Anlage: Stellungnahme zum Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, V. 1.0

1. Sachverhalt

Künstliche Intelligenz (KI) umfasst verschiedene Technologien/ Methoden, die alle das Ziel haben, sich intelligent zu verhalten[1]. Auf welchen Sachverhalt sich die Behörde genau stützt, erläutert sie nicht im Detail. Für die datenschutzrechtliche Bewertung ist ein genaues Verständnis der Technologie jedoch besonders wichtig, da die unterschiedlichen KI-Systeme und -Modelle vielschichtig sind und teils große Unterschiede aufweisen, so dass man sich zunächst die genaue Funktionsweise von KI vor Augen führen sollte. Hilfreich für den KI-Entwickler wären an dieser Stelle konkretere Angaben zu den KI-Systemen und -Modellen, die die Behörde ihrer Ausarbeitung zu Grunde gelegt hat, und deren konkreten Funktionsweisen, um den KI-Entwickler in die Lage zu versetzen, zu prüfen, ob die Ausführungen auf sein konkretes KI-System übertragbar sind. Wir stützen unsere weiteren Ausführungen auf die Annahme, dass Ausgangspunkt der behördlichen Ausführungen im Wesentlichen ein abgeschlossen trainiertes KI-Modell ist, das auf der Machine Learning-Methode des Deep Learnings auf Grundlage eines tiefen neuronalen Netzes basiert, bei dem die KI aus Trainingsdaten lernt, das Gelernte selbständig überträgt und bei der Beantwortung von Anfragen an die KI (auch Prompts/ Arbeitsanweisungen/ Input genannt) anwendet. Im Folgenden erläutern wir zunächst die Rahmenbedingungen bezogen auf dieses KI-Modell.

a) Funktionsweise von KI

Beim Machine Learning über tiefe neuronale Netze (sog. Deep Learning) wird die Funktionalität des menschlichen Gehirns mit Hilfe von lernfähigen Algorithmen nachgeahmt. KI-Modelle können zB Large Language Models sein (sog. LLM wie bspw. ChatGPT) für Textauswertung und -erstellung oder multimodale Modelle für die Text- und Bildverarbeitung sowie Deep-Learning-Text/Bild-zu-Bild Generatoren für Bilder. LLM auf Basis tiefer neuronaler Netze sind – vergleichbar mit dem menschlichen Gehirn – künstliche neuronale Netze, die dazu in der Lage sind, Texte selbständig zu analysieren und Prompts zu beantworten.

b) Trainingsphase

Zunächst muss das KI-Modell mit Informationen versorgt werden. In der Entwicklungsphase werden daher zunächst meist große Datenmengen aus internen oder externen Quellen wie dem Internet (sog. Trainingsdaten) für das Training des KI-Modells verwendet. Die Wörter der Trainingsdaten werden vom KI-Modell ersetzt durch feststehende Zahlen (sog. Tokens). Ein Algorithmus analysiert anschließend die Trainingsdaten, gleicht sie selbständig ab, erkennt hieraus Muster in den Wort- und Buchstabenkombinationen und gewichtet sie. Die Qualität des ausgegebenen Ergebnisses (Output) ist dabei abhängig von der Menge und Qualität der eingegebenen Trainingsdaten.[2] Bspw. prüft der Algorithmus, welche Buchstaben und Worte besonders häufig relevant und wahrscheinlich sind und welche Worte oft

[1] Bitkom, Machine Learning und die Transparenzanforderungen der DS-GVO, 2018, S. 6

[2] BSI, Große KI-Sprachmodelle – Chancen und Risiken für Industrie und Behörden, 2021, S. 15f.; Hoeren/Sieber/Holznagel, Handbuch Multimedia-Recht/Willeke, 59. EL Juni 2023, Teil 29.3 Rn. 6; Begleitforschung Mittelstand-Digital WIK GmbH, Künstliche Intelligenz im Mittelstand Relevanz, Anwendungen, Transfer, S. 10, online abrufbar unter: https://www.mittelstand-digital.de/MD/Redaktion/DE/Publikationen/kuenstliche-intelligenz-im-mittelstand.pdf?__blob=publicationFile&v=5; Alzubaidi et al., A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. J Big Data 10, 46 (2023), S. 13f; Keber/Maslewski, RDV 2023, 273, (274)

