 UAbs. 1 lit. c, e DSGVO kann die Verantwortlichkeit durch Rechtsvorschrift nach Art. 4 Nr. 7 Hs. 2 DSGVO festgelegt werden.²⁰⁴

Eine datenschutzrechtliche Stelle, die nur punktuell Verarbeitungen für einen Verantwortlichen und nach dessen Weisungen ausführt, unterliegt als Auftragsverarbeiter nach Art. 4 Nr. 8, Art. 26 DSGVO einem nur eingeschränkten Pflichtenkatalog. Insbesondere bedarf die Übermittlung personenbezogener Daten zwischen Verantwortlichem und Auftragsverarbeiter keiner datenschutzrechtlichen Rechtsgrundlage nach Art. 6, 9 DSGVO.²⁰⁵

b. Einordnung der Rollen der Ministerien, der ausführenden Stellen sowie der anwendenden Justizbediensteten und Richter

In dem untersuchungsgegenständlichen GSJ-Projekt bedarf es insbesondere einer Einordnung der Rollen der Ministerien, der ausführenden Stellen sowie der ein trainiertes KI-System anwendenden Gerichte. Hierbei ist

201 Vgl. *EuGH*, ZD 2024, 209 (Rn. 43); *EDPB*, Leitlinien 07/2020 zu den Begriffen „Verantwortlicher“ und „Auftragsverarbeiter“ in der DSGVO, Rn. 55.

202 *EuGH*, NJW 2018, 2537 (Rn. 31 ff.) – Wirtschaftsakademie; ZD 2024, 328 (Rn. 59 ff.) – IAB Europe.

203 *EDPB*, Leitlinien 02/2019, Rn. 48-50.

204 *T. Radtke*, Gemeinsame Verantwortlichkeit unter der DSGVO, S. 105 f.

205 *Ausf. Gabel/Lutz*, in: *Taeger/Gabel*, Art. 28 DSGVO Rn. 11 m.w.N.

zwischen mehreren Projektphasen mit unterschiedlichen Verarbeitungsvorgängen zu differenzieren.

Anders als in der Konstellation von sog. aufgedrängten personenbezogenen Daten²⁰⁶ wird es vorliegend in Ansehung des abstrakten Wissens der Ministerien um stattfindende Verarbeitungen unschädlich sein, dass die Verarbeitung personenbezogener (Trainings-)Daten gerade nicht gewollt ist.

aa. (Teilweise) Anonymisierung

Eine (teilweise) Anonymisierung stellt ebenfalls eine Verarbeitung dar.²⁰⁷ Die Entscheidung über die Anonymisierung steht in engem Zusammenhang mit der späteren Verwendung, die durch die Anonymisierung ermöglicht werden soll. Für eine Anonymisierung, die aus der ex ante-Perspektive ausschließlich die Verwendung zum KI-Training und dem späteren Einsatz in der Justiz ermöglichen soll und dementsprechend durch ein Ministerium im Interesse der eigenen Justiz beauftragt wird, ist das jeweilige Ministerium der datenschutzrechtlich Verantwortliche. Denn das Ministerium entscheidet in diesen Fällen nicht nur über den Zweck, sondern auch über die heranzuziehende Datengrundlage in Form der Entscheidung, mithin auch über Kategorien der personenbezogenen Daten (z.B. über den Ausschluss von Arzthaftungssachen mit möglichen Gesundheitsdaten), sowie den Kreis und die Kategorien betroffener Personen.

Die ausführenden Stellen können vertraglich an die Weisungen der verantwortlichen Ministerien gebunden und sodann als Auftragsverarbeiter nach Art. 4 Nr. 8, Art. 28 DSGVO eingesetzt werden. Soweit die ausführenden Stellen allerdings personenbezogene Daten in Abweichung von den Weisungen oder aus Eigeninteresse an einer weiteren Forschungsnutzung personenbezogener Daten verarbeiten, werden sie selbst als Verantwortliche anzusehen sein bzw. gelten als solche (vgl. Art. 28 Abs. 10 DSGVO). Die ausführenden Stellen müssen in der Folge insbesondere entsprechend der jeweiligen Auftragsverarbeitungsvereinbarung hinreichende technische und organisatorische Maßnahmen zur Datensicherheit treffen.

206 Hierzu etwa *L. Meyer*, RD 2025, 125.

207 S. schon oben unter D.II.5.b zur Veröffentlichung teilanonymisierter Gerichtsentscheidungen sowie zur Anonymisierung als Verarbeitung nachfolgend unter DV.4.c.aa.

bb. KI-Training

Ähnliche Erwägungen gelten mit Blick auf das KI-Training einschließlich der vorbereitenden Zusammenstellung der Trainingsdaten. Wenn und soweit das Training einschließlich des notwendigen Debuggings allein für den späteren Einsatz in der Justiz erfolgt, werden die Ministerien als Verantwortliche und die ausführenden Stellen als Auftragsverarbeiter zu qualifizieren sein. Ein nach eigenem Ermessen der ausführenden Stellen durchgeführtes Debugging ändert nichts an dem Risiko der Verarbeitungsvorgänge für die betroffenen Personen und bewegt sich als Entscheidung über unwesentliche Mittel der Verarbeitung zudem im Rahmen des Spielraums eines Auftragsverarbeiters, soweit und solange die Ministerien den Zweck des KI-Trainings für das GSJ-Sprachmodell sowie die Speicherdauer, Datenempfänger und die Kategorien der betroffenen Personen selbst vorgeben.

Da beide Ministerien allfällige Gerichtsentscheidungen übermitteln und den Einsatz des trainierten KI-Modells in ihrem Verantwortungsbereich beabsichtigen, kommt mit Blick auf das Training eine gemeinsame Verantwortlichkeit in Betracht. Sobald sich die personenbezogenen Daten aus unterschiedlichen Gerichtsentscheidungen im Rahmen des Trainings des Sprachmodells nicht (mehr) klar voneinander trennen lassen, sondern ein gemeinsamer Datenpool im Interesse beider Ministerien entsteht, wird eine gemeinsame Verantwortlichkeit der Ministerien anzunehmen sein. Somit bedarf es einer Verantwortlichkeitszuteilung zwischen beiden Ministerien, die den Anforderungen des Art. 26 Abs. 1 S. 2, 3, Abs. 2 S. 1 DSGVO entspricht und deren wesentlicher Inhalt den betroffenen Personen nach Art. 26 Abs. 2 S. 2 DSGVO zur Verfügung gestellt wird. Außerdem resultiert hieraus eine gesamtschuldnerische (zivilrechtliche) Haftung (Art. 26 Abs. 3, Art. 82 Abs. 4 DSGVO).

Die gemeinsame Beauftragung eines Auftragsverarbeiters durch die Ministerien ist aufgrund der gemeinsamen Verantwortlichkeit mit zusammenhängenden Interessen möglich,²⁰⁸ wenn und soweit sich die Ministerien auf einheitliche Weisungen verständigen.

208 T. Radtke, Gemeinsame Verantwortlichkeit unter der DSGVO, S. 416.

cc. Speicherung des KI-Modells

Die Speicherung des KI-Modells ist datenschutzrechtlich (nur) relevant, wenn und soweit Daten nach allgemeinem Ermessen einer natürlichen Person zugeordnet werden können.²⁰⁹ Selbst wenn sich in den Trainingsdaten noch personenbezogene Daten befinden, kann das Training als Anonymisierung dergestalt wirken, dass die betreffenden Daten nicht (mehr) mit hinreichender Wahrscheinlichkeit zusammengeführt und einer natürlichen Person zugeordnet werden können. Hierzu trägt nicht nur das Training selbst bei, sondern auch allfällige Maßnahmen auf Ebene der Ausgabe (z.B. Ausgabefilter).

Falls eine solche Zuordnung möglich ist, enthält das Sprachmodell – eingebettet in das KI-System – personenbezogene Daten und es stellt sich die Frage nach der Verantwortlichkeit für die Speicherung des Modells als Verarbeitung personenbezogener Daten. Die Speicherung dürfte regelmäßig durch ein Ministerium als Verantwortlichen²¹⁰ unter Einbindung eines Rechenzentrumsanbieters als Auftragsverarbeiter gesteuert werden. Ein gemeinsamer Betrieb eines KI-Systems durch beide Ministerien, die zuvor bereits die Trainingsdaten zur Verfügung gestellt haben und das KI-System nun zu eigenen Zwecken einsetzen, wird wiederum eine gemeinsame Verantwortlichkeit darstellen.

dd. Einsatz des KI-Systems

Durch den Einsatz des KI-Systems kann es sowohl zu einer Verarbeitung personenbezogener (Trainings-)Daten kommen, wenn bereits das Sprachmodell selbst einen Personenbezug aufweist, als auch wenn das Sprachmodell mangels Wahrscheinlichkeit der Identifizierungen keinen Personenbezug aufweist, ein solcher unwahrscheinlicher Fall allerdings aufgrund der konkreten Eingabeaufforderung im Einzelfall auftritt.

In der ersten Konstellation setzt sich die – gegebenenfalls gemeinsame – Verantwortlichkeit des Ministeriums fort. Denn das Ministerium hat hier maßgeblich über den Einsatz des Sprachmodells mit Personenbezug sowie die Zusammensetzung der (personenbezogenen) Trainingsdaten entschieden. Hinzutreten können weitere Verantwortliche aus dem Nutzerkreis

209 ErwGr. 26 S. 3 DSGVO sowie s. oben unter D.I.1.b.

210 Zugleich als Anbieter des KI-System i.S.d. KI-VO.

(z.B. das jeweilige Gericht²¹¹). Soweit diese weiteren Verantwortlichen allerdings eine Eingabeaufforderung übergeben, die gerade keine Ausgabe personenbezogener Trainingsdaten erwarten lässt, wird die Bewertung für die Verarbeitung dieser sog. aufgedrängten personenbezogenen Daten²¹² anders ausfallen haben. Denn den Nutzern fehlt in diesem Fall das kognitive Element für eine Entscheidung über die Verarbeitung und die Verantwortlichkeit verbleibt deshalb bei den Ministerien.

In der zweiten Konstellation gleicht die Bewertung grundsätzlich den vorherigen Ausführungen. In dieser Konstellation liegt eine Verantwortlichkeit des Nutzers allerdings näher. Dieser Befund betrifft zumindest die aufgrund der Eingabeaufforderung erwartbare oder gezielte Extraktion von personenbezogenen Trainingsdaten durch eine entsprechende Eingabeaufforderung. Für eine erwartbare Ausgabe personenbezogener Trainingsdaten ist regelmäßig das Vorliegen einer gemeinsamen Verantwortlichkeit anzunehmen, denn es wirkt sich weiterhin die Entscheidung zur Zusammenstellung der Trainingsdaten durch die Ministerien im Zusammenspiel mit der konkreten Eingabeaufforderung aus.

Für eine gezielte, missbräuchliche Extraktion personenbezogener Trainingsdaten durch eine Eingabeaufforderung kann darüber hinaus sogar eine alleinige Verantwortlichkeit des Nutzers anzunehmen sein. Denn wenn der Nutzer auf die Extraktion konkreter Datenkategorien und Kategorien betroffener Personen unter Umgehung etwaiger Schutzmaßnahmen abzielt, die er zu eigenen Zwecken aus dem Sprachmodell extrahiert, legt der Nutzer nicht nur einen eigenen Zweck fest, sondern verengt auch die wesentlichen Mittel der Verarbeitung erheblich. Zusammengenommen ist die Verarbeitung daher nur noch auf die Entscheidung des Nutzers zurückzuführen und dient nicht (mehr) dem Zweck des Betriebs des Sprachmodells.

ee. Veröffentlichung des KI-Modells oder KI-Systems

Im Fall der Veröffentlichung des KI-Modells oder KI-Systems²¹³ verbleibt es mit Blick auf die Trainingsdaten bei einer (gemeinsamen) Verantwortlichkeit der veröffentlichenden Ministerien. Mit Blick auf die Nutzer gelten

211 *Schild*, in: BeckOK Datenschutzrecht, Syst. E. Rn. 9.

212 Ausf. zu den Lösungsansätzen in diesem Fall *L. Meyer*, RDi 2025, 125; *T. Radtke*, in: Gersdorf/Paal, Art. 6 DSA Rn. 14 f.

213 S. zu den drei verschiedenen Stufen unter D.IV.3.

die zuvor gemachten Ausführungen entsprechend. Soweit sich die Nutzer im Rahmen der Privathaushaltsausnahme bewegen (Art. 2 Abs. 2 lit. c DSGVO) oder Nutzer aus deren Sicht nicht zu erwartende Ausgaben personenbezogener Trainingsdaten als sog. aufgedrängte personenbezogene Daten erhalten, kommen diese Nutzer schon gar nicht als Verantwortliche in Betracht.

Im Fall der gezielten Extraktion personenbezogener Trainingsdaten durch entsprechende Eingabeaufforderungen kommt allerdings eine (gemeinsame) Verantwortlichkeit der Nutzer in Betracht.

4. Rechtsgrundlage nach Art. 6 DSGVO

Für die vorgenannten Verarbeitungsvorgänge bedarf es nach Art. 6 Abs. 1 DSGVO jeweils einer Rechtsgrundlage.

a. Einschränkungen aufgrund des Vorbehalts des Gesetzes (Art. 6 Abs. 1 UAbs. 2 DSGVO)

Nach Art. 6 Abs. 1 UAbs. 2 DSGVO können sich Behörden „in Erfüllung ihrer Aufgaben“ nicht auf berechtigte Interessen nach Art. 6 Abs. 1 UAbs. 1 lit. f DSGVO berufen, sondern vielmehr nur auf den Rechtfertigungskatalog des Art. 6 Abs. 1 UAbs. 1 DSGVO im Übrigen. Der europäische Gesetzgeber begründet dies damit, dass es über Art. 6 Abs. 1 lit. c, e DSGVO „dem Gesetzgeber obliegt, per Rechtsvorschrift die Rechtsgrundlage für die Verarbeitung personenbezogener Daten durch die Behörden zu schaffen“ (ErwGr. 47 S. 5 DSGVO). Die Vorschrift soll einfachgesetzlich den Vorbehalt des Gesetzes sichern²¹⁴ und versperrt insoweit auch die Berücksichtigung der behördlichen Interessen durch Dritte im Rahmen des Art. 6 Abs. 1 lit. f DSGVO²¹⁵ sowie der Interessen Dritter durch Behörden.²¹⁶

Der Begriff der öffentlichen Aufgaben ist unionsautonom auszulegen und muss sich eindeutig, insbesondere aufgrund hoheitlicher Befugnisse, von

214 Albers/Veit, in: BeckOK Datenschutzrecht, Art. 6 DSGVO Rn. 65; Schulz, in: Gola/Heckmann, Art. 6 DSGVO Rn. 60; Buchner/Petri, in: Kühling/Buchner, Art. 6 DSGVO Rn. 157; Schantz, in: Simitis/Hornung/Spiecker gen. Döhmman, Art. 6 Abs. 1 DSGVO Rn. 97.

215 Heberlein, in: Ehmann/Selmayr, Art. 6 DSGVO Rn. 38.

216 Schulz, in: Gola/Heckmann, Art. 6 DSGVO Rn. 60.

einem Verhältnis zwischen Privaten abgrenzen.²¹⁷ Dieses Verhältnis umfasst neben der Eingriffsverwaltung auch Aufgaben der Leistungsverwaltung sowie die Öffentlichkeitsarbeit,²¹⁸ nicht dagegen ein privatrechtliches Handeln der Behörde.²¹⁹ Öffentliche Stellen, die keine Behörde sind, werden von der Vorschrift nicht adressiert.²²⁰

Es ist nicht abschließend geklärt, ob auch staatliche Hochschulen²²¹ und Stellen mit judikativer Funktion²²² dem Begriff der Behörde unterfallen. Für ein weites Verständnis mit Blick auf die hier besonders relevanten judikativen Stellen lässt sich in systematischer Hinsicht anführen, dass Art. 6 Abs. 1 UAbs. 2 DSGVO solche Konstellationen erfassen soll, in denen Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO einschlägig ist; das betrifft auch die Gerichte.²²³ Allerdings differenziert die DSGVO an anderen Stellen gerade zwischen Behörden und Gerichten (ErwGr. 97 S. 1 DSGVO)²²⁴ und sieht Sonderregelungen für Gerichte vor, wie etwa mit Blick auf die Rechtsgrundlage (Art. 9 Abs. 2 lit. f DSGVO) und die Zuständigkeit der Aufsichtsbehörden (Art. 55 Abs. 3, ErwGr. 20 DSGVO).²²⁵

Diese systematischen Argumente und die auch durch die DSGVO anerkannte Unabhängigkeit der Justiz legen nahe, dass für die Kerntätigkeit der Justiz in Form der unabhängigen Rechtsprechung entweder bereits regelmäßig eine spezifische mitgliedstaatliche Rechtsgrundlage vorliegen dürfte oder sich das Gericht jedenfalls gegebenenfalls (auch) auf berechnigte Interessen nach Art. 6 Abs. 1 UAbs. 1 lit. f DSGVO stützen kann. Diese Feststellung wird allerdings gerade nur die Kerntätigkeit einer unabhängigen Justiz umfassen, sprich eine Verarbeitung im Rahmen der Rechtsprechung.²²⁶

Ein durch ein Ministerium organisiertes Projekt zur Arbeitserleichterung wirkt sich zwar auf die (vorbereitende) Rechtsprechungstätigkeit aus, be-

217 Heberlein, in: Ehmann/Selmayr, Art. 6 DSGVO Rn. 39.

218 BVerwG, BeckRS 2018, 35172 (Rn. 26); Heberlein, in: Ehmann/Selmayr, Art. 6 DSGVO Rn. 39; Schulz, in: Gola/Heckmann, Art. 6 DSGVO Rn. 60.

219 Schulz, in: Gola/Heckmann, Art. 6 DSGVO Rn. 60; einschränkend Reimer, in: Sydow/Marsch, Art. 6 DSGVO Rn. 91; Taeger, in: Taeger/Gabel, Art. 6 DSGVO Rn. 122.

220 Heberlein, in: Ehmann/Selmayr, Art. 6 DSGVO Rn. 39.

221 Bejahend Reimer, in: Sydow/Marsch, Art. 6 DSGVO Rn. 59, 90 ff. m.w.N.

222 Bejahend Reimer, in: Sydow/Marsch, Art. 6 DSGVO Rn. 90; implizit die andere Ansicht zugrunde legend Freye/Schnebbe, ZD 2020, 502 (504); ebenfalls a.A. Labusga/Petit, NJW 2022, 300 (Rn. 21 ff.).

223 Reimer, in: Sydow/Marsch, Art. 6 DSGVO Rn. 90.

224 Labusga/Petit, NJW 2022, 300 (Rn. 22).

225 Labusga/Petit, NJW 2022, 300 (Rn. 22).

226 Vgl. Labusga/Petit, NJW 2022, 300 (Rn. 23).

trifft aber nicht den Kernbereich unabhängiger Rechtsprechung in einem konkreten Verfahren – und kann daher als möglicher Bestandteil der allgemeinen öffentlichen Aufgabenerfüllung wegen Art. 6 Abs. 1 UAbs. 2 DSGVO nicht auf Art. 6 Abs. 1 UAbs. 1 lit. f DSGVO gestützt werden.

Die Veröffentlichung des trainierten Sprachmodells fällt erst recht nicht in den Kerntätigkeitsbereich einer unabhängigen Justiz, stellt aber zugleich auch noch kein privatrechtliches, den Art. 6 Abs. 1 UAbs. 2 DSGVO ausschließendes Handeln dar. Denn der Zugriff auf Trainingsdaten in der vorliegend in Rede stehenden Quantität ist gerade Ausfluss der öffentlich-rechtlichen Stellung der Justiz und der hiermit verbundenen Möglichkeit, einen Zugriff auf unveröffentlichte Gerichtsentscheidungen zu nehmen.

b. Überblick über die Rechtsgrundlagen aus Art. 6 Abs. 1 UAbs. 1 DSGVO

Für die öffentlichen Stellen scheiden im hier untersuchungsgegenständlichen Szenario neben Art. 6 Abs. 1 UAbs. 1 lit. f DSGVO (berechtigte Interessen)²²⁷ auch die Rechtsgrundlagen aus lit. a (Einwilligung)²²⁸ aus praktischen Gründen bzw. aus lit. b (Vertragserfüllung) mangels eines abgeschlossenen Vertrags mit der betroffenen Person aus. In Betracht kommen also vor allem lit. c (Erfüllung einer rechtlichen Verpflichtung) und lit. e (Erfüllung einer öffentlichen Aufgabe), die insbesondere über das Merkmal der Berechtigung bzw. Verpflichtung voneinander abzugrenzen sind.²²⁹

c. Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO (öffentliches Interesse)

Für das dieser Untersuchung zugrundeliegende GSJ-Projekt ist vor allem an die Erfüllung einer Aufgabe im öffentlichen Interesse i.S.d. lit. e zu denken, konkret einerseits betreffend die (Teil-)Anonymisierung, die auch der öffentlichen Aufgabe an der Veröffentlichung der Gerichtsentscheidungen dient, und andererseits betreffend das KI-Training und den Einsatz

227 Hierzu etwa *Paal*, ZfDR 2024, 129 (147 ff.); *Dewitte*, in: AI Meets the GDPR, S. 146 ff.; *Dieker*, ZD 2024, 132 (134 f.); *Schäfer*, ZD 2025, 12 (14); *Golland*, EuZW 2024, 846 (848 f.); *Ashkar*, ZD 2023, 523 (525 f.); *BayLDA*, Datenschutzkonforme Künstliche Intelligenz, S. 9; *Engeler/Rolfes*, ZD 2024, 423 (425 f.); *Mertens/D. Meyer*, K&R 2023, 563 (565 ff.); *Hüger*, Künstliche Intelligenz und Diskriminierung, S. 329 ff.