aufeinanderfolgen und überträgt diese Muster und inhaltliche Informationen in die Parameter des LLM[3]. Der Algorithmus lernt selbstständig dazu und kann eigene Vorhersagen treffen; er sagt das nächstbeste Wort voraus, wobei die erlernten Muster Grundlage dafür sind.[4] Der Algorithmus sagt das jeweils nächste Wort auf Basis der gesamten Sequenz (zB im Satz, Absatz, Text) voraus.[5] Dieses Training führt dazu, dass sich die künstlichen Neuronen immer weiter vernetzen, so dass ein tiefes neuronales Netz mit mehreren Schichten (hidden layer) entsteht (sog. Blackbox-KI, da für den Benutzer die Funktionsweise und Rechenoperationen der KI nicht sichtbar und nicht nachvollziehbar sind[6]). Dabei stellt ein Neuron ein Wort und die Verbindung zwischen zwei Neuronen die Wahrscheinlichkeit dar, dass diese Wörter aufeinanderfolgen (sog. Gewichtung).[7] KI-Modelle wie neuronale Netzwerke sind darauf ausgelegt, die erlernten Muster und Strukturen zu speichern, die das Modell dazu befähigen, Texte zu „verstehen" und generieren zu können. Es geht nicht um die Wiedergabe der konkreten Trainingsdaten, sondern um die Erstellung eines neuen Textes, der auf Basis der ausgewerteten statistischen Wortzusammenhänge dem am nächsten kommt, was dem Anfragenden wahrscheinlich am nützlichsten ist.[8] Ein LLM ist daher keine Suchmaschine und kein Nachschlagewerk. Die Blackbox kann nicht ausgelesen werden und man kann auch nicht mehr bzw. nur schwer feststellen, welche Trainingsdaten in die Modell-Parameter eingeflossen sind[9]. Davon unterscheidet sich eine Datenbank.[10] Eine Datenbank gibt als Output die Informationen aus, die konkret in ihr gespeichert sind. Hier basiert der Output auf dem Input bzw. den tatsächlich gespeicherten Daten. Man kann in die Datenbank hineinschauen und darin gespeicherte Daten verändern oder löschen. Eine KI hingegen liefert einen Output entweder, weil die Information binär in ihren Parametern gespeichert ist, weil er das Ergebnis einer Wahrscheinlichkeitsberechnung ist, oder die KI denkt sich eine Information aus. Hier ist unklar, was die KI als Output ausgibt, das ist nicht vorhersehbar. Der Output entspricht nicht zwangsläufig dem Input, geschweige denn den tatsächlich in der KI gespeicherten Informationen. Auch wenn die KI keine Informationen zu einer Person gespeichert hat, kreiert sie zu dieser einen Output. Sie gibt jedoch nicht an, wie ihr Output entstanden ist .

c) Prompten und Output

Wenn nun eine Anfrage oder Arbeitsanweisung an die KI gestellt wird, bewertet der Algorithmus den Prompt und sagt im Rahmen der Erstellung des Outputs voraus, welche Antwort auf den Prompt wahrscheinlich nützlich ist. Auf diese Weise wird eine Antwort generiert, die auf den erlernten Mustern und Wahrscheinlichkeiten beruht. Das ausgeworfene Ergebnis ist nicht unbedingt das Richtige[11], sondern das Wahrscheinlichste bzw. vermeintlich Nützlichste im Sinne einer „bestmöglichen Vermutung"[12]. So werden teilweise Antworten generiert, die inhaltlich falsch sind (sog. Halluzinationen).[13] Ein treffenderes Ergebnis kann erzielt werden, indem man dem Algorithmus einen Kontext bereitstellt (zB genauere

[3] Pesch/Böhme, MMR 2023, 917, (918)
[4] Bitkom, Große Sprachmodelle – Ein Überblick, 2023, S. 4
[5] Seemann, Künstliche Intelligenz, Large Language Models, ChatGPT und die Arbeitswelt der Zukunft, Forschungsförderung Working Paper, Hans-Böckler-Stiftung, Nummer 304, September 2003, S. 11
[6] Bitkom, Blick in die Blackbox – Nachvollziehbarkeit von KI-Algorithmen in der Praxis, 2019, S. 8; Datatilsynet, Artificial intelligence and privacy, 2018, S. 12
[7] Bitkom, Große Sprachmodelle – Ein Überblick, 2023, S. 7
[8] Pesch/Böhme, MMR 2023, 917 (919)
[9] Carlini et al., Proceeding of the 30th USENIX Security Symposium 2021, 2633; Pesch/Böhme, MMR 2023, 917 (919)
[10] A.A. Pesch/Böhme, MMR 2023, 917 (921)
[11] Siehe dazu die Beispiele von ChatGPT und Bard bei Pesch/Böhme, MMR 2023, 917 (919 Fn. 33)
[12] Baumgartner/Brunnbauer/Cross, MMR 2023, 543, (546)
[13] BSI, Große KI-Sprachmodelle – Chancen und Risiken für Industrie und Behörden, 2021, S. 10

Angaben über den Zusammenhang des Prompts oder bestimmte Annahmen, die er zu Grunde legen soll), sobald aber weitere Inhalte außerhalb des Kontexts herangezogen werden, wird der Wahrheitsgehalt der Ausgabe wieder verfälscht. Trotzdem kann es auf Grund verschiedener Umstände vorkommen, dass der Output gleich den Trainingsdaten ist, wenn diese zB im Fall übermäßiger Anpassung (sog. Overfitting) bei Trainings mit den gleichen Daten „erinnert" werden (sog. Memorization), oder aber bei größeren Modellen, die für das Erinnern anfälliger sind.[14] Mit gezielten Abfragen und Attacken können teilweise Daten abgerufen werden, z. B. Namen, Telefonnummern und Social Media Accounts.[15] Durch Sicherheitsmechanismen wie Filter wird zwar versucht, dass bestimmte Inhalte (zB vertrauliche Informationen oder personenbezogene Daten) geblockt und nicht in den Output aufgenommen werden[16], sichere Maßnahmen zur Vermeidung dieser Outputs existieren bislang jedoch noch nicht.[17]

[14] Carlini et al., Proceeding of the 30th USENIX Security Symposium 2021, S. 2633 ff.; Nasr, Carlini et al., Extracting Training Data from ChatGPT, 2023
[15] Carlini et al., Proceeding of the 30th USENIX Security Symposium 2021, S. 2633 (2642)
[16] BSI, Große KI-Sprachmodelle – Chancen und Risiken für Industrie und Behörden, 2021, S. 14f.
[17] Carlini et al., Proceeding of the 30th USENIX Security Symposium 2021, S. 2633 (2644)

2. Zu Kapitel II. Personenbezogene Daten und der Einsatz von Künstlicher Intelligenz

Die zentrale Frage für die Anwendung des Datenschutzrechts ist, ob personenbezogene Daten betroffen sind. Im Zusammenhang mit der Bewertung von Verarbeitungsvorgängen durch eine KI können an unterschiedlichen Stellen personenbezogene Daten verarbeitet werden. Chronologisch orientiert an den einzelnen Verarbeitungsschritten kann ein Personenbezug bei folgenden Daten diskutiert werden: Trainingsdaten (1.), Betriebsdaten im KI-Modell (2.) bzw. das KI-Modell selbst (3.), Eingabedaten im Prompt (4.) und Daten im Output (5.). Diese Daten werden in den folgenden Ausführungen genauer betrachtet. Außerdem können sich personenbezogene Daten zu den Benutzern der KI bspw. in den Metadaten oder Logfiles des KI-Systems sowie den Benutzeraccounts befinden. Hier gibt es bezogen auf ein KI-System grundsätzlich keine Besonderheiten im Vergleich zu ihrer Verarbeitung in anderen Systemen, daher bedarf es an dieser Stelle keiner weiteren Betrachtung. Die Behörde greift an dieser Stelle zu kurz. Sie macht keine differenzierten Ausführungen zum Personenbezug der einzelnen in Betracht kommenden Daten in den unterschiedlichen Verarbeitungsstadien der KI, zwischen denen sie in ihrem Diskussionspapier differenziert[18]. Jedoch gerade auch in Bezug auf den Output ist die Klärung des Personenbezugs wichtig. Eine spezifischere Betrachtung wäre hier wünschenswert.

Außerdem stellt die Aufsichtsbehörde bei der Beurteilung des Personenbezugs darauf ab, ob und inwieweit ein abgeschlossen trainiertes KI-Modell die Identifizierbarkeit aktuell und in Zukunft zulässt.[19] Im weiteren Verlauf ihres Papieres trifft sie jedoch nicht nur Aussagen zu den Datenverarbeitungen in einem trainierten Modell, sondern auch zur vorgelagerten Phase der Verarbeitung der Trainingsdaten, ohne jedoch zum Personenbezug von Trainingsdaten Ausführungen zu machen.[20] Auch bei der Bewertung der Rechtsgrundlagen sind die Ausführungen nicht trennscharf an einem abgeschlossen trainierten KI-Modell orientiert. Hier ist eine genauere Auseinandersetzung mit dem Personenbezug dieser Daten angezeigt.

a) Personenbezug von Trainingsdaten

In den Trainingsdaten können auch personenbezogene Daten enthalten sein.[21] Dies wird vor allem dann angenommen werden können, wenn Quellen aus dem Internet zum Training herangezogen werden. Wird ein LLM zB mit Büchern oder Zeitungsartikeln und den personenbezogenen Daten darin trainiert, so werden auch die personenbezogenen Daten der Autoren verarbeitet.[22]

b) Personenbezug von Betriebsdaten im KI-Modell

Ein Personenbezug sei nach Ansicht des LfDI BW möglich bei im KI-Modell selbst enthaltenen personenbezogenen Daten.[23] Diese Aussage begründet die Behörde jedoch nicht weiter. Genau darauf kommt es hier aber an. Insbesondere werden hierzu auch gegenteilige Auffassungen vertreten, nämlich, dass im KI-Modell selbst keine personenbezogenen Daten gespeichert sind.[24] So führt bspw. die

[18] LfDI BW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, V. 1.0, S. 7ff.
[19] LfDI BW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, V. 1.0, S. 6
[20] LfDI BW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, V. 1.0, Ziffer III. 1. und 2., Seite 7f.
[21] Marx, Sütthoff, KI und Datenschutz: Zur Reichweite der Löschungspflicht des Verantwortlichen, Recht & Verwaltung 2022, Kapitel II.
[22] Pesch/Böhme, MMR 2003, 917 (919)
[23] LfDI BW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, V. 1.0, S. 6
[24] Franke, RDi 2023, 565