228 *Ashkar*, ZD 2023, 523 (525).

229 Vgl. *Albers/Veit*, in: BeckOK Datenschutzrecht, Art. 6 DSGVO Rn. 49.

des KI-Systems, die der effektiven Ausübung der Rechtsprechungstätigkeit dienen könnten.

Grundsätzlich kann es sich bei diesen Betätigungen um Aufgaben im öffentlichen Interesse²³⁰ i.S.d. lit. e (siehe auch Ziff. 1.2.1 Verwaltungsvorschrift) handeln. Die Veröffentlichungspflicht²³¹ im Hinblick auf Gerichtsentscheidungen könnte insoweit eine Pflicht zur Verarbeitung personenbezogener Daten im Wege der Anonymisierung und Veröffentlichung i.S.d. lit. c implizieren;²³² nimmt man eine solche Verarbeitungspflicht an, gelten die nachfolgend dargestellten Maßgaben entsprechend. In beiden Konstellationen sind die ergänzenden Anforderungen an die Bestimmtheit der Rechtsgrundlage aus Art. 6 Abs. 3 DSGVO zu beachten.²³³ Danach bedarf es einer Rechtsvorschrift mit festgelegtem Datenverarbeitungsbezug, die „klar und präzise und [deren] Anwendung [...] für die Rechtsunterworfenen [...] vorhersehbar“ ist (ErwGr. 41 S. 2 DSGVO).

In Anwendung des Merkmals der Erforderlichkeit und der Anforderungen an den Zusammenhang mit einer Aufgabe im öffentlichen Interesse letztlich eine Interessenabwägung durchzuführen (hierfür spricht auch die Gleichstellung von lit. e und f in Art. 21 Abs. 1 S. 1 DSGVO). Je geringer die Auswirkungen auf die betroffenen Personen sind, desto geringer fallen die Anforderungen an den Zusammenhang zur öffentlichen Aufgabe aus.

aa. (Teil-)Anonymisierung der Gerichtsentscheidungen

Im GSJ-Projekt stellt sich insoweit die Frage, ob die Anonymisierung auch dann für eine Aufgabe im öffentlichen Interesse erforderlich ist, wenn die Anonymisierung zwar auch der späteren Veröffentlichung dienen kann, aber im Zuge der Vorbereitung für ein KI-Training zur späteren Optimierung der Arbeitsabläufe der Justiz erfolgt.

An dieser Stelle gilt es, sich zunächst in Erinnerung zu rufen, dass es sich auch bei einer Anonymisierung um eine grundsätzlich rechtfertigungs-

230 *BVerfG*, NJW 1997, 2694 (2694); *OVG Lüneburg*, NJW 1996, 1489 (1489).

231 *BVerfG*, NJW 1997, 2694; *BVerwG*, NJW 1997, 2694; *BGH*, NJW 2018, 3123 (Rn. 14): „die aus dem Rechtsstaatsgebot einschließlich der Justizgewährungspflicht, dem Demokratiegebot und dem Grundsatz der Gewaltenteilung folgt“; *OLG Celle*, NSTZ 1990, 553; Überblick auch bei *Ludyga*, ZUM 2021, 887 (887 f.); *Köhnlein*, in: Graf, § 23 EGGVG Rn. 181.

232 *VG Stuttgart*, ZD 2022, 583 (Rn. 33).

233 Zum Verhältnis von Abs. 2 und 3 *Roßnagel*, in: Simitis/Hornung/Spiecker gen. Döhmman, Art. 6 Abs. 2 DSGVO Rn. 7 ff.

bedürftige Verarbeitung handelt.²³⁴ In Ansehung der datenschutzrechtlich gerade erwünschten Folgen einer Anonymisierung dürften die Anforderungen an eine solche Rechtfertigung (z.B. nach Art. 6 Abs. 1 UAbs. 1 lit. c, e, f DSGVO) sehr gering ausfallen, unabhängig davon, welcher Lösungsweg konkret gewählt wird.²³⁵

Im GSJ-Projekt ist für die Anonymisierung zunächst der konkrete Zweck der Verarbeitung zur Anonymisierung zu identifizieren, um zu prüfen, ob auch dieser Zweck der Aufgabenerfüllung i.S.d. Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO zuzuordnen ist.

Die eindeutige Zweckfestlegung nach Art. 5 Abs. 1 lit. b DSGVO erfordert eine möglichst kleinschrittige Betrachtung,²³⁶ losgelöst von Fernzielen,²³⁷ wie etwa dem späteren Einsatz als Trainingsdatum für ein Sprachmodell. Der vorliegend in Rede stehende Zweck ist vor allem die Anonymisierung als Aufbereitung der Entscheidung zur Weiterverarbeitung. Die weiteren Verarbeitungen wären – im Fall einer erfolgreichen Anonymisierung – sodann gerade keine Verarbeitungen i.S.d. DSGVO (mehr) und daher nicht zu berücksichtigen. Diese Entscheidungsaufbereitung als Voraussetzung der Weiterverarbeitung der amtlichen Entscheidungen in jeglicher Form (z.B. in der Rechtsanwendung) steht in unmittelbarem Zusammenhang mit der öffentlichen Aufgabe. Einzelne Gerichte sehen dementsprechend auch bei einer nur teilweisen Anonymisierung von Entscheidungen einen Zusammenhang mit einer öffentlichen Aufgabe.²³⁸

Eine abweichende Auffassung lässt sich darauf stützen, dass vorliegend eben gerade nicht mehr zwangsläufig die Justiz selbst, sondern gegebenenfalls die ausführenden Stellen die Anonymisierung vornehmen. Eine solche Auslagerung ist zwar im Rahmen der Auftragsverarbeitung und Zurechnung zu dem jeweiligen Verantwortlichen unschädlich, wäre aber anders zu bewerten im Fall einer gemeinsamen Verantwortlichkeit zwischen den Ministerien und den ausführenden Stellen.

234 Artikel-29-Datenschutzgruppe, WP 216, S. 7 f.; Paal, ZfDR 2024, 129 (136) m.w.N.; bejahend Burghoff, ZD 2023, 658 (660); BfDI, Positionspapier zur Anonymisierung unter der DSGVO unter besonderer Berücksichtigung der TK-Branche, S. 5 ff.; Hornung/B. Wagner, ZD 2020, 223; Roßnagel/Geminn, ZD 2021, 487; Stürmer, ZD 2020, 626 (629); ablehnend hingegen Thüsing/Rombey, ZD 2021, 548. S. auch ErwGr. 61 a.E. KI-VO.

235 Zu den Lösungswegen im Einzelnen Paal, ZfDR 2024, 129 (136) m.w.N.

236 Ausf. Jansen/T. Radtke, PinG 2024, 61 (63) m.w.N.

237 Selbst das Fernziel ist zumindest von der DSGVO anerkannt, etwa indem Art. 23 Abs. 1 lit. f DSGVO auch auf „eine ordnungsgemäße Rechtspflege abzielt“, s. EuGH, ZD 2023, 396 (Rn. 38) – Norra Stockholm Bygg AB.

238 VG Stuttgart, ZD 2022, 583.

Die teilweise Anonymisierung in Vorbereitung einer Weiterverarbeitung der Gerichtsentscheidung ist also im Rahmen des Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO grundsätzlich zulässig. Hieraus lässt sich aber noch nicht bereits ableiten, dass auch eine Veröffentlichung der mittels Erlanger Tools anonymisierten Entscheidungen mit gegebenenfalls geringerem Anonymisierungsgrad gegenüber einer händischen Anonymisierung zulässig ist.²³⁹

bb. Weiterverarbeitung teilanonymisierter Entscheidungen: Training i.w.S., Speicherung und Einsatz des Sprachmodells

Für die folgenden Schritte in Form des Trainings i.w.S., der Speicherung des Sprachmodells und dem Einsatz des auf dem Sprachmodell basierenden KI-Systems stellt sich die Frage, ob diese Verarbeitungstätigkeiten (noch) zur Wahrnehmung der öffentlichen Aufgabe(n)²⁴⁰ der Gerichte und Ministerien erforderlich sind. Insoweit ist zu untersuchen, ob die Weiterverarbeitung von teilanonymisierten Gerichtsentscheidungen zur Optimierung von Arbeitsabläufen in der Justiz (noch) den der Justiz zugewiesenen Aufgaben im öffentlichen Interesse dient.

Im Ausgangspunkt dürfte weitgehend Einigkeit darüber bestehen, dass die Veröffentlichung teilanonymisierter Gerichtsentscheidungen zur Berücksichtigung in der Fachwelt²⁴¹ ebenso wie die Vorlage von Gerichtsakten als Arbeitsnachweis in richterlichen Bewerbungsverfahren oder die Umwandlung von in Gerichtsakten enthaltenen Informationen zu Statistikzwecken öffentlichen Aufgaben der Justiz zuzuordnen ist.²⁴² Gerade die Statistikzwecke dienen zwar noch (jedenfalls teilweise) der Transparenz gegenüber der Öffentlichkeit, ermöglichen aber vor allem die Prüfung und Anpassung von Arbeitsabläufen und der Arbeitsorganisation in der Justiz.

239 S. zu den Anforderungen oben unter D.II.5.b.

240 Für die Übermittlung zwischen den Ministerien zu Zwecken der (zulässigen) Zusammenarbeit ist ebenfalls ein Aufgabenzusammenhang anzunehmen, vgl. auch Art. 5 Abs. 1 S. 1 Nr. 1 BayDSG und § 6 DSG NRW.

241 Z.B. eine zwar bei Betrachtung des Texts vollständige Anonymisierung, die aber die Herstellung eines Personenbezugs unter Zuhilfenahme von Presseberichterstattung zulässt. S. zuvor unter D.II.5.b.

242 S. bspw. die Anordnung über die Erhebung von statistischen Daten in der Verwaltungsgerichtsbarkeit (VwG-Statistik), https://media.frag-den-staat.de/files/foi/570849/VwG-Anordnung_2020.pdf: Danach sind u.a. das Aktenzeichen einer Sache, die Verfahrensart, die Art der Verfahrensbeendigung sowie Informationen zur Prozesskostenhilfe zu erheben. Diese Informationen sind (zunächst) aufgrund des Bezugs zu einer konkreten Akte personenbezogen.

Die Aufgaben der Justiz mit Datenverarbeitungsbezug im öffentlichen Interesse beschränken sich also nicht nur auf die eigentliche Rechtsprechung, sondern umfassen darüber hinaus auch Verarbeitungsvorgänge in engem Zusammenhang mit der Rechtsprechung sowie personellen und organisatorischen Angelegenheiten der Justiz. Eine Zusammenarbeit der Gerichte, gegebenenfalls unter Einbindung oder ausgehend von den jeweiligen Justizministerien, ist – wie in Statistiksachen – unschädlich. Denn ungeachtet der richterlichen Unabhängigkeit im Rahmen der Rechtsprechung i.e.S. werden sich insoweit stets organisatorische und institutionelle Anforderungen (schon insbesondere aus dem Beamtenrecht) ergeben, die zwar über das einzelne Gericht hinausreichen, aber einen hinreichenden Bezug zu öffentlichen Aufgaben aufweisen.

Die hier untersuchungsgegenständliche Verarbeitung zum Training eines Justiz-Sprachmodells als Arbeitshilfe zur Sachverhaltsdurchdringung in neuen Verfahren (siehe die Use-Cases 1-3) wird daher ebenso wie die Führung einer digitalen Akte oder einer aktenübergreifenden Statistik in einem hinreichenden Zusammenhang mit der öffentlichen Aufgabenerfüllung stehen und zur Aufgabenerfüllung erforderlich sein.²⁴³

Die Erforderlichkeit i.S.d. Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO zielt hierbei nicht bloß ab auf den möglichen Einsatz eines milderen Mittels,²⁴⁴ sondern darüber hinaus auch auf einen in Ansehung der mit der Datenverarbeitung verbundenen Risiken hinreichenden Funktionszusammenhang zwischen Verarbeitung und Aufgabenwahrnehmung. Daher wird es für das untersuchungsgegenständliche Projekt unschädlich sein, dass bereits ein geringer Anteil an Gerichtsentscheidungen in anonymisierter Form öffentlich zugänglich ist. Selbst wenn man an dieser Stelle eine Erforderlichkeit im Sinne des deutschen Verfassungsrechts verlangen würde (i.e. die Prüfung eines milderen, gleich effektiven Mittels), dürfte dem Erfordernis bei einer erheblich gesteigerten Datengrundlage (z.B. einer Vervielfachung der zur Verfügung stehenden Urteile um einen Faktor deutlich > 5) zum effektiveren Training eines Sprachmodells Genüge getan sein.

In diesem Sinne erkennt auch der europäische Gesetzgeber selbst in der KI-VO den Bedarf nach großen Datenmengen zum KI-Training an (z.B. in ErwGr. 97 S. 3, 98, 105 S. 2 KI-VO: „riesigen Mengen“ sowie vgl. Art. 3

243 Wohl krit. im Hinblick auf die Aufgabenerfüllung durch ein KI-Training, allerdings ohne Berücksichtigung der hier vorgenommenen Teilanonymisierung und weiterer Maßnahmen, *LfDIBW*, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, S. 31.

244 *Albers/Veit*, in: BeckOK Datenschutzrecht, Art. 6 DSGVO Rn. 59.

Nr. 29-32 und Art. 10 KI-VO), da sich aus der Verwendung dieser großen Datenmengen ein signifikanter Vorteil gegenüber kleineren Datenmengen ergeben kann (z.B. entsprechend der wahrheitsähnlichkeitsbasierten Ausgabe im Hinblick auf die Genauigkeit der Ausgabe, einen sog. Bias oder die Unterbindung von Ausgaben von – personenbezogenen – Daten betreffend einzelne Entscheidungen aus dem Trainingsdatensatz).²⁴⁵

Weiterhin ist zu beachten, dass die hier grundsätzlich einschlägigen Generalklauseln aus Art. 4 Abs. 1 BayDSG bzw. § 3 Abs. 1 DSG NRW nur eine eingriffsarme Verarbeitung tragen können. Die untersuchungsgegenständliche Verarbeitung umfasst nur einen Restbestand an überwiegend nicht-sensiblen personenbezogenen Daten,²⁴⁶ die innerhalb der Einflussphäre der für die Justiz Verantwortlichen verbleiben. Die Grundrechte (insbesondere Art. 8 GRCh) und Interessen betroffener Personen sind also allenfalls geringfügig betroffen und Schäden in Form finanzieller Nachteile oder drohender Diskriminierungen (vgl. ErwGr. 75 DSGVO) sind unwahrscheinlich.

Diesen Befund bestätigen weitere Umstände des Einzelfalls, die sich bereits auf Ebene der Annahme personenbezogener Daten²⁴⁷ aus- und an dieser Stelle fortwirken:²⁴⁸ Umfasst sind nach einer Bearbeitung durch das Erlanger Tool vor allem Informationen, die nicht als Kennungen unmittelbar auf eine Person hinweisen (d.h. insbesondere nur selten Namen oder Anschriften, sondern überwiegend Sachverhaltsangaben, Aktenzeichen und Daten). Diese Informationen erfordern einen erhöhten Aufwand als Ausgangspunkt für eine Identifizierung und schränken schon hierdurch die Eingriffsintensität ein.

Vor allem aber ist der Kreis der Empfänger erheblich begrenzt: Für das Training verbleiben die Daten im Einflussbereich der Ministerien unter Einschaltung der ausführenden Stellen als Auftragsverarbeiter. Erst im Fall eines Datenlecks im Rahmen der Ausgabe des eingesetzten KI-Systems würde der Empfängerkreis erweitert, indem beispielsweise Gerichte verse-

245 Vgl. die Forschung zu Duplikaten im Trainingsdatensatz *Kandpal/Wallace/Raffel*, *Deduplicating Training Data Mitigates Privacy Risks in Language Models*; sowie im Überblick zur Datenqualität *Albalak et al.*, *A Survey on Data Selection for Language Models*.

246 Vgl. auch *EDPB*, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, Rn. 73.

247 S. oben unter D.I.1.b.bb.

248 Vgl. auch *Hornung/B. Wagner*, CR 2019, 565 (Rn. 50).

hentlich und ganz vereinzelt²⁴⁹ Zugriff auf verbliebene personenbezogene Daten aus dem Trainingsdatensatz erhielten. Selbst dieser erweiterte Empfängerkreis wird in Ansehung der besonderen Stellung der Gerichte (vgl. Art. 1 Abs. 3, Art. 20 Abs. 3 GG), der Richterschaft und Beamtenschaft, des dienstlichen Kontexts und des großen Interesses an der effizienten Bearbeitung der jeweils aktuellen Akte allerdings nicht mit einer signifikanten Eingriffstiefe für betroffene Personen verbunden sein. Überdies wird dieser (justizinterne) Empfängerkreis auch im Rahmen der berechtigten Erwartungen betroffener Personen liegen (ErwGr. 47 S. 3 DSGVO).

Nach alledem wird die Verarbeitung einzelner personenbezogener Daten für das KI-Training, die Speicherung und der justizinterne Einsatz des Sprachmodells in einem KI-System – gestützt auf Art. 6 Abs. 1 UAbs. 1 lit. e, Abs. 2, 3 DSGVO i.V.m. Art. 4 Abs. 1 BayDSG bzw. § 3 Abs. 1 DSG NRW – als grundsätzlich zulässig anzusehen sein. Das gilt in Ansehung des eng begrenzten Empfängerkreises, der sehr geringen Quantität personenbezogener Daten aufgrund des Einsatzes von Anonymisierungstools und weiterer Maßnahmen sowie des engen Zusammenhangs mit dem Einsatz innerhalb der Justiz zur Unterstützung bei Arbeitsabläufen auf Basis der Erkenntnisse aus vergangenen Gerichtsverfahren.

cc. Veröffentlichung des Sprachmodells

Die Veröffentlichung des Sprachmodells als KI-System, des eigentlichen Modells mitsamt der Gewichte und der vollständigen Trainingsdaten unterfällt jeweils nicht mehr der Aufgabenerfüllung i.S.d. Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO.

Die Veröffentlichung, gleich in welcher Form, dient zwar als Arbeitshilfe für die Praxis der Rechtsanwendung, aber auch der Transparenz der Arbeitsweise der Justiz sowie gegebenenfalls wissenschaftlichen Forschungszwecken.

Die Transparenz öffentlichen Handelns ist beispielsweise in den Informationsfreiheits- und Transparenzgesetzen der Länder als legitimes Ziel angelegt. Wissenschaftliche Forschungszwecke sind an verschiedenen Stel-

249 Das erfordert erneut entsprechende Maßnahmen auf Ausgabebene (z.B. Ausgabe-
filter).

len in der DSGVO privilegiert.²⁵⁰ Es müssten aber gerade diese Zwecke im Zusammenhang mit der Erfüllung einer Aufgabe der Justiz bzw. Ministerien im öffentlichen Interesse stehen, für die die Verarbeitung personenbezogener Daten erforderlich ist. Unabhängig davon, ob überhaupt noch ein Aufgabenbezug angenommen werden kann, dürfte jedenfalls die maßgebliche Erforderlichkeit ausscheiden. Denn die personenbezogenen Daten werden gegebenenfalls einem unbeschränkten Empfängerkreis zugänglich gemacht, wovon eine erhebliche Erhöhung des Risikos für die betroffenen Personen ausgeht. Nur in der niedrigsten Veröffentlichungsstufe, sprich dem bloß eingeschränkten Zugriff der Öffentlichkeit auf ein KI-System, können diese Auswirkungen zwar erheblich begrenzt werden, räumen aber gleichwohl nicht die Zweifel im Hinblick auf den Zusammenhang zur öffentlichen Aufgabenerfüllung aus.

Je näher das Erlanger Tool oder ein anderes Instrument zur automatisierten Anonymisierung an einen Anonymisierungsstandard heranreicht, der auch bei sorgfältiger händischer Anonymisierung erzielt worden wäre,²⁵¹ desto eher wird tendenziell auch die Zulässigkeit der Veröffentlichung des Sprachmodells hinzunehmen sein. Denn wenn die Veröffentlichung der – gegebenenfalls nur weitgehend – anonymisierten Gerichtsentscheidung zulässig wäre, birgt die Zusammenführung dieser ohnehin veröffentlichungsfähigen Gerichtsentscheidungen im Rahmen des KI-Trainings kaum (mehr) erhöhte Risiken für betroffene Personen und kann daher zugleich einen geringen Zusammenhang mit der öffentlichen Aufgabenerfüllung ausreichen lassen.

d. Art. 6 Abs. 1 UAbs. 1 lit. c DSGVO (rechtliche Pflicht)

Wie gezeigt werden konnte, dürfte für das untersuchungsgegenständliche Projekt regelmäßige Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO anstelle von lit. c in Betracht kommen.²⁵² Falls im Einzelfall eine vollständige Anonymisierung bewirkt werden können sollte, kann die Anonymisierung einer Löschung

250 S. z.B. Art. 5 Abs. 1 lit. b, e, Art. 9 Abs. 2 lit. j, Art. 14 Abs. 5 lit. b, Art. 17 Abs. 3 lit. b, Art. 21 Abs. 6, Art. 85, 89 DSGVO sowie ErwGr. 33, 50, 52, 53, 62, 65, 133, 156 ff. DSGVO.

251 S. oben unter D.II.5.b.

252 S. oben unter D.V.4.c.

gleichstehen und auf Art. 6 Abs. 1 UAbs. 1 lit. c i.V.m. Art. 17 DSGVO gestützt werden.²⁵³

e. Zweckänderung (Art. 5 Abs. 1 lit. b, Art. 6 Abs. 4 DSGVO)

Weiterhin – oder nach einer Ansicht alternativ zu den Voraussetzungen aus Art. 6 Abs. 1 DSGVO – sind als Ausprägung aus dem in Art. 5 Abs. 1 lit. b DSGVO niedergelegten Zweckbindungsgrundsatz die Anforderungen an eine Zweckänderung aus Art. 6 Abs. 4 DSGVO²⁵⁴ zu beachten, wenn für die weitere Verarbeitung von dem Erhebungszweck abgewichen wird. Das betrifft die untersuchungsgegenständliche Konstellation der Weiterverarbeitung personenbezogener Daten durch das Ministerium zu Zwecken des KI-Trainings.

aa. Verhältnis zu Art. 6 Abs. 1 DSGVO

Es sprechen durchaus beachtliche Argumente dafür, Art. 6 Abs. 4 DSGVO als eine Spezialvorschrift zu Art. 6 Abs. 1 DSGVO anzusehen. In diesem Fall würde die zuvor anhand des Art. 6 Abs. 1 DSGVO durchgeführte Prüfung vollständig durch den nach Art. 6 Abs. 4 DSGVO erforderlichen Kompatibilitätstest verdrängt (siehe hierzu nachfolgend unter D.V.4.e.bb(2)). Die Rechtfertigung einer Verarbeitung im GSJ-Projekt wäre mithin an einem weniger strengen Maßstab zu messen. In Literatur²⁵⁵ und Rechtsprechung ist allerdings (noch) nicht abschließend geklärt, ob ein solches Spezialitäts-

253 Stürmer, ZD 2020, 626 (630).

254 Wenn und soweit man die landesrechtlichen Vorschriften Art. 5 Abs. 1 S. 1 Nr. 2 BayDSG bzw. § 8 Abs. 2 DSG NRW als geeignete Rechtsgrundlage entgegen BVerwG, BeckRS 2018, 35172 (Rn. 25 f.) ansieht, treten diese – i.E. ohne Auswirkung – an die Stelle des Art. 6 Abs. 4 DSGVO.

255 Für die Notwendigkeit einer separaten Rechtsgrundlage Albers/Veit, in: BeckOK Datenschutzrecht, Art. 6 DSGVO Rn. 108; und die Rechtsprechung des EuGH hierzu für „unmissverständlich“ haltend Buchner/Petri, in: Kühling/Buchner, Art. 6 DSGVO Rn. 184; Heberlein, in: Ehmann/Selmayr, Art. 6 DSGVO Rn. 26; DSK, Erfahrungsbericht der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder zur Anwendung der DS-GVO, S. 13 f.; Albrecht, CR 2016, 88 (92); Schantz, NJW 2016, 1841 (1844); Pauly/Nabulsi, ZD 2023, 519 (522); a.A. Roßnagel, in: Simitis/Hornung/Spiecker gen. Döhmann, Art. 6 Abs. 4 DSGVO Rn. 12; Schulz, in: Gola/Heckmann, Art. 6 DSGVO Rn. 142; Taeger, in: Taeger/Gabel, Art. 6 DSGVO Rn. 173; Culik/Döpke, ZD 2017, 226 (230); Stürmer, ZD 2020, 626

verhältnis anzunehmen ist oder vielmehr die Anforderungen aus Art. 6 Abs. 4 DSGVO *zusätzlich* zu den Anforderungen aus Art. 6 Abs. 1 DSGVO zu beachten sind, also insgesamt höhere Anforderungen an die zweckändernde Weiterverarbeitung zu stellen sind.

(1) Art. 6 Abs. 4 DSGVO als ergänzende Regelung

Die ein Spezialitätsverhältnis ablehnenden Stimmen verweisen darauf, Art. 6 Abs. 4 DSGVO sei gerade nicht als ein gesonderter Rechtfertigungstatbestand konzipiert, weshalb die Fragen nach der Rechtsgrundlage und der Zweckbindung wie schon unter der DSRL voneinander zu trennen seien.²⁵⁶ Ferner ergebe sich der Befund einer stets notwendigen Rechtsgrundlage auch aus Art. 8 Abs. 2 GRCh, dessen Erfordernisse der europäische Gesetzgeber in Art. 6 Abs. 4 DSGVO achten wollte.²⁵⁷ Zudem wird eine dann möglicherweise konsequente Anwendung des Art. 6 Abs. 4 DSGVO auch auf besondere Kategorien personenbezogener Daten i.S.d. Art. 9 DSGVO kritisiert.²⁵⁸ In historischer Hinsicht wird darauf hingewiesen, dass ein Vorschlag, in Art. 6 Abs. 4 DSGVO auf einzelne Rechtsgrundlagen zu verweisen („Ist der Zweck der Weiterverarbeitung mit dem Zweck, für den die personenbezogenen Daten erhoben wurden, nicht vereinbar, muss auf die Verarbeitung mindestens einer der in Absatz 1 Buchstaben a bis e genannten Gründe zutreffen“), durch das Europäische Parlament im Gesetzgebungsverfahren zurückgewiesen wurde.²⁵⁹

(2) Art. 6 Abs. 4 DSGVO als eigenständige Rechtsgrundlage

Allerdings indiziert die Formulierung in Art. 83 Abs. 5 lit. a DSGVO („Bei Verstößen gegen die folgenden Bestimmungen ...: die Grundsätze für die Verarbeitung, einschließlich der Bedingungen für die Einwilligung, gemäß

(629 f.); P. Richter, DuD 2015, 735 (736); Überblick zum Streitstand auch bei Ziegenhorn/Schulz-Große, ZD 2023, 581 (586) m.w.N.

256 Buchner/Petri, in: Kühling/Buchner, Art. 6 DSGVO Rn. 184.

257 Buchner/Petri, in: Kühling/Buchner, Art. 6 DSGVO Rn. 184; Artikel-29-Datenschutzgruppe, WP 203, S. 12.

258 Pauly/Nabulsi, ZD 2023, 519 (522).

259 Pauly/Nabulsi, ZD 2023, 519 (522); Albrecht, CR 2016, 88 (92). Vgl. auch die DSGVO in der Entwurfsfassung 2017/C 378/55, abrufbar unter: <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=OJ:C:2017:378:FULL&from=DE>.

den Artikeln 5, 6, 7 und 9;“), dass die jeweiligen Absätze des Art. 6 DSGVO zusammen zu betrachten sind, sprich Art. 6 Abs. 4 DSGVO das Erfordernis einer Rechtsgrundlage nach Art. 6 Abs. 1 DSGVO gegebenenfalls verdrängen kann.

Das Erfordernis eines zusätzlichen Erlaubnistatbestands würde außerdem geringfügige Zweckänderungen durch einen Verantwortlichen benachteiligen (z.B. die justizinterne Weiterverarbeitung von personenbezogenen Daten in Gerichtsentscheidungen zu Dokumentations- oder Statistikzwecken), da für diese Verarbeitung sowohl die Voraussetzungen aus Art. 6 Abs. 4 DSGVO als auch aus Art. 6 Abs. 1 DSGVO mit Blick auf die Weitervereinbarung einzuhalten wären.²⁶⁰ Eine Neuerhebung durch einen anderen Verantwortlichen zu einem gänzlich anderen Zweck bedürfte demgegenüber bloß einer Rechtsgrundlage nach Art. 6 Abs. 1 DSGVO. Diese Benachteiligung geringfügiger Zweckänderungen durch einen Verantwortlichen könnte aber aufgrund der eigenständigen Bedeutung der Zweckbindung (Art. 5 Abs. 1 lit. DSGVO) gerechtfertigt sein. Insoweit wäre die Erwartung betroffener Personen geschützt, die weitere Verarbeitung durch einen Verantwortlichen erfolge entsprechend dem ursprünglichen Erhebungszweck.

Hervorzuheben ist allerdings zugunsten eines Spezialitätsverhältnisses, dass ErwGr. 50 S. 2 DSGVO im Einklang mit der Systematik des Art. 6 DSGVO eindeutig formuliert, dass in diesen Fällen „keine andere gesonderte Rechtsgrundlage [...] als diejenige für die Erhebung der personenbezogenen Daten“ erforderlich ist.²⁶¹ Daran vermag auch der *allgemeine* Verweis auf die Einhaltung der Datenschutzgrundsätze in ErwGr. 50 S. 8 DSGVO nichts zu ändern.²⁶² Der Gesetzgeber hat sich zudem gerade gegen einen Verweis auf Art. 6 Abs. 1 DSGVO für die Weiterverarbeitung und damit auch gegen das Erfordernis eines gesonderten Rechtmäßigkeitstatbestands entschieden. Andernfalls, sprich wenn gegebenenfalls erneut eine (zusätzliche) Einwilligung nach Art. 6 Abs. 1 lit. a DSGVO zu prüfen wäre, wäre zudem die gesonderte Nennung der Einwilligung und der Rechtsvorschriften in Art. 6 Abs. 4 Hs. 1 DSGVO anstelle des Kompatibilitätstests nach Art. 6 Abs. 4 Hs. 2 DSGVO nicht erforderlich und liefe faktisch weitgehend leer.

260 Roßnagel, in: Simitis/Hornung/Spiecker gen. Döhmman, Art. 6 Abs. 4 DSGVO Rn. 12.

261 Culik/Döpke, ZD 2017, 226 (230); P. Richter, DuD 2015, 735 (736).

262 So aber etwa DSK, Erfahrungsbericht der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder zur Anwendung der DS-GVO, S. 13.

(3) Berücksichtigung der EuGH-Rechtsprechung

Der EuGH hat sich bisher, soweit ersichtlich, noch nicht eindeutig positioniert, scheint aber zu einem Spezialitätsverhältnis zu tendieren: Soweit Stimmen in der Literatur²⁶³ auf die EuGH-Entscheidung in der Rs. Bara verweisen, stellt der EuGH dort im Kontext einer Übermittlung bloß fest: „Gemäß den Bestimmungen des Kap. II („Allgemeine Bedingungen für die Rechtmäßigkeit der Verarbeitung personenbezogener Daten“) der RL 95/46 muss jede Verarbeitung personenbezogener Daten [...] den in Art. 6 der RL [i.e. Art. 5 DSGVO] aufgestellten Grundsätzen in Bezug auf die Qualität der Daten und einem der in Art. 7 der RL [i.e. Art. 6 DSGVO] angeführten Grundsätze in Bezug auf die Zulässigkeit der Verarbeitung von Daten genügen“.²⁶⁴ Damit ist allerdings gerade nichts zu dem Verhältnis von Art. 6 Abs. 1 DSGVO und Art. 6 Abs. 4 DSGVO (i.V.m. Art. 5 Abs. 1 lit. b DSGVO) gesagt: Einerseits liegt im Fall der Übermittlung an einen neuen Verantwortlichen ohnehin eine neue Erhebung i.S.d. Art. 6 Abs. 4 vor, so dass in jedem Fall Art. 6 Abs. 1 DSGVO für den neuen Verantwortlichen zur Anwendung berufen wäre.²⁶⁵ Andererseits verhält sich der EuGH an der vorbenannten Stelle gerade nicht ausdrücklich zu Art. 6 Abs. 4 DSGVO, der in dieser Form als Ausprägung des Zweckbindungsgrundsatzes in die DSGVO aufgenommen wurde.

Die vorzugswürdige Ansicht, die ein Spezialitätsverhältnis annimmt, dürfte mittlerweile jedenfalls eine Stütze in der EuGH-Rechtsprechung finden.²⁶⁶ In einer Entscheidung aus dem Jahr 2022 befasste sich der EuGH mit der Frage, ob die separate Speicherung zu Testzwecken mit dem Grundsatz der Zweckbindung (Art. 5 Abs. 1 lit. b DSGVO) vereinbar ist.²⁶⁷ In diesem Zusammenhang ging der EuGH nur auf die Rechtmäßigkeit nach Art. 6 Abs. 1 DSGVO für die Erhebung ein²⁶⁸ und erwähnte mit keinem Wort das Erfordernis einer Rechtsgrundlage aus Art. 6 Abs. 1 DSGVO für die zweckändernde Weiterverarbeitung – wie es sich bei einem anderen Normverständnis ansonsten aufgedrängt hätte –, weshalb der Gerichtshof schließlich zu dem Ergebnis kommt, „dass Art. 5 Abs. 1 Buchst. b der Verordnung 2016/679 dahin auszulegen ist, dass der darin vorgesehene

263 Buchner/Petri, in: Kühling/Buchner, Art. 6 DSGVO Rn. 184.

264 EuGH, ZD 2015, 577 (Rn. 30) – Bara.

265 So auch P. Richter, DuD 2015, 735 (736).

266 S. auch EuGH, ZD 2023, 396 (Rn. 43, 46 ff.) – Norra Stockholm Bygg AB.

267 EuGH, ZD 2023, 31.

268 EuGH, ZD 2023, 31 (Rn. 28).

Grundsatz der ‚Zweckbindung‘ es dem Verantwortlichen *nicht verwehrt*, in einer zu Testzwecken und zur Behebung von Fehlern eingerichteten Datenbank personenbezogene Daten zu erfassen und zu speichern, die zuvor erhoben und in einer anderen Datenbank gespeichert wurden, *wenn* diese Weiterverarbeitung mit den konkreten Zwecken *vereinbar ist*, für die die personenbezogenen Daten ursprünglich erhoben wurden, was anhand der in Art. 6 Abs. 4 dieser Verordnung genannten Kriterien und sämtlicher Umstände des Einzelfalls zu beurteilen ist“²⁶⁹ (Hervorhebung d. d. Verf.).

(4) Zwischenergebnis

Insgesamt sprechen also durchaus gewichtige Argumente für ein Spezialitätsverhältnis, sodass die Weiterverarbeitung durch die Ministerien bzw. ihre Auftragsverarbeiter ausschließlich an Art. 6 Abs. 4 DSGVO (und nicht an Art. 6 Abs. 1 DSGVO) zu messen wäre. Allerdings besteht ein nicht unerhebliches Risiko, dass die Rechtsprechung sich in der Zukunft gegen ein solches Spezialitätsverhältnis ausspricht. Eine Verarbeitung im GSJ-Projekt sollte daher nicht nur auf Art. 6 Abs. 4 DSGVO gestützt werden, sondern darüber hinaus die Anforderungen aus Art. 6 Abs. 1 DSGVO beachten.

bb. Kompatibilität im Einzelfall

Entweder als maßgebliche Zulässigkeitsvoraussetzung oder ergänzend zu Art. 6 Abs. 1 DSGVO für den Fall der Zweckänderung durch einen Verantwortlichen, wie im untersuchungsgegenständlichen GSJ-Projekt, sind stets die Kompatibilitätsanforderungen aus Art. 6 Abs. 4 DSGVO zu beachten. Für die Zulässigkeit nach Art. 6 Abs. 4 DSGVO kommt entweder (Hs. 1) eine Rechtsvorschrift in Betracht, „die in einer demokratischen Gesellschaft eine notwendige und verhältnismäßige Maßnahme zum Schutz der in Artikel 23 Absatz 1 genannten Ziele darstellt“ oder (Hs. 2) die Durchführung des Kompatibilitätstests anhand der in lit. a-e nicht abschließend aufgeführten Kriterien.

269 EuGH, ZD 2023, 31 (Rn. 45).

(1) Rechtsvorschrift (Hs. 1)

Die Anforderungen an das Vorliegen einer Rechtsvorschrift i.S.d. Art. 6 Abs. 4 DSGVO sind im Einzelnen umstritten.²⁷⁰ Mit dem EuGH ist allerdings davon auszugehen, dass Art. 6 Abs. 4 DSGVO über Art. 6 Abs. 1 UAbs. 1 lit. c, e DSGVO hinausgeht, indem die Rechtsvorschrift die Zweckänderung berücksichtigen muss und den in Art. 23 Abs. 1 DSGVO genannten Zielen entsprechen muss.²⁷¹ Art. 23 Abs. 1 lit. f DSGVO nennt unter anderem „den Schutz der Unabhängigkeit der Justiz und den Schutz von Gerichtsverfahren“, der die Justiz vor „internen oder externen Eingriffen [schützt], aber auch auf eine ordnungsgemäße Rechtspflege abzielt“.²⁷²

Es bestehen Zweifel, ob die landesrechtlichen Generalklauseln i.V.m. Aufgabenzuweisungen an die Rechtspflege²⁷³ hinreichend konkret i.S.d. Art. 6 Abs. 4 DSGVO eine Zweckänderung zur Optimierung der Arbeitsabläufe in der Rechtspflege zulassen. Im Ergebnis wäre das vorliegend insbesondere dann zu bejahen, wenn die Auswirkungen auf die betroffenen Personen sich in Grenzen hielten und die Zwecke eng miteinander verknüpft wären. Eine Aussage über diesen Umfang der Auswirkungen und die Zwecknähe enthält bereits der Art. 6 Abs. 4 Hs. 2 DSGVO.

Im Ergebnis ergeben sich somit für die Frage der Reichweite der Rechtsvorschrift i.S.d. Art. 6 Abs. 4 Hs. 1 DSGVO ähnliche Erwägungen wie innerhalb des Kompatibilitätstests nach Art. 6 Abs. 4 Hs. 2 DSGVO. Vor diesem Hintergrund soll nachfolgend dieser Kompatibilitätstest durchgeführt werden.

(2) Kompatibilitätstest (Hs. 2)

Der Kompatibilitätstest als eine der beiden alternativen Zulässigkeitsvoraussetzungen nach Art. 6 Abs. 4 DSGVO erfordert eine normative²⁷⁴ Bewertung der Vereinbarkeit des ursprünglichen und des neuen Zwecks unter Berücksichtigung der in lit. a-e genannten Maßgaben sowie der berechtigten Erwartungen betroffener Personen (ErwGr. 50 S. 6 DSGVO). Die

270 Überblick bei *Albers/Veit*, in: BeckOK Datenschutzrecht, Art. 6 DSGVO Rn. 112.

271 *EuGH*, ZD 2023, 396 (Rn. 34 ff.) – Norra Stockholm Bygg AB.

272 *EuGH*, ZD 2023, 396 (Rn. 38) – Norra Stockholm Bygg AB.

273 S. oben unter DV.4.c.

274 So auch *Ziegenhorn/Schulz-Große*, ZD 2023, 581 (585); zu den Kriterien auch *Ashkar*, ZD 2023, 523 (526 f.).

Weiterverarbeitung zur Teilanonymisierung und zum KI-Training ist daher insbesondere, jedoch nicht nur an den Kriterien der lit. a-e zu prüfen.

Die Verbindung der Zwecke (lit. a) spricht im Hinblick auf den Einsatz in der Justiz und im Zusammenhang mit (anderen) Gerichtsverfahren zwar zunächst für die Kompatibilität. Zugleich ist aber der ursprüngliche Zweck auf ein Gerichtsverfahren, der neue Zweck auf sämtliche zukünftige Gerichtsverfahren und allgemeine Arbeitsabläufe gerichtet.

Im Rahmen des Erhebungszusammenhangs und des Verhältnisses zwischen Verantwortlichem und betroffener Person (lit. b) sind insbesondere die berechtigten Erwartungen betroffener Personen aus diesem Verhältnis heranzuziehen. Abhängig von der Erhebung im Einzelfall und den in Rede stehenden Kategorien betroffener Personen (z.B. Partei, Zeuge oder – nur – im Rahmen eines Schriftsatzes genannte Person) besteht im Regelfall eine Erwartung der Verarbeitung für justizinterne Zwecke, im Einzelfall eine nur geringfügige Erwartung bzw. ein Wissen, überhaupt im Rahmen des Gerichtsverfahrens erwähnt worden zu sein. Für die erstgenannten Konstellationen wird die Erwartung betroffener Personen die Verarbeitung durch die Richterschaft und die Geschäftsstellen sowie gegebenenfalls auch das Justizministerium umfassen, wenn und soweit ein enger Zusammenhang mit der Rechtspflege besteht und keine Weitergabe an Dritte erfolgt (z.B. zu Statistikzwecken). Die untersuchungsgegenständliche Zweckänderung zu Zwecken der verfahrensübergreifenden Optimierung von Arbeitsabläufen durch Anonymisierung, ein KI-Training und den KI-Einsatz dürfte daher vielfach im Rahmen des Erwartbaren liegen; dies gilt unabhängig davon, ob der Einsatz für die konkrete Software (in Form einer KI-Anwendung) erwartet wurde.

Die Art der personenbezogenen Daten (lit. c) hängt insbesondere davon ab, ob besondere Kategorien personenbezogener Daten i.S.d. Art. 9 DSGVO verarbeitet werden. Sofern eine mit Blick auf Art. 9 DSGVO sinnvolle Clusterung der Urteile erfolgt,²⁷⁵ wird auch an dieser Stelle davon auszugehen sein, dass regelmäßig keine besonderen Kategorien personenbezogener Daten verarbeitet werden. Die Sensibilität der verarbeiteten Daten wird dann im Übrigen als eher gering einzuschätzen sein, zumal die Informationen unter Umständen zumeist bereits aus anderen Quellen bekannt sein können (z.B. aus einer medialen Berichterstattung).

Die Folgen der beabsichtigten Weiterverarbeitung (lit. d) weisen Überschneidungen auf mit dem Vorhandensein geeigneter Garantien (lit. e).

275 S. hierzu nachfolgend unter D.V.5.

Durch die anzuwendenden Garantien in Form der Teilanonymisierung, der weiteren Gewährleistung der Zweckbindung durch technische und organisatorische Maßnahmen sowie einen auf die Justiz beschränkten Empfängerkreis sind keine schwerwiegenden Folgen für betroffene Personen zu erwarten, soweit überhaupt (vereinzelte) personenbezogene Daten vorliegen.