dänische Aufsichtsbehörde aus, dass das KI-Modell kein personenbezogenes Datum, sondern nur das aggregierte Ergebnis einer Datenverarbeitung darstelle.[25] Im KI-Modell werden die ursprünglichen Trainingstexte grundsätzlich nicht gespeichert und daher auch nicht mehr ausgeworfen. Allein aus der Tatsache, dass sich Teile der Trainingsdaten in den Parametern des KI-Modells niedergeschlagen haben, die auf einen Prompt wieder ausgegeben werden könnten, lässt sich kein Personenbezug des Modells begründen.[26] Eine tatsächliche Identifizierung von Personen durch KI-Modelle ist aufgrund der Komplexität und Abstraktheit der in den Modellen gespeicherten Informationen sehr schwer. Gleichwohl kann es vorkommen, dass der Output Teile der Trainingsdaten beinhaltet, wenn sie nach Berechnung des Algorithmus die wahrscheinlichste nützlichste Buchstaben-/ Wortkombination darstellen. Modelle könnten während des Trainings auch zufällige Muster oder sogar durch Overfitting lernen und spezifische Trainingsdaten zu stark berücksichtigen, so dass sie einzelne personenbezogene Daten erinnern können.[27] Hierzu könnte der LfDI daher noch weitergehende Ausführungen machen.

c) Personenbezug des KI-Modells/ Mittelbare Identifizierbarkeit durch Attacken auf das KI-Modell

Bei der Frage der mittelbaren Identifizierbarkeit überlegt die Behörde, ob durch Attacken auf das KI-Modell herausgefunden werden könne, ob sich unter den Trainingsdaten solche mit Personenbezug befunden haben, oder ob aus den Lernergebnissen des Modells ein Rückschluss auf die Trainingsdaten gezogen werden könne.[28] Dann könne das KI-Modell selbst als personenbezogenes Datum angesehen werden.[29] Diese Aussage wird jedoch nur in Ansätzen begründet. Verschiedentlich wurde herausgefunden, dass mit Hilfe von Angriffen auf das KI-Modell personenbezogene Daten im Output hervorgebracht werden können.[30]

Bei der Beurteilung, ob es sich um personenbezogene Daten handelt, kommt es nur auf die Speicherung von Informationen an. Unerheblich ist, ob die Daten auf ihrem Speicherort lesbar und sichtbar sind oder nicht. Dafür könnte man zB anbringen, dass die Datenverarbeitung in diesen Fällen vergleichbar mit einer Verschlüsselung ist. Verschlüsselte Daten haben dann Personenbezug, wenn sie pseudonymisiert sind, also u. a. eine Rückschlüsselung unter Hinzuziehung zusätzlicher Informationen noch möglich ist (Art. 4 Nr. 5 DS-GVO[31], ErwG 16). Nur wenn keine Rückschlüsselung mehr oder sie nur mit einem unverhältnismäßigen Aufwand möglich ist, man also von einer Anonymisierung sprechen kann, dann liegt kein Personenbezug vor. Das KI-Modell speichert grundsätzlich keine Worte, sondern übersetzt diese in Zahlen. Die Blackbox der KI kann nicht (wie zB eine SQL-Datenbank) ausgelesen werden[32]. Wenn personenbezogene Trainingsdaten mit Hilfe von Attacken aber aus dem KI-Modell ausgelesen werden können, so stellen die im Rahmen dieser Attacke ergriffenen Rückschlüsselungsmethoden eine Art Schlüssel wie bei einer klassischen Verschlüsselung dar. Ergänzend muss dann noch die Frage beantwortet werden, ob eine solche Attacke noch einen verhältnismäßigen Aufwand darstellt, um die

[25] Datatilsynet DK, Offentlige myndigheders brug af kunstig intelligens, Oktober 2023, S. 7
[26] Pesch/Böhme, MMR 2023, 917 (920)
[27] Carlini et al., Proceeding of the 30th USENIX Security Symposium 2021, S. 2633 ff.
[28] LfDI BW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, V. 1.0, S. 6, 7
[29] LfDI BW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, V. 1.0, S. 7
[30] LfDI BW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, V. 1.0, S. 6f. m. w. N; Carlini et al., Proceeding of the 32th USENIX Security Symposium 2023, S. 5253 ff.; Carlini et al., Proceeding of the 30th USENIX Security Symposium 2021, S. 2633 ff.
[31] Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung)
[32] Siehe dazu oben unter Sachverhalt

Trainingsdaten sichtbar zu machen. Sollte dies nicht der Fall sein, könnten die Daten ggf. sogar anonymisiert sein. Die Ausführungen der Behörde greifen hier leider nur sehr kurz; eine weitere Auseinandersetzung mit den Rechtsfragen wäre hilfreich.

Auch werden zum Personenbezug von KI-Modellen im Zusammenhang mit dem Angriff auf KI-Modelle andere Ansichten vertreten.[33] Ein Angriff auf die KI könne zwar eine Verletzung des Schutzes personenbezogener Daten darstellen, wenn er zur Reidentifizierung natürlicher Personen führe; das Risiko einer Reidentifizierung führe aber nicht zur Bejahung des Personenbezugs des KI-Modells.[34] Auch aus dem Algorithmus, aus dem die KI besteht, lässt sich grundsätzlich kein Personenbezug herleiten.[35] Die KI stellt allenfalls ein Mittel für die Herstellung eines Personenbezug im Sinne von ErwG 26 S. 3 DS-GVO dar.[36] Außerdem muss nach allgemeinem Ermessen ein Rückbezug der Trainingsdaten auf eine natürliche Person wahrscheinlich sein. Angriffe auf KI-Modelle sind zwar theoretisch möglich, in der Praxis aber oft auf spezifische Szenarien und Modelle beschränkt. Zumindest aktuell ist die Eintrittswahrscheinlichkeit von Angriffen und die Erfolgswahrscheinlichkeit der Extraktion von Trainingsdaten aus der KI noch sehr gering.[37]