Über die genannten Merkmale hinaus können auch die bereits im Rahmen von Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO genannten Abwägungsaspekte fruchtbar gemacht werden.²⁷⁶ Diese Abwägungsaspekte stützen ebenfalls den Befund der Kompatibilität des neuen Zwecks mit dem ursprünglichen Zweck der Verarbeitung.

cc. Schlussfolgerungen für das untersuchungsgegenständliche Projekt

Die Anforderungen des Art. 6 Abs. 4 Hs. 2 DSGVO sind also bei Vorliegen hinreichender Garantien (siehe nachfolgend unter D.VIII) und der sorgfältigen Clusterung mit Blick auf besondere Kategorien personenbezogener Daten grundsätzlich gewahrt. In Ansehung der im Rahmen des Kompatibilitätstests ermittelten Zwecknähe erscheint alternativ auch eine Rechtfertigung nach Art. 6 Abs. 4 Hs. 1 DSGVO i.V.m. landesrechtlichen Vorschriften²⁷⁷ denkbar.

Die Weiterverarbeitung im untersuchungsgegenständlichen Projekt zu KI-Trainingszwecken und dem späteren Einsatz des Sprachmodells in der Justiz kann also entweder – nach bestrittener Auffassung – bereits auf Art. 6 Abs. 4 DSGVO gestützt werden oder aber die Verarbeitung erfüllt zumindest die zusätzlichen Voraussetzungen aus Art. 6 Abs. 4 DSGVO und ist zulässig, soweit jeweils auch eine Rechtsgrundlage aus Art. 6 Abs. 1 DSGVO einschlägig ist.

5. Rechtsgrundlagen für besondere Datenkategorien nach Art. 9, 10 DSGVO

Art. 9 DSGVO sieht das Erfordernis einer gesonderten Rechtsgrundlage vor für die „Verarbeitung personenbezogener Daten, aus denen die rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschau-

²⁷⁶ S. oben unter DV.4.c.bb.

²⁷⁷ S. oben unter DV.4.c.

liche Überzeugungen oder die Gewerkschaftszugehörigkeit hervorgehen, sowie [für] die Verarbeitung von genetischen Daten, biometrischen Daten zur eindeutigen Identifizierung einer natürlichen Person, Gesundheitsdaten oder Daten zum Sexualleben oder der sexuellen Orientierung einer natürlichen Person“. Art. 10 DSGVO modifiziert diese Anforderungen mit Blick auf Daten über strafrechtliche Verurteilungen und Straftaten.

Art. 9, 10 DSGVO sehen für eine Abwägung, wie etwa unter Art. 6 Abs. 1 UAbs. 1 lit. f DSGVO, nur einen sehr eingeschränkten Spielraum vor. Vor diesem Hintergrund werden für eingriffsarme Verarbeitungsvorgänge (z.B. eine Anonymisierung) eine teleologische Reduktion des Art. 9 Abs. 1 DSGVO²⁷⁸ bzw. vergleichbare Lösungsansätze für zufällig verarbeitete sensible Daten vorgeschlagen.²⁷⁹ Solche teleologisch orientierten Überlegungen erhalten Aufwind durch ein weites Verständnis des Begriffs der Gesundheitsdaten durch den EuGH, der bereits Daten über eine Bestellung nicht verschreibungspflichtiger Medikamente als Gesundheitsdaten einordnet.²⁸⁰

a. Relevanz für das untersuchungsgegenständliche Projekt

Für das GSJ-Projekt ist mit Blick auf nach einer Teilanonymisierung gegebenenfalls weiterhin vorhandene personenbezogene Daten zu fragen, ob eine Rechtsgrundlage aus dem Katalog des Art. 9 Abs. 2 DSGVO einschlägig ist. In Ansehung des Ausschlusses von Gerichtsentscheidungen und Aktenauszügen in Strafsachen (vgl. Art. 10 DSGVO) sowie insbesondere auch in Arzthaftungssachen (i.e. betreffend Gesundheitsdaten i.S.d. Art. 9 Abs. 1 DSGVO) und Familiensachen ist bereits in konzeptioneller Hinsicht die Wahrscheinlichkeit des Vorliegens besonderer Kategorien personenbezogener Daten erheblich reduziert.

Weiterhin in Betracht kommen allerdings gleichwohl das Vorhandensein von vereinzelt enthaltenen Gesundheitsdaten und Daten zum Sexualleben oder der sexuellen Orientierung (z.B. im Hinblick auf deliktsrechtliche Anspruchsgrundlagen, gegebenenfalls i.V.m. §§ 823 ff. BGB oder im Rahmen des Schadensersatzumfangs nach §§ 249 ff. BGB, insbesondere mit Blick auf ein Schmerzensgeld, § 253 Abs. 1, 2 BGB) sowie von Daten über politische Meinungen oder Gewerkschaftszugehörigkeiten (z.B. im Arbeitsrecht mit

278 Stürmer, ZD 2020, 626 (630); Hornung/B. Wagner, ZD 2020, 223 (228).

279 Etwa Britz/Indenhuck/Langerhans, ZD 2021, 559; Kohn/Schleper, ZD 2023, 723.

280 EuGH, NJW 2025, 33 (Rn. 88).

Gewerkschaftsbezug sowie geltend gemachte Unterlassungsansprüche im Deliktsrecht im Hinblick auf getätigte Aussagen). Darüber hinaus können auch und gerade im Zusammenhang mit deliktsrechtlichen Ansprüchen vereinzelt Verweise auf Strafverfahren und strafrechtliche Verurteilungen enthalten sein, die Art. 10 DSGVO unterfallen. Insoweit bietet sich eine weitergehende Clusterung an, wie sie im Nachfolgenden vorgeschlagen wird.

b. Personenbezogene Daten über politische Meinungen, Gesundheit und sexuelle Orientierung (Art. 9 DSGVO)

Zunächst sind die besonderen Kategorien personenbezogener Daten i.S.d. Art. 9 DSGVO in den Blick zu nehmen.

aa. Art. 9 Abs. 2 lit. e DSGVO (offensichtliche Öffentlichmachung)

Eine Verarbeitung ist nach Art. 9 Abs. 2 lit. e DSGVO zulässig, wenn sie sich auf personenbezogene Daten bezieht, „die die betroffene Person offensichtlich öffentlich gemacht hat“. Eine solche „offensichtliche Öffentlichmachung“ setzt die Bereitstellung der Daten an einen unbestimmten Personenkreis voraus, wofür die Unterbreitung von Informationen an die Gerichtsöffentlichkeit bzw. oftmals nur Schriftsatzöffentlichkeit als bestimmten Personenkreis nicht genügen wird. Ferner dürfte vorliegend die entsprechend dem Zweck der Regelung vorausgesetzte freiwillige²⁸¹ Öffentlichmachung bei Informationen im Rahmen eines Gerichtsverfahrens zumindest zweifelhaft sein.

Der EuGH berücksichtigt insoweit zudem einschränkend den Kontext der Öffentlichmachung, z.B. im Fall der im Rahmen einer Podiumsdiskussion öffentlich gemachten sexuellen Orientierung mit Blick auf die Verarbeitung der Daten über die sexuelle Orientierung im Kontext eines sozialen Netzwerks.²⁸²

Vor diesem Hintergrund werden im GSJ-Projekt verarbeitete besondere Kategorien personenbezogener Daten regelmäßig nicht bereits deswegen

281 Am Beispiel der Impressumspflicht DSK, Orientierungshilfe der Aufsichtsbehörden zur Verarbeitung von personenbezogenen Daten für Zwecke der Direktwerbung unter Geltung der Datenschutz-Grundverordnung, S. 15; Reichert/K. Radtke/Eske, ZD 2024, 483 (485).

282 EuGH, NJW 2025, 207 (Rn. 76 ff.).

als öffentlich gemacht anzusehen sein, weil sie der Gerichtsöffentlichkeit präsentiert wurden. Im Einzelfall, ohne belastbar alle in dem GSJ-Projekt in Betracht kommenden Konstellationen erfassen zu können, ist allerdings denkbar, dass die Information bereits auf anderem Wege (z.B. medial oder über das Internet) öffentlich gemacht wurde und daher Art. 9 Abs. 2 lit. e DSGVO zur Anwendung gelangt.

bb. Art. 9 Abs. 2 lit. f DSGVO (Rechtsverteidigung und justizielle Tätigkeit der Gerichte)

Nach Art. 9 Abs. 2 lit. f DSGVO ist eine Verarbeitung besonderer Kategorien personenbezogener Daten zulässig, soweit diese erforderlich ist „zur Geltendmachung, Ausübung oder Verteidigung von Rechtsansprüchen oder bei Handlungen der Gerichte im Rahmen ihrer justiziellen Tätigkeit“. Die Vorschrift stellt eine spezielle Ausprägung der in Art. 23 Abs. 1 lit. f DSGVO niedergelegten und weitergefassten Öffnungsklausel zum „Schutz der Unabhängigkeit der Justiz und den Schutz von Gerichtsverfahren“ dar. Art. 9 Abs. 2 lit. f DSGVO hat vor allem Bedeutung mit Blick auf die Tatbestandsalternative der Rechtsverfolgung und den individuellen Justizgewährleistungsanspruch.²⁸³ Das Merkmal der justiziellen Tätigkeit knüpft an diese Tatbestandsalternative an und stellt sicher, dass die Justiz die für die Entscheidungsfindung im Einzelfall notwendigen Informationen verarbeiten kann.²⁸⁴

Die Verwaltungstätigkeit der Gerichte ist nach dem Wortlaut des Art. 9 Abs. 2 lit. f DSGVO („justizielle Tätigkeit“) allerdings gerade nicht adressiert,²⁸⁵ mithin auch nicht die Organisation, Entwicklung und Beschaffung von Hilfsmitteln für die justizielle Tätigkeit. Wegen der erhöhten Schutzbedürftigkeit der betroffenen Personen, wie sie in Art. 9 Abs. 1 DSGVO zum Ausdruck kommt, kann das untersuchungsgegenständliche KI-Training somit nicht auf diesen Ausnahmetatbestand gestützt werden.

Etwas anderes mag allerdings gelten, soweit der Anonymisierungsstandard einer zulässigen Veröffentlichung erreicht wird:²⁸⁶ In diesem Fall stellt sich die durch jedermann mögliche Weiterverarbeitung als eingriffsintensiv

283 Weichert, in: Kühling/Buchner, Art. 9 DSGVO Rn. 83; Kampert, in: Sydow/Marsch, Art. 9 DSGVO Rn. 33.

284 Mester, in: Taeger/Gabel, Art. 9 DSGVO Rn. 27.

285 Kampert, in: Sydow/Marsch, Art. 9 DSGVO Rn. 33.

286 S. unter D.II.5.b.

dar und ein hinreichender Zusammenhang mit der justiziellen Tätigkeit der Gerichte bleibt gewahrt.

cc. Art. 9 Abs. 2 lit. g DSGVO (Öffnungsklausel für ein erhebliches öffentliches Interesse)

Art. 9 Abs. 2 lit. g DSGVO enthält eine Öffnungsklausel, zu deren Wahrnehmung die nationale Rechtsvorschrift einem erheblichen öffentlichen Interesse dienen muss. Die landesrechtlichen Generalklauseln (z.B. aus Art. 4 Abs. 1 BayDSG) bedienen allerdings nicht ein solches erhebliches Interesse. Das zeigt sich in systematischer Hinsicht etwa an Art. 8 BayDSG, der die Verarbeitung besonderer Kategorien personenbezogener Daten adressiert.

Entsprechendes ergibt sich für das DSG NRW, das in § 16 DSG NRW Regelungen für die Verarbeitung besonderer Kategorien personenbezogener Daten vorsieht. In beiden Vorschriften finden sich im Ergebnis keine Tatbestände, auf die sich die untersuchungsgegenständliche Verarbeitung belastbar stützen lässt.

Allerdings ist eine Verarbeitung beispielsweise zu Zwecken des Rechts der sozialen Sicherheit (Art. 8 Abs. 1 S. 1 Nr. 1 BayDSG, § 16 Abs. 1 Nr. 4 DSG NRW), der Arbeitsmedizin (Art. 8 Abs. 1 S. 1 Nr. 3 BayDSG), der öffentlichen Gesundheit (Art. 8 Abs. 1 S. 1 Nr. 4 BayDSG, § 16 Abs. 1 Nr. 3 DSG NRW) sowie zur Abwehr erheblicher Nachteile für die öffentliche Sicherheit (Art. 8 Abs. 1 S. 1 Nr. 5 BayDSG i.V.m. Art. 6 Abs. 2 Nr. 3 lit. a BayDSG, § 16 Abs. 1 Nr. 1 DSG NRW) zugelassen.²⁸⁷

Die unionsrechtlich in Art. 10 Abs. 5 KI-VO vorgesehene Rechtsgrundlage²⁸⁸ für die Verarbeitung besonderer Kategorien personenbezogener Daten erlaubt nur die Verarbeitung „für die Erkennung und Korrektur von Verzerrungen im Zusammenhang mit Hochrisiko-KI-Systemen“. Die vorliegend für das GSJ-Projekt in Rede stehende Verarbeitung erfolgt allerdings gerade nicht gezielt zum Zwecke einer Verzerrungserkennung.

287 Ähnl. auf Bundesebene nach § 22 BDSG.

288 Steinrötter/Markert, RDi 2024, 400 (Rn. 24).

dd. Art. 9 Abs. 2 lit. j DSGVO (wissenschaftliche Forschungszwecke)

Art. 9 Abs. 2 lit. j DSGVO enthält eine Öffnungsklausel für wissenschaftliche Forschungszwecke, von der die nationalen und Landesgesetzgeber in § 27 BDSG²⁸⁹ sowie in Art. 8 Abs. 1 S. 1 Nr. 5 BayDSG i.V.m. Art. 6 Abs. 2 Nr. 3 lit. c BayDSG und § 17 DSG NRW Gebrauch gemacht haben. Die hier maßgeblichen landesrechtlichen Vorschriften verlangen jeweils im Einklang mit Art. 89 Abs. 1 DSGVO²⁹⁰ eine Interessenabwägung sowie gegebenenfalls weitere technische und organisatorische Maßnahmen.

Für die Anwendbarkeit von Art. 9 Abs. 2 lit. j DSGVO kommt es entscheidend darauf an, ob die betreffenden Verarbeitungen wissenschaftlichen Forschungszwecken dienen. Die Regelung ist unionsautonom auszulegen, wobei der Begriff der wissenschaftlichen Forschungszwecke wegen der unterschiedlichen Konzeption beider Rechtsgebiete nicht zwangsläufig in allen Einzelheiten mit dem nachfolgend unter E.II.2.c.dd erörterten urheberrechtlichen Verständnis gleichläuft.²⁹¹ Im Rahmen sowohl des Datenschutzrechts als auch des Urheberrechts wird die Verfolgung wissenschaftlicher Forschungszwecke allerdings jedenfalls dann abzulehnen sein, wenn Handlungen mit Forschungsbezug (z.B. Verarbeitungen oder Vervielfältigungen zur Entwicklung „innovativer“ Produkte) unmittelbar und vorrangig auf die Entwicklung eines Produkts für den operativen Einsatz abzielen.

Der Begriff der wissenschaftlichen Forschungszwecke ist (auch) im Kontext der DSGVO unter Berücksichtigung der grundrechtlich garantierten akademischen Freiheit (Art. 13 S. 2 GRCh) auszulegen; umfasst ist hiervon „alles, was nach Inhalt und Form als ernsthafter, planmäßiger Versuch zur Ermittlung der Wahrheit anzusehen ist“.²⁹² Der europäische Gesetzgeber geht in ErwGr. 159 S. 2 DSGVO ausdrücklich von einer weiten Auslegung aus,²⁹³ wonach der Begriff beispielsweise „die technologische Entwicklung und die Demonstration, die Grundlagenforschung, die angewandte Forschung und die privat finanzierte Forschung“ umfassen soll. Die Inbezug-

289 S. insb. zur Unsicherheit, ob die Regelung eine belastbare Rechtsgrundlage darstellt, *LfDIBW*, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, S. 35.

290 Zu der Rechtsnatur des Art. 89 Abs. 1 DSGVO *Matzke*, in: BeckOK Datenschutzrecht, Art. 89 DSGVO Rn. II.

291 Hierzu ausf. unter E.II.2.c.dd.

292 Vgl. *BVerfG*, BeckRS 1973, 104803 (Rn. 75); *Weichert*, ZD 2020, 18 (19); *Rofsnagel*, ZD 2019, 157 (158); *T. Radtke*, ZGE 17 (2025), 1 (39 f.) m.w.N.

293 *LfDIBW*, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, S. 35.

nahme der angewandten Forschung legt nahe, dass auch Forschungsprojekte dem Anwendungsbereich unterfallen können, die auf einen praktischen Einsatz gerichtet sind.

In systematischer Hinsicht ist bei einer Gesamtbetrachtung mit den übrigen Öffnungsklauseln des Art. 9 Abs. 2 DSGVO aber jedenfalls ein Bezug zu einem konkreten Forschungsprojekt erforderlich, das nicht bloß Durchgangsstadium zu einem von vornherein feststehenden Produktiveinsatz ist. Denn eine zu weite Auslegung würde dazu führen, dass ansonsten letztlich die Entwicklung jeder Software und die Untersuchung ihrer Effektivität in der Praxis bereits als wissenschaftliche Forschung zu qualifizieren wäre, womit es bereits nicht mehr der übrigen Ausnahmen in Art. 9 Abs. 2 DSGVO bedürfte.

Vor diesem Hintergrund ist im GSJ-Projekt zwar grundsätzlich denkbar, einzelne Verarbeitungen, die schwerpunktmäßig der Forschung dienen, auf die Rechtsgrundlage des Art. 9 Abs. 2 lit. j DSGVO zu stützen (z.B. Debugging-Maßnahmen zu optimiertem Lernverhalten). Keinesfalls aber können sämtliche Verarbeitungen im Trainingsprozess und im späteren Einsatz hierauf gestützt werden.

c. Personenbezogene Daten über strafrechtliche Verurteilungen und Straftaten (Art. 10 DSGVO)

Sofern vereinzelt eine Verarbeitung personenbezogener Daten über strafrechtliche Verurteilungen und Straftaten²⁹⁴ erfolgt, z.B. im Zusammenhang mit der Beiziehung von strafrechtlichen Akten im Zivilverfahren, ist Art. 10 S. 1 DSGVO zu beachten. Nach dieser Vorschrift darf die Verarbeitung nur unter behördlicher Aufsicht, aber ohne das Erfordernis einer über Art. 6 Abs. 1 DSGVO hinausgehenden Rechtsgrundlage,²⁹⁵ oder gestützt auf eine entsprechende Rechtsvorschrift erfolgen. Die Einzelheiten des Anwendungsbereichs von Art. 10 DSGVO mit Blick auf Zeugen, schuldloses Handeln und mutmaßliche Straftaten sind umstritten,²⁹⁶ wirken sich vorliegend mit Blick auf die Beschränkung auf zivilrechtliche Verfahren aber ohnehin

294 Ordnungswidrigkeiten sind nach vorzugswürdiger, aber bestrittener Ansicht aufgrund des Wortlauts nicht mitgemeint, weil die DSGVO gerade zwischen Straftaten und bspw. Bußgeldern nach Art. 83 DSGVO unterscheidet. Zum Meinungsstand *Nolde*, in: Taeger/Gabel, Art. 10 DSGVO Rn. 9 m.w.N.

295 *Petri*, in: Simitis/Hornung/Spiecker gen. Döhmman, Art. 10 DSGVO Rn. 1.

296 Überblick etwa bei *Nolde*, in: Taeger/Gabel, Art. 10 DSGVO Rn. 9 ff.

kaum aus. Besonders für etwaige Strafverfahren gilt es vorgelagert, die Voraussetzungen nach Art. 10 DSGVO zu prüfen und zu wahren.

Die behördliche Aufsicht schließt zwar nicht per se die Weitergabe an Dritte aus, verlangt aber eine in diesen Fällen fortbestehende Aufsicht und effektive Steuerung durch die Behörde.²⁹⁷ Eine Sachverhaltskonstellation, die einer nicht-behördlichen Stelle eine eigenständige Gestaltung der Verarbeitung ermöglicht und eine gemeinsame Verantwortlichkeit begründet, wird diesen Anforderungen nicht gerecht. Die hier zwischen den Ministerien und den ausführenden Stellen anzunehmende Auftragsverarbeitung erfüllt grundsätzlich das Merkmal der behördlichen Aufsicht und damit die Anforderungen aus Art. 10 S. 1 Var. 1 DSGVO.

Eine spezifische Regelung i.S.d. zweiten Tatbestandsalternative des Art. 10 S. 1 DSGVO ist für das untersuchungsgegenständliche GSJ-Projekt nicht ersichtlich.

d. Verhältnis zu Anforderungen an die Zweckänderung (Art. 5 Abs. 1 lit. b, Art. 6 Abs. 4 DSGVO)

Die Frage nach dem Verhältnis von Anforderungen an die Zweckänderung nach Art. 6 Abs. 4 DSGVO und allgemeinen Rechtsgrundlagen²⁹⁸ wirkt sich auch für Art. 9 DSGVO aus. Für die Kategorien personenbezogener Daten als maßgebliches Abwägungskriterium in Art. 6 Abs. 4 lit. c DSGVO könnte auch in diesen Fällen keine separate Rechtsgrundlage aus Art. 9 Abs. 2 DSGVO für eine Zweckänderung bei der Weiterverarbeitung durch den ursprünglichen Verantwortlichen erforderlich sein.

Im Ergebnis bedarf diese Frage vorliegend jedoch keiner Entscheidung. Denn selbst wenn man eine separate Rechtsgrundlage aus Art. 9 Abs. 2 DSGVO für nicht erforderlich hält, sind ähnliche Erwägungen mit hohem Gewicht in die Abwägung nach Art. 6 Abs. 4 DSGVO einzubeziehen. Damit sind nach beiden Lösungswege strenge Anforderungen an die Verarbeitung besonderer Kategorien personenbezogener Daten zu stellen.