Gleichwohl ist nicht zu verkennen, dass ein Prompt mit einer Frage nach Informationen zu einer bestimmten Person dazu führen kann, dass vom KI-Modell aus den Trainingsdaten ggf. erinnerte Informationen zu dieser Person ausgegeben werden, welche dann (ggf. mit Hilfe von Suchmaschinen) identifiziert werden kann.[38] Immer dann, wenn KI-Modelle mit personenbezogenen Daten trainiert wurden, kann nicht ausgeschlossen werden, dass auch personenbezogene Daten verarbeitet werden.[39]

Für unterschiedliche Szenarien kann die Frage des Personenbezugs unterschiedlich zu beantworten sein. Es muss daher im Einzelfall geprüft werden, ob die ausgegebenen Daten tatsächlich einen Personenbezug aufweisen. Die Klärung der Frage des Personenbezugs ist essenziell für die Beantwortung der weiteren Rechtsfragen; eine eingehendere Betrachtung durch die Behörde ist insofern wünschenswert.

d) Personenbezug der Eingabedaten im Prompt

Es können sich personenbezogene Daten in den Prompts befinden, die dann von der KI weiterverarbeitet werden. Da die Prompts bei statistischen KI-Modellen jedoch nicht dazu genutzt werden, das Modell weiter zu trainieren, ergeben sich für den Ausgangsfall keine Besonderheiten bei der Beurteilung der Datenschutzkonformität.

e) Personenbezug von Outputs

Es stellt sich hier die Frage, ob fiktive Daten auch personenbezogen im Sinne der DSGVO sind. Für die Beurteilung des Personenbezugs an sich ist es unerheblich, ob die Daten, die einer natürlichen Person zugeordnet werden, richtig oder falsch sind, solange die Person, der man sie zuordnet, identifizierbar ist. Auch wenn direkt identifizierende Informationen nicht vorhanden sind, kann u.U. aus der Summe

[33] Datatilsynet DK, Offentlige myndigheders brug af kunstig intelligens, Oktober 2023, S. 7; Franke, RDi 2023, 565 (566); Marx, Sütthoff, KI und Datenschutz: Zur Reichweite der Löschungspflicht des Verantwortlichen, Recht & Verwaltung 2022, Kapitel III. 2

[34] Datatilsynet DK, Offentlige myndigheders brug af kunstig intelligens, Oktober 2023, S. 7

[35] Marx, Sütthoff, KI und Datenschutz; Kapitel III. 2.: Zur Reichweite der Löschungspflicht des Verantwortlichen, Recht & Verwaltung 2022t

[36] Franke, RDi 2023, 565 (566); Marx, Sütthoff, KI und Datenschutz: Zur Reichweite der Löschungspflicht des Verantwortlichen, Recht & Verwaltung 2022, Kapitel III. 2

[37] Franke, RDi 2023, (565 f.)

[38] Pesch/Böhme, MMR 2003, 917 (920 f.)

[39] Hoeren/Sieber/Holznagel, Handbuch Multimedia-Recht/Willeke, 59. EL Juni 2023, Teil 29.3 Rn. 8

der „anonymen“ Angaben eindeutig auf eine konkrete Person geschlossen werden (wie wenn zB der Autor einer Biografie Personen aus seinem Umfeld nicht konkret benennt, sie aber so beschreibt, dass sie auf Grundlage der Menge dieser beschreibenden Informationen identifiziert werden können). Hier wird auch die Frage zu beantworten sein, worauf bei der Bewertung des Personenbezugs des Outputs abzustellen ist: auf das Verständnis des Anbieters der KI, ob er die KI mit Daten zu einer realen Person trainiert hat (im vorgenannten Beispiel der Autor), oder auf objektive Kriterien, ob Informationen zu einer realen Person in der KI als Datenbasis gespeichert sind (im Beispiel bei objektiver Möglichkeit der Rückbeziehung auf eine natürliche Person), oder auf den Empfängerhorizont, ob der Anwender der KI aus den Merkmalen auf eine reale Person schließen kann (im Beispiel der Leser).