297 Frenzel, in: Paal/Pauly, Art. 10 DSGVO Rn. 6; Schiff, in: Ehmann/Selmayr, Art. 10 DSGVO Rn. 7.

298 S. oben unter DV.4.e.

e. Schlussfolgerungen für das untersuchungsgegenständliche Projekt

Da nach den vorherigen Ausführungen kein Rechtfertigungstatbestand aus Art. 9 Abs. 2 DSGVO vorliegen dürfte, auf den die Verarbeitungen des GSJ-Projekts belastbar in allen Fällen gestützt werden könnten, empfiehlt sich eine weitergehende Filterung und gegebenenfalls händische Anonymisierung mit Blick auf besondere Kategorien personenbezogener Daten.

Wenn und soweit durch das Erlanger Tool bereits ein Anonymisierungsstandard erreicht wurde, der den allgemeinen Anforderungen an eine Urteilsveröffentlichung entspricht, kann grundsätzlich auch die Weiterverarbeitung zu Zwecken des KI-Trainings als eingriffsarme Anonymisierungstechnik i.w.S. erfolgen.

Zunächst empfiehlt sich eine Clusterung nach:

- auszuschließenden Verfahrenskategorien (z.B. Asylrecht, Arzthaftungssachen, Strafrecht, Familienrecht) und
- zu berücksichtigenden Verfahrenskategorien (z.B. Mietrecht, Verkehrsrecht).

Beispielhaft können die folgenden Merkmale als Ausgangspunkte für eine weitergehende Filterung und gegebenenfalls händische Anonymisierung herangezogen werden:

- Geltendmachung von deliktsrechtlichen Schadensersatzansprüchen (insbesondere § 823 Abs. 2 BGB i.V.m. StGB);
- Geltendmachung von Schmerzensgeld nach § 253 Abs. 1, 2 BGB;
- Geltendmachung von Unterlassungsansprüchen mit Blick auf (politische) Aussagen, wobei gegebenenfalls eine offensichtliche Öffentlichmachung nach Art. 9 Abs. 2 lit. e DSGVO in Betracht kommen kann;
- Arztberichte innerhalb der Aktenauszüge;
- Beigezogene Straftaten als Teil der zivilrechtlichen Aktenauszüge;
- Vorkommen von Begriffen wie „Gewerkschaft“ sowie geläufigen Gewerkschafts- und Parteibezeichnungen innerhalb der Aktenauszüge.

6. Weitere Pflichten aus der DSGVO

Die weiteren Pflichten aus der DSGVO unterliegen keinen projektbezogenen Besonderheiten, sind aber von Relevanz mit Blick auf den Einsatz eines KI-Systems und werden daher nachfolgend im Überblick erläutert, soweit diese DSGVO-Pflichten gerade Fragen mit Blick auf das Training

des Sprachmodells und dessen Einsatz als KI-System in der Justiz aufwerfen.

a. Informationspflichten (Art. 13, 14 DSGVO)

Die Informationspflicht aus Art. 14 Abs. 1-4 DSGVO findet im Ergebnis aufgrund der in Art. 14 Abs. 5 lit. b Hs. 1 Var. 2 DSGVO niedergelegten Ausnahme des unverhältnismäßigen Aufwands der Informationserteilung keine Anwendung. Es sind allerdings geeignete Maßnahmen zum Schutz betroffener Personen zu treffen, wobei diese geeigneten Maßnahmen die Bereitstellung von Informationen für die Öffentlichkeit einschließen.²⁹⁹

Die Informationspflicht, auch im Hinblick auf eine Zweckänderung (siehe Art. 14 Abs. 4 DSGVO), richtet sich mangels Direkterhebung durch das Ministerium als erste relevante Stelle im Rahmen dieser Untersuchung nach Art. 14 DSGVO. Konkret wird regelmäßig eine Direkterhebung durch das jeweilige Gericht oder die Parteien und deren Bevollmächtigten erfolgt sein.

Ausnahmen von der Informationspflicht sind insbesondere vorgesehen, wenn die Erteilung der Informationen einen unverhältnismäßigen Aufwand erfordern würde (Art. 14 Abs. 5 lit. b Hs. 1 Var. 2 DSGVO) – so vor allem auch bei wissenschaftlichen Forschungszwecken. Die Artikel-29-Datenschutzgruppe verlangt insoweit einen erhöhten Begründungsaufwand für die Unverhältnismäßigkeit bei anderen als wissenschaftlichen Forschungszwecken.³⁰⁰

Die hier untersuchungsgegenständliche KI-Entwicklung dient zwar nicht primär einem Forschungsprojekt, weist aber eine Nähe zur Forschung i.S.d. in Art. 14 Abs. 5 lit. b Hs. 1 Var. 2 DSGVO verankerten Regelbeispiels auf. Diese Forschungsnähe zeigt sich nicht nur an dem Einsatz innovativer Technologien, sondern vor allem an dem Zugriff auf große Mengen an Daten, um hieraus losgelöst von den einzelnen betroffenen Personen neue Informationen bzw. Inhalte zu entwickeln. Weiterhin sind die hohe Anonymisierungsrate, die nur sehr vereinzelt mögliche Datenextraktion und das Vorhandensein von zusätzlichen Schutzmaßnahmen zu berücksichtigen. Nach alledem wird das untersuchungsgegenständliche GSJ-Projekt im Zuge

299 Diese Informationsbereitstellung kann sich gegebenenfalls auf die urheberrechtliche Bewertung auswirken, indem vermehrt mit der Erklärung von Vorbehalten i.S.d. § 44b Abs. 3 UrhG zu rechnen ist, s. E.II.2.c.cc(3).

300 Artikel-29-Datenschutzgruppe, WP 260 rev.01, Rn. 61

einer Gesamtbetrachtung ebenfalls auf diese Ausnahme gestützt werden können.

Die Ausnahme der ernsthaften Beeinträchtigung einer Verwirklichung der Ziele der Verarbeitung (Art. 14 Abs. 5 lit. b Hs. 2 DSGVO) zielt vor allem auf Verarbeitungen, die gerade eine Geheimhaltung erfordern (z.B. interne Ermittlungen), und wird daher im vorliegenden Fall grundsätzlich nicht zur Anwendung gelangen.

Der Verantwortliche muss in jedem Fall weitere Schutzmaßnahmen prüfen. Hierzu zählt insbesondere die regelmäßig erforderliche³⁰¹ öffentliche Bereitstellung dieser Informationen, z.B. über eine Internetseite. Die weiteren Maßnahmen können etwa eine Datenschutzfolgenabschätzung sowie technische und organisatorische Maßnahmen umfassen; insoweit kann auf die nachfolgenden Ausführungen unter c. verwiesen werden.

b. (Weitere) Betroffenenrechte und Datenschutzgrundsätze

Darüber hinaus sind die weiteren Anforderungen aus den Art. 15 ff. DSGVO zu beachten, die beispielsweise auf Berichtigungs- und Löschungsersuchen abstellen, zugleich aber auch eigeninitiativ über die Datenschutzgrundsätze i.V.m. Art. 24 ff. DSGVO den Verantwortlichen auferlegt werden können. Diese Anforderungen an die Ausgestaltung stellt den Umgang mit KI-Systemen, wie er im GSJ-Projekt vorgesehen ist, vor Herausforderungen.³⁰²

Für diese Herausforderungen können in der Praxis zwar belastbare Lösungen gefunden werden, die aber wegen der charakteristischen Besonderheiten von KI-Systemen das Risiko eines Verstoßes gegen Art. 5, 15 ff. DSGVO vielfach nicht werden restlos ausräumen können. Im GSJ-Projekt dürfte es aufgrund der grundsätzlich nur justizinternen Weiterverarbeitung der KI-Ausgabe³⁰³ allerdings ohnehin allenfalls selten zur Geltendmachung von Betroffenenersuchen kommen.

Im Fall der Geltendmachung ist der Verantwortliche nach Art. 11 Abs. 1, 2 DSGVO gegebenenfalls nicht zur eigenständigen Verknüpfung der Trai-

301 Paal/Hennemann, in: Paal/Pauly, Art. 14 DSGVO Rn. 41; a.A. Schmidt-Wudy, in: BeckOK Datenschutzrecht, Art. 14 DSGVO Rn. 99.

302 Berz/Engel/Hacker, ZUM 2023, 586 (589); Werry, MMR 2023, 911 (914).

303 Denkbar bleibt in den Use-Cases allerdings eine Übernahme generierter KI-Ausgaben in eine Gerichtsentscheidung, die dann der Öffentlichkeit zugänglich gemacht werden.

ningsdaten im Einzelfall zur Identifizierung der betroffenen Personen verpflichtet, muss aber zusätzliche, von der betroffenen Person bereitgestellte Informationen zur Identifizierung und Bearbeitung eines Ersuchens berücksichtigen.³⁰⁴

aa. Auskunft und Datenkopie (Art. 15 DSGVO)

Nach Art. 15 Abs. 1, 3 DSGVO hat die betroffene Person ein Recht auf Auskunft und auf eine Kopie der verarbeiteten personenbezogenen Daten. Im Hinblick auf den Einsatz des KI-Systems dürfte vor allem die nach Art. 15 Abs. 1 S. 1 Hs. 1 DSGVO vorzunehmende Bestätigung über die Verarbeitung personenbezogener Daten die Verantwortlichen vor Herausforderungen stellen. Denn der Verantwortliche hat regelmäßig gerade keine Kenntnis davon, ob Daten einer einzelnen Person verarbeitet werden oder nicht. Erst wenn dem Verantwortlichen ein Datenleck als Ergebnis einer KI-Ausgabe bekannt wird, kann der Verantwortliche die betreffende Datenverarbeitung bestätigen, gegebenenfalls anhand des Trainingskorpus lokalisieren und die weiteren Informationen nach Art. 15 Abs. 1 DSGVO bereitstellen.

Mithin besteht das Risiko, einem Betroffenenersuchen nach Art. 15 DSGVO mangels Kenntnis des Verantwortlichen über die konkreten Datenverarbeitungen nicht hinreichend entsprechen zu können. Für diesen Fall kann unter bestimmten Voraussetzungen ein Unmöglichkeitseinwand ins Feld geführt werden, der sich in der Rechtsprechung des EuGH zumindest für die Auskunft über einzelne Informationskategorien abzeichnet.³⁰⁵ Ob der EuGH einen solchen Einwand in diesem Umfang anerkennt, bleibt allerdings abzuwarten. Der EuGH könnte insoweit auch bloß eine aufgrund der Technologie selbst herbeigeführte Unmöglichkeit erkennen.³⁰⁶

Das vorgenannte Risiko eines Verstoßes gegen Art. 15 DSGVO wird aber durch folgende Überlegung zumindest begrenzt: Das Sprachmodell weist regelmäßig für sich genommen keinen Personenbezug auf,³⁰⁷ sondern es kommt erst durch eine Zusammenführung verschiedener Informationen in der KI-Ausgabe zu einer Verarbeitung i.S.d. DSGVO. Diese Zusammenführung im Rahmen der Ausgabe wird grundsätzlich durch verschiedene Maßnahmen verhindert. Sofern es zu einer solchen Zusammenführung

304 Hierzu etwa *Ashkar*, ZD 2023, 523 (528).

305 *EuGH*, NJW 2023, 973 (Rn. 48) – Österreichische Post.

306 *Mertens/D. Meyer*, K&R 2023, 563 (568).

307 S. unter D.IV.1.

kommt, ist durch technische und organisatorische Maßnahmen sicherzustellen, dass die Verantwortlichen über bekanntgewordene Datenlecks und damit auch über das Vorliegen einer Verarbeitung informiert werden, z.B. durch die Möglichkeit für die das KI-System einsetzende Richterschaft, die KI-Ausgabe zur Überprüfung zu melden.

bb. Berichtigung und Mitteilungspflicht (Art. 16 und 19 DSGVO)

Nach Art. 16 DSGVO hat die betroffene Person das Recht, die Berichtigung unrichtiger bzw. Vervollständigung unvollständiger, sie betreffender personenbezogener Daten zu verlangen. Die Erfüllung dieses Rechts kann aus den vorgenannten Gründen in Ansehung von KI-Systemen (ebenfalls) nicht nur unerhebliche Schwierigkeiten für Verantwortliche aufwerfen.

Allerdings ist zu bedenken, dass die Richtigkeit oder Vollständigkeit der personenbezogenen Daten aus den Trainingsdaten selten mit Erfolg wird bestritten werden können.³⁰⁸ Denn die vollständige Wiedergabe einer Gerichtsentscheidung in den Trainingsdaten als Dokumentation der Entscheidung des Gerichts ist gerade nicht unrichtig. Denn die Richtigkeit verlangt eine Übereinstimmung der Informationen³⁰⁹ mit der Realität³¹⁰ unter Berücksichtigung des Informationskontexts:³¹¹ Die Information muss als Tatsache geäußert werden,³¹² z.B. als durch objektive oder durch ein Gericht ordnungsgemäß festgestellte Tatsache,³¹³ oder im Sinne eines graduellen Ansatzes zumindest einen Tatsachengehalt in Form von – gegebenenfalls auch wertenden – Prognosen enthalten.³¹⁴

308 Anders z.B. beim Scraping *Mertens/D. Meyer*, K&R 2023, 563 (569).

309 Also nicht bloß den Prozess der mathematisch richtigen Erzeugung, *Engeler/Rolfes*, ZD 2024, 423 (428).

310 *Herbst*, in: Kühling/Buchner, Art. 5 DSGVO Rn. 60; *Voigt*, in: Taeger/Gabel, Art. 5 DSGVO Rn. 30.

311 *Stevens*, CR 2020, 73 (Rn. 8) m.w.N.

312 *Herbst*, in: Kühling/Buchner, Art. 5 DSGVO Rn. 60.

313 Vgl. auch *Roßnagel*, in: Simitis/Hornung/Spiecker gen. Döhmman, Art. 5 DSGVO Rn. 140.

314 *Hoeren*, ZD 2016, 459 (462) unter Berücksichtigung der unterschiedlichen Sprachfassungen; *Hallinan/Zuiderveen Borgesius*, International Data Privacy Law 10 (2020), 1; *Stevens*, CR 2020, 73 (Rn. 9); *Schantz*, in: BeckOK Datenschutzrecht, Art. 5 DSGVO Rn. 27; im Hinblick auf die Datengrundlagen für automatisierte Entscheidungen *Artikel-29-Datenschutzgruppe*, WP251 rev.01, S. 12 f.; diff. *Pesch/Böhme*, MMR 2023, 917 (923); die Vorschrift hingegen auf reine Tatsachenbehauptungen beschränkend *Herbst*, in: Kühling/Buchner, Art. 5 DSGVO Rn. 60; *Werry*,

Erst die neue Zusammensetzung durch die Generierung der KI-Ausgabe kann zu einer Unrichtigkeit führen (z.B. in Gestalt einer „Halluzination“),³¹⁵ sofern man annimmt, das KI-System stelle insoweit überhaupt eine Behauptung mit Tatsachengehalt auf.³¹⁶ Ein solcher Tatsachengehalt wird insbesondere in Abhängigkeit davon anzunehmen sein, wie zuverlässig das KI-System eingeschätzt³¹⁷ und wofür es eingesetzt wird. Im untersuchungsgegenständlichen GSJ-Projekt dürften gelegentliche personenbezogene³¹⁸ Ausgaben aus den Trainingsdaten sowie den Eingabedaten aus der neuen Akte in den Use-Cases einer Sachverhaltszusammenstellung einer neuen Akte als Tatsache wahrgenommen werden. Die Unrichtigkeit der Angaben kann aber gegebenenfalls durch diese entsprechende Temperatur-Einstellung eingedämmt werden.³¹⁹

Unter dieser Prämisse müsste die betreffende KI-Ausgabe, soweit sie denn gespeichert wurde,³²⁰ durch den Verantwortlichen bearbeitet oder gelöscht und müssten etwaige Empfänger nach Art. 19 DSGVO hierüber in Kenntnis gesetzt werden. Sofern eine Berichtigung im Gesamtkontext der KI-Ausgabe nicht zielführend erscheint, wird es mit Blick auf die gerade unbeabsichtigte und aus Sicht des Verantwortlichen zu vermeidende Verarbeitung sowie der regelmäßig verwirklichten Befriedigung der Interessen der betroffenen Personen durch die Löschung als zulässig anzusehen sein, erst recht und nur die Löschung oder gegebenenfalls Einschränkung der Verarbeitung (Art. 17 und 18 DSGVO) vorzunehmen.³²¹

cc. Löschung und Mitteilungspflicht (Art. 17 und 19 DSGVO), auch im Zusammenhang mit einem Widerspruch (Art. 21 DSGVO)

Die Löschung ist von Gesetzes wegen (Art. 5 Abs. 1, Art. 25 Abs. 1 DSGVO) oder auf Ersuchen einer betroffenen Person hin auch und gerade dann

MMR 2023, 911 (914); wohl auch *Roßnagel*, in: Simitis/Hornung/Spiecker gen. Döhmman, Art. 5 DSGVO Rn. 140.

315 Hierzu schon oben unter D.IV.1.b.dd(2) und D.IV.2.

316 Hierzu auch *EDPB*, Report of the work undertaken by the ChatGPT Taskforce, Rn. 30 f.; dies i.E. annehmend *Pesch/Böhme*, MMR 2023, 917 (922).

317 *Pesch/Böhme*, MMR 2023, 917 (922).

318 Also gerade nicht Angaben bezogen auf eine fiktive Person, s. *LfDIBW*, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, S. 9.

319 S. bspw. <https://platform.openai.com/docs/api-reference/chat/create>.

320 Vgl. *Paal*, in: *Paal/Pauly*, Art. 16 DSGVO Rn. 15.

321 Vgl. auch *EDPB*, Report of the work undertaken by the ChatGPT Taskforce, Rn. 34.

vorzunehmen, wenn die betroffene Person wirksam einen Widerspruch eingelegt hat (Art. 21 Abs. 1 DSGVO) und keine vorrangigen berechtigten Gründe des Verantwortlichen vorliegen (Art. 17 Abs. 1 lit. c DSGVO). Derartige Fallkonstellationen dürften bei der justizinternen Verarbeitung kaum vorkommen.

Falls in einem KI-basiert vorformulierten Sachverhaltsabschnitt einer Gerichtsentscheidung (siehe den zweiten Use-Case) personenbezogene Daten aus den Trainingsdaten enthalten sind und über die Gerichtsentscheidung der Öffentlichkeit zugänglich gemacht werden, mag die Löschung nach Art. 17 DSGVO allerdings relevant werden. Die nach Art. 6 Abs. 1 UAbs. 1 lit. e, Abs. 4 DSGVO (oder Art. 6 Abs. 1 UAbs. 1 lit. f DSGVO) vorgenommene Interessenabwägung³²² ist dann gegebenenfalls für den konkreten Einzelfall zu überprüfen.

Der Verantwortliche³²³ muss nach dem Eingang eines berechtigten Löschersuchens sicherstellen, dass das Datenleck den jeweiligen Gerichtsentscheidungen aus dem Trainingsdatensatz zugeordnet wird und die Urteile in dem Training künftiger KI-Modelle ausgeschlossen oder vollständig anonymisiert werden. Das gilt in Ansehung der erzielten Ausgabe als – gegebenenfalls erneute reproduzierbare – Verarbeitung personenbezogener Daten unabhängig davon, ob das Sprachmodell selbst als eine Speicherung personenbezogener Daten anzusehen ist.

Nach Art. 19 DSGVO sind erforderlichenfalls etwaige Empfänger über die Löschung zu informieren. In diesem Zusammenhang sollte der Verantwortliche als technische und organisatorische Maßnahme zur Sicherstellung der Rechtmäßigkeit der Verarbeitung (Art. 5 Abs. 1 lit. a, Art. 24 f. DSGVO) gegenüber den Nutzern des KI-Systems eine (dienst-)vertragliche Meldepflicht etablieren oder zumindest auf die Meldemöglichkeit prominent hinweisen.

Die Löschung muss grundsätzlich binnen einer Frist von einem Monat nach Art. 12 Abs. 3 S. 1 DSGVO erfolgen; die Löschungsfrist verlängert sich ausnahmsweise „unter Berücksichtigung der Komplexität und der Anzahl von Anträgen“ um zwei weitere Monate. Diese Ausnahme ist restriktiv anzuwenden (vgl. ErwGr. 59 DSGVO). Zwar kann die Schwierigkeit, die konkreten Informationen auszulesen, im Hinblick auf einzelne Anträge eine

322 S. oben unter DV.4.c.

323 In der Regel die jeweiligen Ministerien, s. oben unter DV.3.b.dd.

Fristverlängerung rechtfertigen,³²⁴ nicht aber pauschal der Einsatz eines besonderen (KI-)Systems eine Fristverlängerung zum Regelfall erheben.

Der Verantwortliche kann sich insoweit jedoch gegebenenfalls anderweitig behelfen: Unter Beachtung des oben aufgestellten Maßstabs³²⁵ kann die primäre Anonymisierung und sekundäre eingeschränkte justizinterne Verarbeitung grundsätzlich auf Art. 6 Abs. 1 UAbs. 1 lit. e, f DSGVO oder Art. 6 Abs. 4 DSGVO gestützt werden. Das gilt im Einzelfall auch für besondere Kategorien personenbezogener Daten nach Art. 9 Abs. 2 lit. f DSGVO. Einzelne Löschungen können auf Ausgabebene realisiert werden: System-Prompts und Ausgabenfilter, die die konkrete Verarbeitung im Rahmen der Ausgabe verhindern, können faktisch einer Löschung gleichstehen oder gar eine Anonymisierung herbeiführen, falls keine wahrscheinlichen Mittel zur Reproduktion der personenbezogenen Daten in der Ausgabe in Betracht kommen.