Was im Output an Informationen ausgeworfen wird, ist nur ein geringer Teil dessen, was in der Trainingsphase an Trainingsdaten im KI-Modell verarbeitet wurde. Allein die Tatsache, dass im Output zB Namen enthalten sein können, reicht nicht aus, um einen Personenbezug dieser Daten anzunehmen. Denn es ist unklar, in welchem Kontext die Daten im Output ursprünglich standen, als sie als Trainingsdaten in die KI eingespielt wurden. Dieser Zusammenhang, in dem die personenbezogenen Trainingsdaten in den Trainingstexten standen, ist aber für die Qualifikation von Daten als personenbezogen besonders wichtig.[40] Auch wenn im Output ein Datum wie zB ein Name enthalten ist, der auf eine natürliche Person hinweisen könnte, muss es sich nicht um eine konkrete natürliche Person handeln, da die Namenskombination vielleicht einfach nur die wahrscheinlichste war, die der Algorithmus berechnet hat. Wenn die KI zB mit Texten trainiert wurde, in denen häufig der Name Michael Müller enthalten war, kann es sein, dass sie im Output diese Vor-/ Nachnamens-Kombination wieder ausgibt. Es heißt aber nicht, dass dieses Datum auch noch auf eine konkrete Person hinweist, da es aus dem ursprünglichen Kontext entfernt wurde und ohne diese Zusatzinformationen nicht mehr auf eine konkrete Person zurückgeschlüsselt werden kann. Es können daher zusätzliche Informationen notwendig sein, um die im Output angezeigten Daten einer konkreten natürlichen Person zuzuordnen. Auch ist nicht klar, ob es sich bei den ausgegebenen personenbezogenen Daten tatsächlich um solche existierender natürlicher Personen handelt, oder ob auch hier auf Grund der Wahrscheinlichkeitsberechnung ein fiktives Datum vom Algorithmus generiert wurde.

f) Anonyme Daten, die zu personenbezogenen Daten werden

Der LfDI wirft in der Kurz-Checkliste die Frage als Prüfungspunkt auf, ob „anonyme Daten verarbeitet werden, die zu personenbezogenen Daten werden können".[41] Diesbezüglich wäre es hilfreich, wenn die Behörde konkretisieren könnte, an welchen Anwendungsfall sie hier gedacht hat. Denn grundsätzlich wird man davon ausgehen können, wenn anonymisierte Trainingsdaten verwendet werden, dass auch die Datenverarbeitung in der KI sowie auch der Output keine personenbezogenen Daten enthalten werden. Mit Zusatzinformationen aus weiteren Informationsquellen und mit Hilfe anderer technischer Mittel können die Informationen jedoch verschnitten werden, was dazu führen kann, dass auf diese Weise ein Personenbezug hergestellt wird.

3. Grundsatz der Datenrichtigkeit

Den Grundsatz der Datenrichtigkeit nach Art. 5 Abs. 1 lit. d DS-GVO klammert die Behörde ganz aus, wenngleich er gerade bei der Bewertung von statischen, abschließend trainierten KI-Modellen eine

[40] Art.-29-Datenschutzgruppe, Stellungnahme 4/2007 zum Begriff „personenbezogene Daten“, WP 136, S. 15 allgemein zur Bestimmung des Personenbezugs anhand von Kontextinformationen

[41] LfDI BW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, V. 1.0, S. 27

wichtige Rolle spielt. Dieser Punkt wird auch im Rahmen der Erforderlichkeitsprüfung bei der Bewertung der Rechtsgrundlagen relevant.

Nach diesem Grundsatz müssen Daten sachlich richtig und erforderlichenfalls auf dem neuesten Stand sein; dh mit der Realität übereinstimmen[42]. Am Grundsatz der Datenrichtigkeit zu messen sind daher vor allem Tatsachenbehauptungen, die dem Beweis zugänglich sind.[43] Es können aber auch Schätzungen oder auf Wahrscheinlichkeiten beruhende Ergebnisse relevant werden, wenn nicht ausdrücklich kenntlich gemacht wird, dass sie nicht auf validen Tatsachen beruhen.[44] Hier ist zu fragen, ob diese Ergebnisse auf mathematisch-statistisch exakten Methoden der Datenverarbeitungen basieren. Falls nicht, kann der Grundsatz der Datenrichtigkeit dann verletzt sein.[45] Per se problematisch ist bei KI-Modellen, die auf Wahrscheinlichkeitsbewertungen basieren, dass die ausgegebenen Daten vielfach nicht (ganz) richtig sein werden, da die Ergebnisse der KI eben nicht auf Fakten, sondern auf Wahrscheinlichkeiten beruhen[46], und der Anwender dies nicht weiß bzw. auch nicht richtig einschätzen kann. Selbst wenn die KI angibt, dass ein bestimmtes Zitat von einem konkreten Dichter stammt, kann diese Aussage erfunden sein, damit nicht den Tatsachen entsprechen und unrichtig sein.

Die Frage ist, ob man den Grad der Validität der Daten am jeweiligen intendierten Verarbeitungszweck orientieren kann. Für die Auswertung von Lebensläufen durch eine KI im Rahmen eines Bewerbungsverfahrens auf ihre Übereinstimmung mit dem Stellenprofil ist es wichtig, besonders valide Daten zu generieren. Hier kommt es darauf an, dass die Aussagen, die die KI trifft, auch möglichst genau und richtig sind, um auf dieser Basis eine Entscheidung treffen zu können. Für die Erstellung von Nutzerprofilen für werbliche Maßnahmen reichen jedoch weniger sichere Fakten dahingehend, dass der Empfänger Interesse an der Werbung haben könnte. Wenn die KI neue kreative Texte wie Geschichten oder Werbeanzeigen generieren soll, kommt es ggf. auch gar nicht auf einen Wahrheitsgehalt an. Die Anforderungen an den Grad der Richtigkeit können daher unterschiedlich sein; um diese zu bestimmen, muss auf den konkreten Verarbeitungszweck abgestellt werden.