Den Befund der Zulässigkeit dieses Vorgehens stützt überdies § 35 Abs. 1 BDSG, der eine Ausnahme von dem Recht auf Löschung definiert. Für Daten i.S.d. Art. 9, 10 DSGVO, die trotz der ergriffenen Maßnahmen offengelegt werden, kann im Einzelfall abhängig von der Komplexität eine Fristverlängerung nach Art. 12 Abs. 3 S. 2 DSGVO gerechtfertigt sein.

dd. Einschränkung der Verarbeitung und Mitteilungspflicht (Art. 18 und 19 DSGVO)

Die Einschränkung der Verarbeitung (Art. 18 DSGVO) als „Einfrieren“ zur Klärung oder zu Dokumentationszwecken und zur Mitteilung nach Art. 19 DSGVO lässt sich entsprechend umsetzen, z.B. wenn die Richtigkeit einer Ausgabe bestritten wird oder die betroffene Person Widerspruch eingelegt hat und (noch) nicht feststeht, ob die berechtigten Gründe des Verantwortlichen gegenüber denen der betroffenen Person überwiegen. Wie bereits unter DV.5.b aufgezeigt, dürften diese Fälle im GSJ-Projekt praktisch allenfalls (sehr) selten vorkommen.

324 EDPB, Leitlinien 01/2022 zu den Rechten der betroffenen Person – Auskunftsrecht, Rn. 163.

325 S. oben unter DV.4.

c. Datenschutzfolgenabschätzung und Pflicht zur Implementierung technischer und organisatorischer Maßnahmen

Als eine besondere technische und organisatorische Maßnahme ist die Datenschutzfolgenabschätzung (Art. 35 Abs. 1 DSGVO) hervorzuheben, die insbesondere die Anonymisierungsmaßnahmen (einschließlich der Ausgabefilter) und das verbleibende Restrisiko nachvollzieht. Ein solches Erfordernis ergibt sich zwar nicht zwingend aus Art. 35 Abs. 3 DSGVO, kann aber gleichwohl mit Blick auf die Verwendung neuer Technologien i.S.d. Art. 35 Abs. 1 DSGVO³²⁶ sowie die Quantität der möglicherweise verarbeiteten personenbezogenen Daten einschließlich der zu vermeidenden Verarbeitung besonderer Kategorien personenbezogener Daten i.S.d. Art. 9, 10 DSGVO angezeigt sein (vgl. Art. 35 Abs. 3 lit. b DSGVO).³²⁷

Darüber hinaus sind zur Umsetzung der Datenschutzgrundsätze, so auch und gerade der Datenminimierung, Speicherbegrenzung, Integrität und Vertraulichkeit (Art. 5 Abs. 1 lit. c, e, f DSGVO) technische und organisatorische Maßnahmen zu ergreifen (Art. 24, 25, 32 DSGVO). Das gilt auch mit Blick auf die landesrechtlichen Vorgaben (z.B. § 3 DSG NRW). Diese Verpflichtung umfasst die regelmäßige Überprüfung, ob das eingesetzte Anonymisierungstool, die Auswahl der ein- und ausgeschlossenen Urteilskategorien, die Ausgabefilter, der System Prompt sowie die eingesetzte Software und Übermittlungswege dem Stand der Technik und IT-Sicherheitsanforderungen entsprechen.³²⁸ Diese Überprüfung sollte proaktiv in regelmäßigen Abständen und zudem als Reaktion auf etwaige Datenschutzverletzungen i.S.d. Art. 33, 34 DSGVO erfolgen. Grundlage für diese Überprüfung ist eine sorgfältige Dokumentation einschließlich der Informationen über die verwendeten Trainingsdaten³²⁹ und erfolgten Anonymisierungsmaßnahmen.

Zudem empfiehlt es sich, den Nutzern des KI-Systems eine einfache Möglichkeit zur Verfügung zu stellen, um KI-Ausgaben gegebenenfalls zur Überprüfung an die Verantwortlichen zu melden. Eine generelle Markie-

326 Ashkar, ZD 2023, 523 (529).

327 Vgl. Artikel-29-Datenschutzgruppe, WP248 rev.01, S. 10 ff.; Dewitte, in: AI Meets the GDPR, S. 155.

328 Zur Handlungspflicht bei Änderungen des Stands der Technik Hornung/B. Wagner, CR 2019, 565 (Rn. 25 ff.).

329 BayLDA, Datenschutzkonforme Künstliche Intelligenz, S. 5; EDPB, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, Rn. 56-58.

rung der KI-Ausgaben mit einem Wasserzeichen³³⁰ kann darüber hinaus eine Nachvollziehbarkeit des Ursprungs der KI-Ausgabe bei unbeabsichtigter Weiterverarbeitung sicherstellen.

Nur kurz erwähnt sei an dieser Stelle³³¹ der Abschluss von etwaigen notwendigen datenschutzrechtlichen Vereinbarungen (insbesondere einer Auftragsverarbeitungsvereinbarung) sowie Nutzungsvereinbarungen bzw. Nutzungshinweisen.

d. Grundsatz der Speicherbegrenzung

Das Trainingskorpus kann für weitere Trainings des Sprachmodells gespeichert werden, da die Verarbeitung erforderlich ist für das KI-Training.³³² Eine restriktive Auslegung der Erforderlichkeit i.S.d. Art. 5 Abs. 1 lit. c, e DSGVO ist nicht geboten, wenn der Verantwortliche durchgehend die vorgenannten Anforderungen an die technischen und organisatorischen Maßnahmen erfüllt.

Mit Blick auf Besonderheiten einzelner Sachverhalte bedarf es jedoch gegebenenfalls einer Prüfung einzelner Datencluster, wenn sich die Sensibilität personenbezogener Daten durch Zeitablauf erhöht und sich die Abwägung nach Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO im Laufe der Zeit unter Berücksichtigung des Zwecks tendenziell zugunsten der betroffenen Person verschiebt. Von Bedeutung sind insoweit beispielsweise Informationen über eine finanzielle Notlage,³³³ wie etwa im Zusammenhang mit der Privatinsolvenz, Entscheidungen der Sozialgerichtsbarkeit oder vollstreckungsgerichtliche Entscheidungen.

Für verschiedene weitere Urteilscluster mag zwar die Intensität eines Datenlecks mit der Zeit zunehmen; dieser Effekt dürfte aber zugleich durch eine höhere Anonymisierungswahrscheinlichkeit mit Zeitablauf aufgehoben werden. Sofern beispielsweise einer Person aufgrund einer Ortsbeschreibung im Urteil bestimmte Informationen zugeordnet werden konnten, dürfte diese Zuordnung mit Zeitablauf (in Verbindung etwa mit Um-

330 EDPB, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, Rn. 107.

331 S. zu weiteren möglichen Maßnahmen auch EDPB, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, Rn. 102: Wartefrist und Opt-Out-Möglichkeit.

332 Paal, ZfDR 2024, 129 (141).

333 Vgl. *EuGH*, NJW 2024, 418 – Schufa; NJW 2014, 2257 – Google Spain.

zügen und Veränderungen der örtlichen Umstände) schwerer fallen und gegebenenfalls sogar ausscheiden.

7. Pflichten unter der JI-RL

Die JI-RL sieht besondere datenschutzrechtliche Regelungen vor für die „Verarbeitung personenbezogener Daten durch die zuständigen Behörden zum Zwecke der Verhütung, Ermittlung, Aufdeckung oder Verfolgung von Straftaten oder der Strafvollstreckung, einschließlich des Schutzes vor und der Abwehr von Gefahren für die öffentliche Sicherheit“, die in den §§ 45 ff. BDSG und landesrechtlichen Regelungen³³⁴ umgesetzt wurden. Hierunter fallen insbesondere Strafsachen und die zugehörigen Ermittlungsakten, die im GSJ-Projekt jeweils ausgenommen sind.

Vor diesem Hintergrund ist keine nähere Betrachtung der betreffenden Regelungen für das untersuchungsgegenständliche Projekt angezeigt.

8. Ergänzende Anforderungen mit Blick auf Eingabedaten

Mit Blick auf personenbezogene Daten in den Eingabeaufforderungen des entwickelten KI-Systems sind ebenfalls die oben dargestellten datenschutzrechtlichen Anforderungen zu beachten. Als Hilfsmittel im Rahmen der Bearbeitung eines konkreten Falls in Betracht kommt grundsätzlich auch die nicht-anonymisierte Eingabe im Zusammenhang mit den öffentlichen Aufgaben der Rechtspflege i.S.d. Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO i.V.m. landesrechtlichen Generalklauseln. Allerdings sind die zuvor aufgestellten Maßstäbe entsprechend anzuwenden, wenn die Eingabedaten unmittelbar und ohne weitere Anonymisierung zum Zwecke des weiteren Trainings des KI-Systems verwendet werden. Eine solche Verwendung wirft nicht nur Bedenken mit Blick auf eine Rechtsgrundlage nach Art. 6 DSGVO auf, sondern löst zusätzlich gegebenenfalls ergänzende Anforderungen aus der gemeinsamen Verantwortlichkeit aus.³³⁵

334 S. zum Verhältnis insb. § 1 Abs. 1 S. 1 Nr. 2 BDSG.

335 Hierzu ausf. *Hüger/T. Radtke*, KIR 2025, 154.

VI. Besonderheiten bei der Einbeziehung von Aktenauszügen

Für die Verwertung vollständiger Aktenauszüge und deren Bearbeitung durch das Erlanger Tool sind die Besonderheiten der Informationsanordnung zu berücksichtigen.

1. Inhalte von gerichtlichen Akten

Zunächst gilt es, sich hierzu der typischen Akteninhalte zu vergewissern, wobei diese typischen Akteninhalte hier in Ansehung der Vielfältigkeit nicht abschließend, sondern nur exemplarisch beschrieben werden können:

- Anwaltliche Schriftsätze (z.B. Klage und Klageerwiderung);
- Anlagen zu anwaltlichen Schriftsätzen (z.B. Wiedergabe von Zeitungsartikeln, Internetseiten, Fotografien, eidesstattliche Versicherungen und Zeugenaussagen);
- Urkunden und amtliche Auskünfte;
- Sitzungsprotokolle;
- Hinweis-, Beweis- und sonstige Beschlüsse;
- Terminladungen und zusammenhängende gerichtliche Verfügungen;
- Beigezogene Akten (z.B. Ermittlungsakten der Staatsanwaltschaft);
- Vermerke;
- Gutachten (z.B. Unfallgutachten oder psychologische Gutachten);
- Empfangsbekanntnisse;
- Gerichtliche (End-)Entscheidung, z.B. in Form eines Urteils;
- Handschriftliche Anmerkungen auf Aktenbestandteilen, Post-Its o.ä., die gegebenenfalls die digitale Bearbeitung erschweren.

Diese Aktenbestandteile können in unterschiedlicher Quantität und Qualität personenbezogene Daten enthalten. Das betrifft insbesondere Kennungen, wie etwa Namen, Anschriften, E-Mail-Adressen, Telefonnummern und Identifikationsnummern (z.B. der Parteien, Bevollmächtigter und Kanzleiangestellter, Zeugen, Gutachter sowie Angehörige der Staatsanwaltschaft, der übrigen Rechtspflege oder sonstiger Behörden). Hinzu kommen auch hier³³⁶ – in einer gegenüber einer Gerichtsentscheidung unstrukturierten Form – Informationen über Aktenzeichen, Ort- und Zeitangaben sowie

336 S. oben unter D.I.3.

sonstige Sachverhaltsumstände, die unter Zuhilfenahme weiterer Quellen die Identifizierung natürlicher Personen ermöglichen.

Aus dem Aufbau und Inhalt der Aktenauszüge ergeben sich also vor allem technische Herausforderungen betreffend eine qualitativ gleichbleibende Anonymisierung.

2. Rechtliche Leitplanken für die (Teil-)Anonymisierung der Aktenauszüge

Für diese technischen Herausforderungen gilt es, rechtliche Leitplanken zu herauszuarbeiten, die einerseits quantitativen Anforderungen im Hinblick auf den Recall-Wert bzw. Anonymisierungsanteil genügen, andererseits aber auch qualitative Anforderungen an die zu anonymisierenden Datenkategorien beinhalten.

a. Kein Veröffentlichungsstandard

Im Ausgangspunkt ist festzuhalten, dass die Veröffentlichung von Aktenauszügen gerade keine öffentliche, verfassungsrechtlich gebotene Aufgabe ist. Anders als bei Gerichtsentscheidungen³³⁷ existiert also kein Anonymisierungsgrad, ab dem ohnehin eine Veröffentlichung und damit regelmäßig auch eine Weiterverarbeitung zulässig wäre.

b. Quantitativ: Recall

Die für die Anonymisierung von Gerichtsentscheidungen angenommenen Recall-Werte können zwar grundsätzlich den datenschutzrechtlichen Anforderungen auf der ersten Stufe der Anonymisierung und der zweiten Stufe der Rechtfertigung nach Art. 6 Abs. 1 UAbs. 1 lit. e, f, Abs. 4 DSGVO genügen, sofern man das KI-Training sowie technische und organisatorische Maßnahmen berücksichtigt. Es dürfte aber der Recall-Wert i.H.v. 94-96 % für Kennungen und i.H.v. 68-96 % für sonstige identifizierende Merkmale signifikant unterschritten werden. Dieser reduzierte Anonymisierungsgrad kann sich sowohl auf die Frage auswirken, ob die Daten mangels hinreichender Identifizierungswahrscheinlichkeit anonym sind, als auch auf die

³³⁷ S. hierzu ausf. unter D.II.5.b.

Rechtfertigung nach Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO bei vereinzelt vorkommenden personenbezogenen Daten.

Von besonderer Relevanz bleibt auch bei Aktenauszügen ein hoher Recall-Wert mit Blick auf Kennungen. Für die übrigen identifizierenden Merkmale dürfte bei der justizinternen Verarbeitung keine hinreichende Wahrscheinlichkeit bestehen, dass mittels weiterer Quellen ein Personenbezug hergestellt wird. Für diese sonstigen identifizierenden Merkmale dürfte daher auch ein Recall-Wert von unter 50 % nicht schaden, solange die qualitativen Anforderungen gewahrt bleiben.

Enthaltene Kennungen führen hingegen unmittelbar zu einem Personenbezug. Je häufiger es zu Verarbeitungen von derartigen Kennungen kommt, desto schwerer fällt die Annahme eines belastbaren Zusammenhangs der Verarbeitung der verfahrensfremden Trainingsdaten in einem neuen Verfahren als Wahrnehmung einer öffentlichen Aufgabe i.S.d. Art. 6 Abs. 1 UAbs. 1 lit. e, Abs. 4 DSGVO.

Geringere Abweichungen des Recall-Werts für Kennungen nach unten müssen sich zunächst nicht unmittelbar auf die Frage der Anonymisierung und der Rechtfertigung auswirken. Aufgrund des Trainings des Sprachmodells und der Abhängigkeit der Ausgabe von der konkreten Eingabeaufforderung ist es selbst bei einem Recall-Wert von beispielsweise nur 90 % denkbar, dass es in weniger als 0,1 % der Fälle des konkreten KI-System-Einsatzes zu einer Ausgabe dieser Kennungen kommt. In den konkreten Use-Cases 1-3, in denen regelmäßig ein umfangreicher Eingabekontext und eine auf diese abstellende Eingabeaufforderung übergeben werden, dürfte es – vorbehaltlich einer technisch-empirischen Untersuchung – höchst unwahrscheinlich sein, dass in ein Relationsgutachten oder eine Sachverhaltszusammenfassung ein vollständiger Name oder eine Anschrift „aus“ den Trainingsdaten einfließt.

Eine signifikante Unterschreitung dieses Recall-Werts von 90 % mit Blick auf Kennungen führt aber zu einer erhöhten Wahrscheinlichkeit von nicht mehr nach Art. 6 DSGVO zulässigen Verarbeitungen personenbezogener Daten. Ein solcher Mindest-Recall-Wert dürfte sich auch unter Berücksichtigung weiterer technischer und organisatorischer Maßnahmen nach dem Stand der Technik erzielen lassen. Sofern Anonymisierungsdefizite aus einer schlecht(er)en digitalen Lesbarkeit (z.B. im Rahmen des Einsatzes von Optical Character Recognition (OCR)-Software) herrühren, können vereinzelt Kennungen schon gar nicht erst erfasst werden – und hierdurch eine höhere Anonymisierung auslösen.

Soweit zugleich Kennungen einfließen, die aufgrund von Formatänderungen nicht erkannt werden (z.B. falsch zugeordnete Zeichen, zusätzlich ergänzte Leerzeichen innerhalb eines Namens, einzelne wegen handschriftlicher Anmerkungen nicht lesbare Adressbestandteile), empfiehlt sich die Implementierung eines gesonderten Mechanismus zur Qualitätssicherung. Dokumentbestandteile, deren Digitalisierung Herausforderungen mit sich bringt, könnten separat identifiziert und gegebenenfalls aussortiert oder nachbearbeitet werden.

Ferner kann der ursprüngliche Recall-Wert stets faktisch durch effektive Ausgabefilter erhöht werden, die insbesondere typische Kennungsformate (z.B. Namen, E-Mail-Adressen, Telefonnummern und Identifikationsnummern) anhand ihres Formats durch sog. reguläre Ausdrücke erkennen und mit deren Vorkommen im Eingabekontext abgleichen. Nur wenn sich diese Angaben auch im Eingabekontext wiederfinden, spricht dem konkreten Verfahren und nicht den Trainingsdaten zuzuordnen sind, wird die Ausgabe insoweit zugelassen.

c. Qualitativ: Clusterung

Mit Blick auf die besonderen Kategorien personenbezogener Daten nach Art. 9 DSGVO erlangen die rechtlichen Leitplanken in qualitativer Hinsicht besondere Bedeutung.

Im Ausgangsfall stellt in Ansehung der justizinternen Verwendung des KI-Systems vor allem eine kombinierte Ausgabe von Kennungen (siehe zuvor zu den quantitativen Anforderungen) und besonderen Kategorien personenbezogener Daten die Verantwortlichen vor erhebliche Herausforderungen. Während die Kennung regelmäßig den Trainingsdaten entstammen dürfte, könnten unter Umständen auch halluzinierte Gesundheitsdaten oder ähnlich sensible Informationen einen besonderen Rechtfertigungsbedarf nach Art. 9 DSGVO auslösen.³³⁸ Diese Kombination kann bereits ausreichend adressiert werden, indem das Vorkommen von Kennungen in den Ausgabedaten reduziert wird.

Darüber hinaus sind an dieser Stelle besondere Kategorien personenbezogener Daten aus den Trainingsdaten in den Blick zu nehmen, deren Vorkommen in einer KI-Ausgabe zusammen mit Kennungen aus den Trainingsdaten strikt zu verhindern ist. Hierzu bedarf es einer entsprechenden Clusterung der entsprechenden Bestandteile der Aktenauszüge.

³³⁸ S. zu Halluzinationen z.B. schon oben unter DV.6.b.bb.

Der Verantwortliche ist insoweit als Ausfluss der Pflicht zur Implementierung von technischen und organisatorischen Maßnahmen nach den Art. 24 ff. DSGVO gefordert, sich einen Überblick über die in den Aktenauszügen typischerweise enthaltenen Dokumentenkategorien und deren Sensibilität mit Blick auf Art. 9 DSGVO zu verschaffen.

Auf dieser Grundlage sind einzelne Dokumentenkategorien entweder vollständig auszunehmen oder deren Anonymisierung unter Berücksichtigung des gesamten Akteninhalts händisch sorgfältig zu prüfen. Das kann z.B. – unter der Prämisse eines nicht hinreichend hohen Recall-Werts – die folgenden Dokumentenkategorien betreffen, die typischerweise eine Vielzahl an besonderen Kategorien personenbezogener Daten enthalten:

- Beigezogene Akten der Staatsanwaltschaft im Zusammenhang mit Verfahren, die typischerweise über die Straftat i.S.d. Art. 10 DSGVO hinaus sensible Informationen enthalten (z.B. Straftaten gegen die sexuelle Selbstbestimmung, schwere Körperverletzungsdelikte);
- Psychologische Gutachten;
- Ärztliche Gutachten.

Weitere Dokumentenkategorien sind gegebenenfalls nur durch einen speziell angepassten Anonymisierungsalgorithmus oder nach händischer Nachbearbeitung in das Trainingskorpus aufzunehmen. Das kann z.B. die folgenden Dokumentenkategorien betreffen:

- Beigezogene Akten der Staatsanwaltschaft im Übrigen;
- Behandlungsberichte und Überblick über Behandlungsmaßnahmen, verschriebene Medikamente u.Ä.

VII. Exkurs: Eigene Forschungszwecke der ausführenden Stellen

Im Rahmen einer Zusammenarbeit mit Forschungsorganisationen, wie etwa den ausführenden Stellen, kann sich ein Bedürfnis nach der Verfolgung eigener Forschungszwecke durch die Forschungsorganisationen ergeben. Zur Untersuchung dieser Konstellation soll daher nachfolgend hypothetisch angenommen werden, die ausführenden Stellen würden (Grundlagen-)Forschung mit sämtlichen den ausführenden Stellen von den Ministerien übermittelten Daten betreiben (nachfolgend als hypothetisches Szenario bezeichnet). In diesem Fall ergeben sich verschiedene Abweichungen zu der vorangehend vorgenommenen datenschutzrechtlichen Beurteilung. Diese Abweichungen betreffen sowohl die datenschutzrechtliche Verant-

wortlichkeit (unter 1.) und das Erfordernis einer Rechtsgrundlage für die Verarbeitungsvorgänge (unter 2.) als auch weitere datenschutzrechtliche Anforderungen (unter 3.).

In dem hypothetischen Szenario bestehen auch und gerade mit Blick auf die Übermittlung personenbezogener Daten aus den Gerichtsentscheidungen und Aktenauszügen an die ausführenden Stellen Zweifel im Hinblick auf das Vorliegen einer belastbaren Rechtsgrundlage aus Art. 6 Abs. 1 UAbs. 1 DSGVO. Etwas anderes dürfte gelten, wenn und soweit vor der Übermittlung eine Anonymisierung vorgenommen wird und der Anonymisierungsgrad den unter D.II.5.b herausgearbeiteten Anforderungen an eine Urteilsveröffentlichung entspricht, spricht über den Einsatz und die Wirkmöglichkeiten des Erlanger Tools in der derzeitigen Form hinausgeht.

1. (Gemeinsame) Verantwortlichkeit der ausführenden Stellen

Die Festlegung eigener (Forschungs-)Zwecke durch die ausführenden Stellen geht regelmäßig mit weiteren eigenen Festlegungen zu wesentlichen Mitteln der Verarbeitung (z.B. zur Speicherdauer und Empfängern) einher und begründet grundsätzlich nach Art. 4 Nr. 7 DSGVO abweichend vom Sachverhalt des GSJ-Projekts eine Verantwortlichkeit der ausführenden Stellen.³³⁹

Da die sich ergänzenden³⁴⁰ Verarbeitungsvorgänge der ausführenden Stellen und der Ministerien gemeinsame Datensätze zum Gegenstand haben,³⁴¹ auf einer für betroffene Personen in diesem Umfang nicht konkret vorhersehbaren Verarbeitung beruhen³⁴² und eine umfängliche Zusammenarbeit zwischen den Ministerien sowie den ausführenden Stellen stattfindet,³⁴³ dürften die Ministerien und die ausführenden Stellen als gemeinsam verantwortlich für die Übermittlung an die ausführenden Stellen und die Weiterverarbeitung für das GSJ-Projekt anzusehen sein. Soweit und sobald die ausführenden Stellen personenbezogene Daten über das Projekt hinaus (nur) für eigene (Grundlagen-)Forschung verarbeiten, geht

339 Ausf. zu den Anforderungen an eine (gemeinsame) Verantwortlichkeit unter D.V.3.a.

340 Vgl. *EuGH*, ZD 2024, 209 (Rn. 43); *EDPB*, Leitlinien 07/2020 zu den Begriffen „Verantwortlicher“ und „Auftragsverarbeiter“ in der DSGVO, Rn. 55.

341 *T. Radtke*, Gemeinsame Verantwortlichkeit unter der DSGVO, S. 193 ff.

342 Vgl. *EuGH*, ZD 2019, 455 (Rn. 75) – Fashion ID.

343 Vgl. *EuGH*, ZD 2024, 209 (Rn. 43 f.); ZD 2024, 328 (Rn. 59 ff.) – IAB Europe.

die gemeinsame Verantwortlichkeit gegebenenfalls in eine eigenständige Verantwortlichkeit der jeweiligen ausführenden Stellen über.³⁴⁴

Vorliegen und Umfang der gemeinsamen Verantwortlichkeit können im GSJ-Projekt beschränkt werden, indem zunächst die ausführenden Stellen streng weisungsgebunden als Auftragsverarbeiter die (Teil-)Anonymisierung durchführen. Hierdurch bestünde dann erst für die anschließende Verarbeitung zu Projekt- und Forschungszwecken mit den zumindest teilweise anonymisierten Daten eine gemeinsame Verantwortlichkeit, wobei die Verarbeitung in Gestalt der Bereitstellung (teil-)anonymisierter Daten an die ausführenden Stellen nach Art. 6 Abs. 1 DSGVO tendenziell eher gerechtfertigt werden kann.

2. Rechtsgrundlage nach Art. 6, 9 DSGVO für die Verarbeitung

Für die Beurteilung der Rechtmäßigkeit nach Art. 6, 9 DSGVO ergeben sich in dem hypothetischen Szenario weitere Abweichungen in der rechtlichen Beurteilung gegenüber dem Sachverhalt des GSJ-Projekts.

Wie im Sachverhalt des GSJ-Projekts kommen wegen Art. 6 Abs. 1 UAbs. 2 DSGVO berechnete Interessen nach Art. 6 Abs. 1 UAbs. 1 lit. f DSGVO nicht als Rechtsgrundlage in Betracht.³⁴⁵ Ein Rückgriff auf Art. 6 Abs. 4 DSGVO als (mögliche) eigenständige Rechtsgrundlage ist aufgrund der Neuerhebung durch die ausführenden Stellen als Verantwortliche ebenfalls ausgeschlossen.³⁴⁶

Im Fall einer gemeinsamen Verantwortlichkeit müssen sich die gemeinsam Verantwortlichen jeweils auf eine Rechtsgrundlage berufen können.³⁴⁷ Eine solche Differenzierung nach den jeweiligen Verarbeitungszwecken ist zudem in Art. 89 Abs. 4 DSGVO angelegt. Insoweit sind die Anforderungen an eine mitgliedstaatliche Rechtsvorschrift i.S.d. Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO aus Art. 6 Abs. 3 DSGVO erneut in Erinnerung zu rufen. Danach bedarf es einer Rechtsvorschrift mit festgelegtem Datenverarbeitungsbezug,³⁴⁸ die „klar und präzise und [deren] Anwendung [...] für die Rechtsunterworfenen [...] vorhersehbar“ ist (ErwGr. 41 S. 2 DSGVO).

344 Vgl. *EuGH*, ZD 2019, 455 (Rn. 70) – Fashion ID.

345 S. oben unter DV.4.a.

346 Hierzu oben unter DV.4.e.

347 *EuGH*, ZD 2019, 455 (Rn. 96 f.) – Fashion ID.

348 *Albers/Veit*, in: BeckOK Datenschutzrecht, Art. 6 DSGVO Rn. 57.

Die Anforderungen an die Konkretisierung der Aufgaben im öffentlichen Interesse i.S.d. Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO in der Rechtsvorschrift stehen in einer Wechselwirkung zur unionsautonom auszulegenden³⁴⁹ Erforderlichkeit einer Verarbeitung für die Wahrnehmung ebendieser Aufgabe. Beispielsweise ist die Verarbeitung personenbezogener Daten zu Zwecken der Forschung und Lehre grundsätzlich zulässig unter Rekurs auf allgemeine Aufgabenzuweisungen der Hochschulen i.V.m. Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO.³⁵⁰ Je gravierender sich eine Verarbeitung für die betroffenen Personen mit Blick auf die Verarbeitungszwecke, die Sensibilität der verarbeiteten Daten und den Verarbeitungsumfang darstellt, desto präziser muss die mitgliedstaatliche Rechtsvorschrift ausgestaltet sein.³⁵¹

In diesem Zusammenhang ist auch und gerade unter Beachtung der Grundsätze der Datenminimierung (Art. 5 Abs. 1 lit. c DSGVO) und Speicherbegrenzung (Art. 5 Abs. 1 lit. e DSGVO) mit Blick auf Forschungszwecke zu prüfen, ob und in welchem Umfang es eines Personenbezugs bedarf oder ob alternativ mit pseudonymisierten oder anonymisierten Daten der verfolgte Zweck gleichermaßen erreicht werden kann (vgl. auch Art. 89 Abs. 1 DSGVO).³⁵²

Das aus Art. 89 Abs. 1 DSGVO ableitbare Anonymisierungserfordernis gleicht gerade das an anderer Stelle in der DSGVO vorgesehene Forschungsprivileg aus (z.B. in Art. 5 Abs. 1 lit. b DSGVO, ErwGr. 33 DSGVO).

a. Übermittlung an die ausführenden Stellen

Gegenüber dem Sachverhalt des GSJ-Projekts ergeben sich insbesondere aus dem Vorgang der Übermittlung personenbezogener Daten an die ausführenden Stellen besondere Herausforderungen: Aufgrund der gemeinsamen Verantwortlichkeit bedarf es sowohl einer Rechtsgrundlage der Ministerien für die Offenlegung durch Übermittlung als auch einer Rechtsgrundlage der ausführenden Stellen für die Entgegennahme in Gestalt der Erhe-

349 *Buchner/Petri*, in: Kühling/Buchner, Art. 6 DSGVO Rn. 118.

350 *Golla*, in: Specht-Riemenschneider/Mantz, § 23, Rn. 83.

351 *Schulz*, in: Gola/Heckmann, Art. 6 DSGVO Rn. 57 mit Verweis auf die ständige Rechtsprechung des BVerfG; *Heberlein*, in: Ehmann/Selmayr, Art. 6 DSGVO Rn. 63; *U. Spies*, ZD 2022, 75 (77); vgl. auch *Taeger*, in: Taeger/Gabel, Art. 6 DSGVO Rn. 96; *Roßnagel*, ZD 2020, 296 (298); s. auch zu § 3 BDSG BT-Drs. 18/11325, 81; im Überblick zu großzügigeren Stimmen *Reimer*, in: Sydow/Marsch, Art. 6 DSGVO Rn. 66 m.w.N.

352 *Golla*, in: Specht-Riemenschneider/Mantz, § 23, Rn. 48.

bung dieser Daten durch die ausführenden Stellen. Während die Erhebung durch die ausführenden Stellen im Wesentlichen den gleichen Anforderungen wie die weitere Verarbeitung unterliegt (s. nachfolgend unter b.), ist an dieser Stelle vor allem die Rechtfertigung der Übermittlung aus Sicht der Ministerien gesondert zu prüfen.

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

In Betracht kommt zunächst ein Rekurs auf die Bestimmungen der Art. 6 Abs. 1 UAbs. 1 lit. c, 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, die eine Übermittlung an Dritte aufgrund einer Interessenabwägung zulassen.

Allerdings hat das BVerwG mit Blick auf die nach Art. 6 Abs. 1 UAbs. 2 DSGVO ausgeschlossene Interessenabwägung die Vereinbarkeit des Art. 5 Abs. 1 S. 1 Nr. 2 BayDSG mit den Vorgaben der DSGVO in Abrede gestellt.³⁵³ Diese zutreffende Argumentation des BVerwG dürfte auf den im Kern inhaltsgleichen § 8 Abs. 2 Nr. 2 DSG NRW übertragbar sein.

Vor diesem Hintergrund dürften diese landesrechtlichen Vorschriften nicht als eine belastbare Rechtsgrundlage herangezogen werden können für die Übermittlung an (gemeinsam) Verantwortliche, die eigene Zwecke verfolgen.

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

Zu denken ist ferner an 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, die als Generalklauseln auf eine dem Verantwortlichen übertragene Aufgabe abstellen. Als eine solche Aufgabe der Justizministerien im öffentlichen Interesse lässt sich auch die Forschung bezogen auf die Verbesserung von Arbeitsabläufen und die Entwicklung von Software für den Einsatz in der Justiz einordnen.

Wenn und soweit allerdings den ausführenden Stellen im Hinblick auf die Zwecke und den Umfang der Verarbeitung ein Gestaltungsspielraum zukommt (z.B. für die grundlegende KI-Forschung ohne spezifischen Er-

353 BVerwG, BeckRS 2018, 35172 (Rn. 25 f.); vgl. auch zu § 25 BDSG Frenzel, in: Paal/Pauly, § 25 BDSG Rn. 2.

kenntnisgewinn für die Justiz), wird die hierauf gerichtete Übermittlung personenbezogener Daten nicht erforderlich sein für die Erfüllung öffentlicher Aufgaben der *Ministerien*.

Dieser Befund wird in Ansehung der jeweils und kumulativ erforderlichen Rechtsgrundlage(n) der gemeinsam Verantwortlichen unabhängig davon gelten, ob es sich bei der Forschung um eine öffentliche Aufgabe der *ausführenden Stellen* handelt.

Vor diesem Hintergrund erklären sich die Bemühungen und Überlegungen betreffend die Schaffung eines Forschungsdatengesetzes, welches gerade auf einen erleichterten Forschungszugang zu Daten der öffentlichen Hand und entsprechende Rechtsgrundlagen zielt.³⁵⁴

Sofern die Ministerien ausschließlich mit dem Erlanger Tool vorbereitete und somit teilanonymisierte Datensätze an die ausführenden Stellen übermitteln, sinken im Rahmen der Erforderlichkeit die Anforderungen an den Funktionszusammenhang zwischen Verarbeitung und den Aufgaben im öffentlichen Interesse. Dieser Befund betrifft insbesondere Konstellationen, in denen keine sensiblen personenbezogenen Daten verarbeitet werden (z.B. im Miet- und Verkehrsrecht unter Ausschluss von Fällen, die im Rahmen der Schadensbemessung allfällige Gesundheitsangaben enthalten).

Selbst in diesen Konstellationen kann es aber potenziell zum Vorhandensein von Zehntausenden betroffenen Personen kommen, obwohl aus der Nähe zu Forschungszwecken gerade strenge Anonymisierungsanforderungen aus Art. 89 DSGVO i.V.m. landesrechtlichen Vorschriften³⁵⁵ erwachsen. Vor diesem Hintergrund bestehen nicht unerhebliche Zweifel daran, ob die Verarbeitung des Restbestands – und gegebenenfalls auch überhaupt die Verarbeitung – personenbezogener Daten zur (Grundlagen-)Forschung als erforderlich anzusehen ist für die Aufgabenerfüllung der (Justiz-)Ministerien.

cc. Art. 6 Abs. 1 UAbs. 1 lit. e, Abs. 3 DSGVO i.V.m. Hochschulgesetzen der Länder

Die Hochschulgesetze der Länder Bayern und Nordrhein-Westfalen weisen den Justizministerien zudem keine Aufgabe im öffentlichen Interesse zu, die

354 BMBF, Eckpunkte BMBF Forschungsdatengesetz; hierzu etwa *Specht-Riemenschneider/Wehde*, ZGI 2022, 3 (6); s. auch den aus der Mitte der FDP-Fraktion eingebrachten Gesetzesentwurf, der sich allerdings grundsätzlich auf registerführende Behörden und statistische Daten beschränkt BT-Drs. 20/14262.

355 Hierzu sogleich unter DVII.2.a.cc.

die Datenübermittlung zur Forschungszwecken i.V.m. Art. 4 Abs. 1 BayDSG bzw. § 3 Abs. 1 DSG NRW erfordern würde. Etwaige Aufgabenzuweisungen (z.B. § 2 Abs. 2, § 3 Abs. 1, 3 HG NRW bzw. Art. 2 Abs. 1 S. 2, Abs. 2 S. 2, 3, Art. 3 Abs. 1 S. 1, Art. 6 Abs. 1 S. 1 BayHIG) adressieren grundsätzlich nur die jeweiligen Hochschulen.

Soweit bspw. Art. 6 Abs. 1 S. 1 BayHIG darauf abstellt, dass die Hochschulen „mit dem Bund, den Ländern und anderen juristischen Personen des öffentlichen Rechts“ zur Erfüllung ihrer öffentlichen Aufgaben zusammenwirken, dürfte sich aus dieser Vorschrift zwar ein allgemeines Bekenntnis zu einer Zusammenarbeit entsprechend den vorstehenden Erwägungen zu Art. 4 Abs. 1 BayDSG bzw. § 3 Abs. 1 DSG NRW, nicht aber eine hinreichend konkrete Aufgabenzuweisung an das bayerische Justizministerium ergeben, nicht-anonymisierte oder nur teilweise anonymisierte Daten zur freien Forschung in dem untersuchungsgegenständlichen Umfang zur Verfügung zu stellen.

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

Der Rekurs auf Art. 6 Abs. 1 UAbs. 1 lit. e, Abs. 3 DSGVO i.V.m. § 17 Abs. 1 DSG NRW ermöglicht ebenfalls keine belastbar rechtssichere Übermittlung personenbezogener Daten zu wissenschaftlichen Forschungszwecken in dem hypothetischen Szenario.

Zunächst stellt § 17 Abs. 1 DSG NRW keine Anforderungen an die öffentliche Stelle, sodass die Rechtsgrundlage im Hinblick auf eine Verarbeitung für Forschungszwecke im Ausgangspunkt auch den verantwortlichen Ministerien offensteht. Die Rechtsgrundlage kann im Hinblick auf den dort festgelegten Forschungszweck somit grundsätzlich den Anforderungen der DSGVO aus Art. 6 Abs. 3 DSGVO genügen.³⁵⁶

§ 17 DSG NRW statuiert allerdings weitere Anforderungen, darunter eine Abwägung mit den schutzwürdigen Belangen der betroffenen Personen (§ 17 Abs. 1 DSG NRW), die Einhaltung der allgemeinen Garantien aus § 15 DSG NRW (i.e. technische und organisatorische Maßnahmen einschließlich der nachträglichen Überprüfbarkeit der Verarbeitung und der Sensibilisierung der an Verarbeitungsvorgängen Beteiligten) sowie im Einklang

³⁵⁶ Vgl. zu dem Verhältnis der Alternativen in Art. 6 Abs. 3 S. 2 DSGVO Heberlein, in: Ehmann/Selmayr, Art. 6 DSGVO Rn. 63.

mit Art. 89 Abs. 1 DSGVO in § 17 Abs. 3 S. 1 DSG NRW (ähnlich: Art. 25 Abs. 2 S. 1 BayDSG) eine Anonymisierungspflicht, „sobald dies nach dem Forschungs- oder Statistikzweck möglich ist“.

Entsprechend den zu den anderen Rechtsvorschriften herausgearbeiteten Maßgaben stehen die schutzwürdigen Belange betroffener Personen insbesondere einer Verarbeitung personenbezogener Daten entgegen, wenn die personenbezogenen Daten von vornherein nicht für die Forschungszwecke benötigt werden, zugleich jedoch massenhaft übermittelt werden. Zu dem gleichen Ergebnis führt die in § 17 Abs. 3 S. 1 DSG NRW vorgesehene Anonymisierungspflicht, die nicht auf die technische Möglichkeit einer automatisierten Anonymisierung abstellt³⁵⁷ (z.B. durch das Erlanger Tool), sondern vielmehr nur darauf, ob durch eine Anonymisierung der Forschungszweck beeinträchtigt wird. Eine solche Beeinträchtigung des Forschungszwecks dürfte nach dem untersuchungsgegenständlichen Sachverhalt vorliegend nicht anzunehmen sein.

Im Ergebnis erwachsen in diesem hypothetischen Szenario vor allem aus den an Forschungszwecke gestellten strengeren Maßnahmenanforderungen in Verbindung mit der Übermittlung durch die Ministerien erhebliche datenschutzrechtliche Bedenken mit Blick auf die beabsichtigte Zusammenarbeit im Rahmen des GSJ-Projekts.

Die Bewertung nach bayerischem Landesrecht ist grundsätzlich gleichlaufend (s. auch Art. 25 Abs. 2 BayDSG zum Anonymisierungserfordernis). Für die Verarbeitung personenbezogener Daten in Gerichtsentscheidungen der bayerischen Justiz fehlt es aber an einer Entsprechung zu § 17 Abs. 1 DSG NRW, der über Art. 9 Abs. 2 DSGVO hinaus funktional zugleich als Rechtsvorschrift i.S.d. Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO wirkt. Denn die bayerischen Regelungen privilegieren Forschungszwecke nur mit Blick auf die Zweckbindung (Art. 6 Abs. 2 Nr. 3 lit. c BayDSG) und die Verarbeitung besonderer Kategorien personenbezogener Daten i.S.d. Art. 9 DSGVO (Art. 8 Abs. 1 S. 1 Nr. 5 i.V.m. Art. 6 Abs. 2 Nr. 3 lit. c BayDSG). Für nicht-besondere Kategorien personenbezogener Daten kommt daher nur ein Rückgriff auf Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO, Art. 4 Abs. 1 BayDSG i.V.m. BayHIG in Betracht oder eine mit erheblichen Unsicherheiten behaftete

357 Vgl. zu § 27 BDSG *Koch*, in: BeckOK Datenschutzrecht, § 27 BDSG Rn. 39 ff.; *Hense*, in: Sydow/Marsch, § 27 BDSG Rn. 20: „Ob und inwieweit Anonymisierung unter den Vorzeichen fortschreitender Datenverarbeitungstechniken und damit verbundener Verknüpfungsmöglichkeiten mittels Algorithmen überhaupt eine Option bleibt, sei dahingestellt“; wohl auch *Petri*, in: Simitis/Hornung/Spiecker gen. Döhmnn, § 27 BDSG Rn. 14 f.; mit einem etwas weiteren Verständnis *Buchner/Tinnefeld*, in: Kühling/Buchner, § 27 BDSG Rn. 24.

dogmatische Herleitung in Form eines Erst-Recht-Schlusses der Zulässigkeit aus Art. 8 Abs. 1 S. 1 Nr. 5 i.V.m. Art. 6 Abs. 2 Nr. 3 lit. c BayDSG auch für nicht-besondere Kategorien personenbezogener Daten.

ee. Besonderheiten mit Blick auf besondere Kategorien personenbezogener Daten nach Art. 9 DSGVO

Für die Übermittlung besonderer Kategorien personenbezogener Daten in den Gerichtsentscheidungen kommen als Rechtsgrundlage zwar grundsätzlich Art. 9 Abs. 2 lit. j DSGVO i.V.m. § 17 Abs. 1 DSG NRW bzw. Art. 8 Abs. 1 S. 1 Nr. 5 i.V.m. Art. 6 Abs. 2 Nr. 3 lit. c BayDSG in Betracht.

Nach den vorherigen Ausführungen bestehen insoweit aber erst recht erhebliche Bedenken, eine umfangreiche Verarbeitung sensibler, nicht vollständig anonymisierter Daten auf eine Forschungsgeneralklausel zu stützen. Das in Art. 89 Abs. 1 DSGVO angelegte und in den landesrechtlichen Vorschriften ausgeprägte Anonymisierungserfordernis wird sich in Ansehung der sensiblen Daten in besonderem Maße auswirken.

b. Weiterverarbeitung durch die ausführenden Stellen

Soweit eine (gerechtfertigte Übermittlung) erfolgt ist, sind die Anforderungen an die Weiterverarbeitung grundsätzlich gleichlaufend mit den Anforderungen aus dem Sachverhalt des GSJ-Projekts (s. oben unter D.V), da die zur Realisierung des GSJ-Projekts notwendigen Verarbeitungsaktivitäten weiterhin zu den auch im Sachverhalt des GSJ-Projekts maßgeblichen Zwecken erfolgen.