Der Grundsatz der Datenrichtigkeit fordert allgemein, dass Maßnahmen ergriffen werden müssen, die sicherstellen, dass die Daten in jedem Teil der Prozesskette richtig sind und so verarbeitet werden, dass sich nicht verfälscht werden, zB Halluzinationen vermieden werden. So ist bei der Auswahl und Menge der Trainingsdaten darauf zu achten, dass diese eine gute Qualität aufweisen und richtig sind[47]: wenn bereits der Input von schlechter Qualität ist, dann ist es auch der Output.

Bei einem abschließend trainierten KI-Modell handelt es sich um ein statisches Modell, welches bis zu einem bestimmten Zeitpunkt trainiert wurde, aber neue Daten nach diesem Trainingsdatum nicht mehr berücksichtigt. Das kann dazu führen, dass es veraltet und die gespeicherten Muster nicht mehr aktuell sind, so dass auch die Ergebnisse nicht mehr richtig sind und sie daher ungenau werden. Es sollte sichergestellt werden, dass der Algorithmus richtig funktioniert und die Arbeitsergebnisse und Vorhersagen richtig sind[48]. Daher muss in regelmäßigen Abständen geprüft werden, ob das Modell neu

[42] Taeger/Gabel, DSGVO/Voigt, 4. Aufl. 2022, Art. 5 Rn. 31; Kühling/Buchner, DSGVO/Herbst, 4. Aufl.2024, Art. 5 Rn. 60
[43] Werry, MMR 2023, 911 (914)
[44] Baumgartner/Brunnbauer/Cross, MMR 2023, 543 (546); Pesch/Böhme, MMR 2003, 917 (921): hergeleitet aus der englischen und französischen Fassung der DS-GVO
[45] Pesch/Böhme, MMR 2003, 917 (921): hergeleitet aus der englischen und französischen Fassung der DS-GVO; Hoeren, ZD 2016, 459 (462)
[46] Pesch/Böhme, MMR 2003 (917 f.)
[47] Baumgartner/Brunnbauer/Cross, MMR 2023, 543 (546)
[48] Taeger/Gabel, DSGVO/Voigt, 4. Aufl. 2022, Art. 5 Rn. 31

trainiert werden muss.[49] Zudem muss gewährleistet werden, dass die Verarbeitung der Daten in der KI die Richtigkeit der Daten nicht beeinträchtigt. In diesem Zusammenhang erhält auch der Grundsatz der Transparenz eine besondere Bedeutung: damit der Anwender weiß, ob der Output auf sicheren Fakten beruht oder nicht, müsste die KI kenntlich machen, ob eine Information auf einer binären Kenntnis basiert, ob sie mit einer bestimmten Methodik eine Wahrscheinlichkeit berechnet oder sich etwas ausgedacht hat.

Wird festgestellt, dass Daten unrichtig sind, müssen sie unverzüglich gelöscht oder berichtigt werden, Art. 5 Abs. 1 lit. d Hs. 2 DS-GVO, Art. 16, 17 DS-GVO. Zur Erfüllung dieser Anforderung hat der Verantwortliche angemessene Maßnahmen zu ergreifen. Auch wenn verschiedene KI-Systeme bereits die Meldung von Fehlern durch den Benutzer ermöglichen, kommt ein KI-Modell hier an seine Grenzen. Da nicht selektiv auf einzelne im Modell gespeicherte Daten zugegriffen werden kann, können falsche Daten auch nicht „einfach" gelöscht oder berichtigt werden.[50] Falsche Daten können zwar so herausgefiltert werden, dass sie einem Benutzer nicht mehr angezeigt werden. Dies erfüllt aber nicht die Anforderungen der DS-GVO an eine Löschung oder Berichtigung. Es könnte allenfalls das gesamte KI-Modell mit frischen und sachlich richtigen Daten neu trainiert werden, ohne hierbei die falschen Daten wieder im Training zu nutzen. Insofern ist jedoch auch noch zu klären, ob sich der Anspruch auf Löschung auch auf die Früchte einer Verarbeitung bezieht, dh auf ein trainiertes KI-Modell.[51]

Die Behörde führt im Zusammenhang mit dem Löschen die Methode des Unlearnings (dh man lässt das KI-Modell bestimmte Muster und Parameter verlernen[52]) an[53]. Ob dieses Verfahren überhaupt geeignet ist und die Anforderungen von Art. 5 Abs. 1 lit. d Hs. 2 DS-GVO, Art. 16, 17 DS-GVO dadurch erfüllt werden, führt sie jedoch nicht näher aus.

[49] HmbBfDI, Checkliste zum Einsatz LLM-basierter Chatbots, 13.11.2023, S. 4; Datatilsynet DK, Offentlige myndigheders brug af kunstig intelligens, Oktober 2023, S. 7

[50] Siehe dazu zB die Aussage von OpenAI, Privacy Policy, 14.11.2023, Ziffer 4., "*Given the technical complexity of how our models work, we may not be able to correct the inaccuracy in every instance.*".