Die ergänzende Verarbeitung zu Forschungszwecken über das GSJ-Projekt hinaus kann auf die landesrechtlichen Forschungsgeneralklauseln bzw. Aufgabenzuweisungen in den Hochschulgesetzen i.V.m. Art. 6 Abs. 1 UAbs. 1 lit. e, Art. 9 Abs. 2 lit. j DSGVO gestützt werden (z.B. Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO, Art. 4 Abs. 1 BayDSG i.V.m. Art. 2 Abs. 1 S. 2, Abs. 2 S. 2 und 3, Art. 3 Abs. 1 S. 1, Art. 6 Abs. 1 S. 1 BayHIG bzw. §§ 3, 17 DSG NRW i.V.m. § 2 Abs. 2, § 3 Abs. 1, 3 HG NRW), wenn bereits die Übermittlung dieser Daten zu Forschungszwecken zulässig war und ein ausreichender Anonymisierungsgrad erreicht wurde. Denn die ausführenden Stellen üben in diesem Fall ihre Kerntätigkeit in der Forschung aus; somit steht gerade nicht (mehr) die Frage im Raum, ob die Verfolgung von Forschungszwecken eine Aufgabe der (Justiz-)Ministerien ist.

3. Datenschutzrechtliche Anforderungen im Übrigen

Im Fall einer gemeinsamen Verantwortlichkeit zwischen den Ministerien und den ausführenden Stellen sind die weitergehenden Pflichten zu einer Vereinbarung (Art. 26 Abs. 1 S. 2, 3, Abs. 2 S. 1 DSGVO) sowie zur Bereitstellung des Wesentlichen der Vereinbarung (Art. 26 Abs. 2 S. 2 DSGVO) zu beachten.

Das gesteigerte Risiko³⁵⁸ für betroffene Personen ist im Rahmen der Datenschutzfolgenabschätzung nach Art. 35 DSGVO zu reflektieren. Diese Verpflichtung gilt umso mehr in Ansehung der Forschungsgeneralklauseln, denen gerade keine konkrete Folgenabschätzung zugrunde liegt, die den vorliegenden Fall mit seinen Besonderheiten ausreichend abdeckt.³⁵⁹

Für die zu Forschungszwecken erfolgenden Verarbeitungen sind die in Ausprägung des Art. 89 DSGVO ergänzenden Anforderungen aus dem Landesrecht zu beachten. So statuieren beispielsweise § 17 Abs. 4 DSG NRW bzw. Art. 25 Abs. 3 BayDSG zusätzliche Anforderungen an die Veröffentlichung von Forschungsdaten, wobei sich diese zusätzlichen Anforderungen gegebenenfalls auf die Veröffentlichung des Sprachmodells auswirken können.

Zudem schränkt § 17 Abs. 5 DSG NRW bzw. Art. 25 Abs. 4 BayDSG (vgl. Art. 89 Abs. 2 DSGVO) zugunsten der ausführenden Stellen Betroffenenrechte mit Blick auf die Verarbeitung zu Forschungszwecken ein.

4. Zwischenergebnis

In dem hier zugrunde gelegten hypothetischen Szenario erwachsen besondere Herausforderungen vor allem aus dem Erfordernis einer Rechtsgrundlage für die Übermittlung nicht-anonymisierter Gerichtsentscheidungen durch die Ministerien an die ausführenden Stellen. An der Möglichkeit zu einem rechtlich belastbaren Rückgriff auf die Aufgabenzuweisungen und Forschungsgeneralklauseln in den landesrechtlichen Vorschriften bestehen insoweit erhebliche Zweifel.

Dieser Befund gilt umso mehr, wenn und soweit die Gerichtsentscheidungen ohne Vorbearbeitung durch ein Anonymisierungstool und unter Einschluss im Einzelfall sensibler Gerichtsentscheidungen übermittelt werden (z.B. Gesundheitsdaten im Rahmen von Schmerzensgeld im Verkehrs-

358 T. Radtke, Gemeinsame Verantwortlichkeit unter der DSGVO, S. 411 ff.

359 Vgl. Reimer, in: Sydow/Marsch, Art. 6 DSGVO Rn. 64.

recht oder sogar über das Miet- und Verkehrsrecht hinaus im Falle von Arzthaftungsentscheidungen).

Selbst wenn eine (Teil-)Anonymisierung mit dem Erlanger Tool erfolgt, dürften die strengeren Anonymisierungsanforderungen an eine Verarbeitung zu Forschungszwecken aus Art. 89 DSGVO mit dem derzeit durch das Erlanger Tool erreichten Anonymisierungsstandard kaum eingehalten werden können.

VIII. Ableitung von datenschutzrechtlichen Handlungsempfehlungen

Aus der vorangegangenen Untersuchung lassen sich konkrete Handlungsempfehlungen für das untersuchungsgegenständliche GSJ-Projekt ableiten, die an dieser Stelle zusammengeführt werden.

1. Datenschutzrechtliche Ausgangslage in drei Stufen: Anonymisierung – Rechtfertigung nach Art. 6 DSGVO – Rechtfertigung nach Art. 9, 10 DSGVO

Mit Blick auf die Anwendbarkeit der DSGVO und die Einhaltung grundlegender datenschutzrechtlicher Anforderungen lässt sich zwischen drei, nachfolgend näher zu behandelnden Stufen unterscheiden. Demgegenüber sind andere Rechtsakte, insbesondere die JI-RL, mit Blick auf den gewählten Zuschnitt des Projekts (i.e. Ausschluss von Strafsachen) von der Vertiefung auszunehmen.

Auf erster Stufe steht die Frage des Vorliegens personenbezogener Daten i.S.d. Art. 4 Nr. 1 DSGVO und mithin einer (erfolgreichen) Anonymisierung. Ein Personenbezug setzt entweder eine eindeutige Kennung voraus oder sonstige Informationen, die nach allgemeinem Ermessen wahrscheinlich zur Identifizierung einer natürlichen Person genutzt werden. Für diese Wahrscheinlichkeit kommt es bei KI-Systemen maßgeblich auf die Ausgabe an, da auf andere Weise den justizinternen Anwendern im vorliegenden Fall kein Zugriff auf die Daten, mithin auch keine Identifizierung möglich ist. Eine (Teil-)Anonymisierung durch das Erlanger Tool kann vor diesem Hintergrund in vielen Fällen zu einer nahezu vollständigen Anonymisierung führen, indem die Informationen zunächst lediglich die Grundlage für ein wahrscheinlichkeitsbasiertes Training des KI-Systems sind und erst im Rahmen der Ausgabe eine risikobehaftete Verarbeitung erfolgen kann.

Soweit keine hinreichenden Maßnahmen zum Ausschluss des Personenbezugs ergriffen wurden, können auf zweiter Stufe ein verbleibender Personenbezug im Modell im Rahmen einer Bearbeitung (z.B. betreffend die Verarbeitung von Kennungen) sowie die vereinzelte Ausgabe von Kennungen gegebenenfalls auf Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO i.V.m. landesrechtlichen Generalklauseln Art. 4 Abs. 1 BayDSG, § 3 Abs. 1 DSG NRW oder nach einer weiteren in Betracht kommenden Auffassung auf Art. 6 Abs. 1 UAbs. 1 lit. f DSGVO oder auf Art. 6 Abs. 4 DSGVO gestützt werden, da insoweit ein enger Zusammenhang mit einer öffentlichen Aufgabe besteht und eine risikoarme justizinterne Verarbeitung erfolgt.

Auf dritter Stufe ist die Verarbeitung besonderer Kategorien personenbezogener Daten i.S.d. Art. 9, 10 DSGVO angesprochen, die nur im Einzelfall auf Rechtfertigungstatbestände, wie etwa Art. 9 Abs. 2 lit. e, j DSGVO, gestützt werden oder die Anforderungen des Art. 10 S.1 DSGVO erfüllen kann. Daher ist im Hinblick auf das Vorliegen dieser Daten besonders sorgfältig zu clustern und zu filtern.

Bei Beachtung der jeweils auf dieser Stufe zu stellenden Anforderungen ist das untersuchungsgegenständliche GSJ-Projekt grundsätzlich datenschutzrechtskonform realisierbar.

Auf eine echte Privilegierungswirkung bereits veröffentlichter Gerichtsentscheidungen können sich Verantwortliche hierbei allerdings nicht berufen. Gleichwohl wird die Weiterverarbeitung einer zulässigerweise, mit entsprechend hohem Anonymisierungsgrad veröffentlichten Gerichtsentscheidung aufgrund eines engeren Funktionszusammenhangs mit öffentlichen Aufgaben und gegebenenfalls nicht signifikant erhöhten Risiken für betroffene Personen grundsätzlich auf Art. 6 Abs. 1, 4 DSGVO, Art. 9 Abs. 2 lit. f DSGVO gestützt werden können.

2. Anforderungen an die Zusammenstellung des Trainingskorpus aus Gerichtsentscheidungen und Aktenauszügen (insbesondere Clustering)

Soweit personenbezogene Daten vorliegen, ergibt sich zunächst vor allem mit Blick auf Art. 9, 10 DSGVO ein Bedarf zur Clustering. Im Rahmen einer solchen Clustering nach Verfahrensarten können insbesondere Miet- und Verkehrssachen grundsätzlich einzuschließen, Arzthaftungssachen, Familien- und Strafsachen demgegenüber grundsätzlich auszuschließen sein.

Aus den Aktenauszügen sind unzureichend bzw. mangelhaft digitalisierte Dokumentbestandteile zu filtern und gegebenenfalls händisch nachzubearbeiten.

Außerdem sind für die Aktenauszüge entsprechende Dokumentenkategorien festzulegen mit:

- Hohem Risiko (regelmäßig vollständiger Ausschluss oder Nachbearbeitung), z.B.:
 - o Beigezogene Akten der Staatsanwaltschaft im Zusammenhang mit Verfahren, die typischerweise über die Straftat i.S.d. Art. 10 DSGVO hinaus sensible Informationen enthalten (z.B. Straftaten gegen die sexuelle Selbstbestimmung und schwere Körperverletzungsdelikte);
 - o Psychologische Gutachten;
 - o Ärztliche Gutachten.
- Mittlerem Risiko (angepasstes anonymisiertes Anonymisierungsverfahren), z.B.:
 - o Beigezogene Akten der Staatsanwaltschaft im Übrigen;
 - o Behandlungsberichte und Überblick über Behandlungsmaßnahmen, verschriebene Medikamente u.Ä.

Darüber hinaus ist nach Merkmalen innerhalb der Gerichtsentscheidungen und Aktenauszüge zu filtern. Das betrifft insbesondere:

- Geltendmachung von deliktsrechtlichen Schadensersatzansprüchen (insbesondere § 823 Abs. 2 BGB i.V.m. StGB);
- Geltendmachung von Schmerzensgeld nach § 253 Abs. 1, 2 BGB;
- Geltendmachung von Unterlassungsansprüchen mit Blick auf (politische) Aussagen, wobei gegebenenfalls eine offensichtliche Öffentlichmachung nach Art. 9 Abs. 2 lit. e DSGVO in Betracht kommen kann;
- Vorkommen von Begriffen wie „Gewerkschaft“ sowie geläufigen Gewerkschafts- und Parteibezeichnungen innerhalb der Aktenauszüge.

Soweit einzelne Textabschnitte sich immer wieder in ähnlicher Form in verschiedenen Gerichtsentscheidungen wiederfinden, ist wegen der verstärkten Repräsentation in den Trainings- und womöglich auch den Ausgabedaten in besonderem Maße darauf zu achten, dass Kennungen und sonstige identifizierende Merkmale identifiziert wurden. Das betrifft auch und gerade Konstellationen, in denen Textbausteine aus einer Gerichtsentscheidung in ähnlicher Form in andere Entscheidungen übernommen wurden.

3. Anforderungen an das Anonymisierungstool (insbesondere Recall-Wert)

Das Anonymisierungstool muss insbesondere für Kennungen einen hohen Recall-Wert erreichen. Diese Anforderung gilt gleichermaßen für Gerichtsentscheidungen und Aktenauszüge. Zwar ist selbst bei einem Recall-Wert von unter 90 % aufgrund des zwischengelagerten Trainings, der eng eingegrenzten Eingabeaufforderung im Rahmen der Use-Cases 1-3 sowie der Ausgabe eine datenschutzkonforme Umsetzung mit hohem Anonymisierungsanteil auf erster Stufe und gerechtfertigten systeminternen Verarbeitungen auf zweiter Stufe denkbar. Zugleich steigt aber mit einem niedrig(er)en Recall-Wert für Kennungen jedenfalls die Wahrscheinlichkeit von Datenschutzverstößen, z.B. mit Blick auf vermehrte und gegebenenfalls nicht mehr zu rechtfertigende Verarbeitungen in der Ausgabe oder – auf dritter Stufe – der Kombination von Kennungen mit besonderen Kategorien personenbezogener Daten.

Für sonstige identifizierende Merkmale kann unter Umständen auch noch ein Recall-Wert von deutlich unter 50 % genügen. Denn insoweit ergibt sich regelmäßig schon auf erster Stufe eine Anonymisierung dadurch, dass die Heranziehung weiterer Quellen zur Identifizierung natürlicher Personen durch justizinterne Nutzer im Fall einer unbeabsichtigten Ausgabe aus den Trainingsdaten unwahrscheinlich ist. Dieser Befund gilt vorbehaltlich der übrigen Maßnahmen, insbesondere der Clusterung mit Blick auf besondere Kategorien personenbezogener Daten.

Das in Rede stehende Anonymisierungstool, übrige Software und Prozesse sind entsprechend dem Stand der Technik auch im Hinblick auf IT-Sicherheitsanforderungen fortlaufend zu evaluieren und gegebenenfalls fortzuentwickeln.

4. Anforderungen an das KI-System und dessen Einsatz (insbesondere System Prompt und Ausgabefilter)

Das KI-System ist nur für den justizinternen Einsatz und auf ausgewählte Use Cases ausgelegt. Diese Beschränkung sollte zugleich durch technische und organisatorische Maßnahmen sichergestellt werden, z.B. durch einen möglichst sicheren Login-Mechanismus nach dem Stand der Technik sowie, nicht notwendigerweise, durch technische Einschränkungen der Eingabeaufforderung.

In den genannten Use-Cases kann es sich beispielsweise zur weiteren, nicht aber zwingend erforderlichen Risikoreduktion anbieten, dem Nutzer

die Eingabeaufforderung verbindlich vorzugeben (z.B. die Auswahl jeweils eines Prompts für die Use-Cases 1-3) und nur die Auswahl des Eingabekontexts zu ermöglichen (z.B. die Auswahl der jeweiligen E-Akte oder eines PDF-Dokuments). Eine etwaige Ausgabe personenbezogener Trainingsdaten sollte zudem auf einfachem Wege durch die Nutzer gemeldet werden können.

Ferner ist das KI-System über den initialen sog. System Prompt in geeigneter Weise anzuhalten, vor allem mit den neuen Eingaben zu arbeiten sowie die Wiedergabe von Kennungen oder sonstigen personenbezogenen Daten aus den Trainingsdaten zu vermeiden.

Besonderes Gewicht kommt vor allem effektiven Ausgabefiltern zu, die schon auf erster Stufe eine weitgehende Anonymisierung herbeiführen können. Die Ausgaben sollten auf typische Kennungsformate (z.B. Namen, E-Mail-Adressen, Telefonnummern und Identifikationsnummern) durch sog. reguläre Ausdrücke durchsucht und diese mit dem Vorkommen im Eingabekontext abgeglichen werden. Nur wenn sich diese Angaben in ähnlicher Form auch im Eingabekontext wiederfinden (z.B. eine dort verwendete Telefonnummer oder ein dort verwendeter Nachname), spricht dem konkreten Verfahren und nicht den Trainingsdaten zuzuordnen sind, sollte die Ausgabe zugelassen werden.

Eine weitergehende, aber ebenfalls nicht zwingend erforderliche Risikoreduktion kann erfolgen, indem über einfache reguläre Ausdrücke hinaus die Ausgabe anhand des Erlanger Tools klassifiziert wird und auf dieser Basis mit dem Eingabekontext abgeglichen wird.

Die KI-Ausgabe kann zudem durch ein digitales Wasserzeichen gekennzeichnet werden. Außerdem empfiehlt sich generell eine niedrige Temperatur-Einstellung, um das Auftreten datenschutzrelevanter Halluzinationen zu reduzieren. Eingaben dürfen nicht ohne weitere Bearbeitung zur Verbesserung und zum weiteren Training des bestehenden Sprachmodells verwendet werden.

5. Anforderungen an die Veröffentlichung des KI-Modells oder KI-Systems

Die öffentliche Zurverfügungstellung des KI-Systems, sprich beispielsweise eines öffentlichen abrufbaren Justiz-Chatbots, berührt vor allem die erste Stufe, i.e. die Anonymisierung. Durch den unbeschränkten, auch justizexternen Adressatenkreis steigt die Wahrscheinlichkeit einer Identifizierung anhand von KI-Ausgaben, z.B. durch Extrahierungsangriffe. Dieser Effekt

kann durch stark beschränkte Eingabemöglichkeiten (z.B. nur zum eigentlichen Eingabekontext) und effektive Ausgabenfilter eingedämmt werden, so dass in Abhängigkeit von der konkreten Umsetzung die öffentliche Zurverfügungstellung des KI-Systems datenschutzrechtlich vertretbar ausgestaltet werden können dürfte.

Bei der Veröffentlichung des trainierten Sprachmodells (d.h. insbesondere der Gewichte), gegebenenfalls einschließlich der Trainingsdaten, werden sämtliche dieser Sicherungsmechanismen jedoch grundsätzlich ausscheiden. Insoweit bestehen datenschutzrechtliche Bedenken gegen die Veröffentlichung, die zumindest auf dritter Stufe mit Blick auf Art. 9, 10 DSGVO durch eine effektive Clusterung zu einem hohen Grad ausgeräumt werden können.

Soweit allerdings ein Anonymisierungsgrad erreicht wird, der dem Standard für eine rechtskonforme Urteilsveröffentlichung i.R.d. Art. 6 Abs. 1 UAbs. 1 lit. c, e, f, Abs. 4 DSGVO entspricht (d.h. Recall-Werte von Kennungen i.H.v. annähernd 99 %), kann auch die Veröffentlichung des hierauf aufbauenden Sprachmodells unter Berücksichtigung der hierdurch nicht stärker tangierten Interessen betroffener Personen auf die Rechtsgrundlage aus Art. 6 Abs. 1 UAbs. 1 lit. c, e, f, Abs. 4 DSGVO gestützt werden. Ein solches Vorgehen ist allerdings nur für ein Sprachmodell eröffnet, das lediglich auf Grundlage der Gerichtsentscheidungen, nicht aber auch auf den Aktenauszügen trainiert ist.

6. Beachtung weiterer datenschutzrechtlicher Pflichten

Darüber hinaus sind weitere datenschutzrechtliche Pflichten zu beachten. Diese Pflichten betreffen vielgestaltige technische und organisatorische Maßnahmen, so etwa

- nach der hiesigen Auffassung eine Auftragsverarbeitungsvereinbarung zwischen den Ministerien und den ausführenden Stellen (Art. 28 DSGVO) sowie eine Vereinbarung über die gemeinsame Verantwortlichkeit zwischen den Ministerien (Art. 26 DSGVO);
- die Durchführung einer Datenschutzfolgenabschätzung (Art. 35 DSGVO);
- die Dokumentation der Zusammenstellung der Trainingsdaten bei gleichzeitiger Festlegung und Beachtung von Löschfristen;

- die fortlaufende Überprüfung des Standards des Anonymisierungstools sowie weiterer technischer und organisatorischer Maßnahmen anhand des Stands der Technik (Art. 24, 25, 32 DSGVO).

Im Hinblick auf die Betroffenenrechte ergeben sich weniger aus dem konkreten Projekt als vielmehr aus dem Einsatz eines KI-Systems verschiedene Herausforderungen.

Für die Informationspflicht kann die Ausnahme des Art. 14 Abs. 5 lit. b Hs. 1 Var. 2 DSGVO zur Anwendung gelangen, erfordert allerdings weitere Maßnahmen zum Schutz der betroffenen Personen (z.B. die öffentliche Information über das Projekt und die damit einhergehende Verarbeitung personenbezogener Daten). Auskunftsansprüche nach Art. 15 DSGVO dürften im GSJ-Projekt aufgrund der justizinternen Verarbeitung praktisch kaum eine Bedeutung erlangen und werden sich zudem auch regelmäßig nicht auf das GSJ-Sprachmodell, sondern nur auf eine etwaige Ausgabe richten, denn das Sprachmodell beinhaltet regelmäßig keine personenbezogenen Daten.

Die Richtigkeit der Trainingsdaten, z.B. im Zusammenhang mit einem Berichtigungsbegehren nach Art. 16 DSGVO, wird in Ansehung der einzunehmenden Perspektive im Projekt selten in Frage stehen. Denn es wird sich regelmäßig um richtige Daten als Niederlegung des gerichtlichen Verfahrensablaufs und der rechtlichen Würdigung des Gerichts handeln. Das Risiko des Auftretens etwaiger Halluzinationen kann durch technische und organisatorische Maßnahmen (z.B. die Temperatur-Einstellung des KI-Systems) reduziert werden.

Eine Löschung, z.B. im Zusammenhang mit einem entsprechenden Begehren nach Art. 17, 21 DSGVO, dürfte sich vor allem auf die Ergebnisse der Use-Cases als Sachverhaltsabschnitt einer neuen Gerichtsentscheidung richten und ist insoweit einer Einzelfallprüfung zu unterwerfen.