[51] Siehe dazu ablehnend Marx, Sütthoff, KI und Datenschutz: Zur Reichweite der Löschungspflicht des Verantwortlichen, Recht & Verwaltung 2022, Kapitel IV

[52] Pesch/Böhme, MMR 2003, 917 (922)

[53] LfDI BW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, V. 1.0, S. 7

4. Zu Kapitel IV: Datenschutzrechtliche Verantwortlichkeit

Die Frage der Verantwortlichkeit ist entscheidend, um den oder die Normadressaten der DS-GVO bestimmen zu können. Richtigerweise stellt der LfDI fest, dass es in Bezug auf Künstliche Intelligenz sowohl alleinige und gemeinsame Verantwortliche geben kann, als auch die Konstellation der Auftragsverarbeitung. Bei dem Einsatz von Künstlicher Intelligenz sind üblicherweise mehrere Parteien in unterschiedlichen Phasen in jeweils unterschiedlicher Weise beteiligt, so dass der Frage der Verantwortlichkeit unserer Einschätzung nach in Bezug auf diese Phasen eine besondere Praxisrelevanz zukommt. Wir schlagen daher vor, hierzu noch ausdifferenzierter zwischen den unterschiedlichen Phasen der Verarbeitung zu unterscheiden und die Verantwortlichkeit der Parteien in den jeweiligen Phasen genauer zu betrachten.

Besonders relevant ist dabei unserer Einschätzung nach die Phase der Nutzung von KI-Systemen, wenn einerseits die KI-Systeme personenbezogene Daten beinhalten,[54] und anderseits der Anwender bei der Nutzung des KI-Systems personenbezogene Daten verarbeitet, indem er diese zB in einem Prompt eingibt. Aufgrund der in der Rechtsprechung und bei Aufsichtsbehörden weit gefassten Definition gemeinsamer Verantwortung, bei der schon konvergierende Entscheidungen zu einer gemeinsamen Verantwortlichkeit führen können,[55] ist es hilfreich, insbesondere für diese Fälle eine Abgrenzung der Verantwortlichkeiten vorzunehmen. Nach unserer Einschätzung kann es in diesen Fällen nicht schon dadurch zu einer gemeinsamen Verantwortung kommen, dass ein Anwender ein vom Anbieter bereitgestelltes KI-System durch die Eingabe personenbezogener Daten nutzt, in dem ggf. personenbezogene Daten enthalten sind (zB ein Unternehmen, dass mit einem KI-Model eine Einschätzung zu einem Bewerber erstellen lässt, oder KI-Model nutz, um Kundenkorrespondenz zu beantworten). In diesen Fällen wird es bei der Nutzung derzeit schon angebotener generativer KI (zB LLM) von den großen Anbietern (beispielsweise ChatGPT oder Google Bard) an einem wechselseitigen „entscheidenden Einfluss“ fehlen.[56] Darüber hinaus müsste die wechselseitige Beeinflussung auch gerade von den beteiligten Parteien bezweckt sein.[57] Auch an einer solchen Zweckbestimmung wird es regelmäßig fehlen es, wenn ein Anwender zB ein bereitgestelltes generatives KI-Modell nutzt. Von einer gemeinsamen Verantwortung in diesem Sinne kann unserer Einschätzung nach nur dann ausgegangen werden, wenn die Interaktion des Anwenders (zumindest indirekt) das vom Anbieter bereitgestellte KI-System im Sinne einer konvergenten Entscheidung i.) beeinflusst und ii.) der Anwender diese Beeinflussung auch gerade bezweckt (zB wenn der Anwender personenbezogene Daten nicht nur für seine eigenen Zwecke eingibt, sondern er beabsichtigt, dass auf der Grundlage dieser Daten, für seine der Zwecke weitere Funktionalitäten in dem KI-System bereitgestellt oder verbessert werden).

Außerdem halten wir die Konstellation, in der mehrere Verantwortliche zwar gemeinsam beteiligt sind, aber als jeweils eigene Verantwortliche tätig sind (zB ein Verantwortlicher, der eine direkte End-Kundenbeziehung unterhält, stellt einem anderen Verantwortlichen, der eine Künstliche Intelligenz trainieren möchte, Daten zur Verfügung, ohne an der Entwicklung selbst mittel- oder unmittelbar beteiligt zu sein), für praxisrelevant, da es in diesen Fällen ggf. zu Datenübermittlungen kommt.

[54] S.o. 2. b)
[55] EDSA-Leitlinie 07/2020 Version 2.0 Rn. 54; Kühling/Buchner, DS-GVO/Hartung, 4. Aufl. 2024, Art. 26 Rn. 16 ff.
[56] Vgl. EDSA-Leitlinie 07/2020 Version 2.0 Rn. 54
[57] EDSA-Leitlinie 07/2020 Version 2.0 Rn. 62

Zur Veranschaulichung hierzu eine tabellarische Übersicht (Tabelle 1):

Phasen der Verarbeitung	Alleinige Verantwortung einer Partei	Alleinige Verantwortung mehrerer Parteien (C2C)	Gemeinsame Verantwortung	Auftragsverarbeitung (C2P)
Erhebung von Trainingsdaten	Die Stelle, die die KI trainiert, erhebt eigene personenbezogene Daten für das Training.	Eine oder mehrere Stellen übermitteln personenbezogene Daten, für deren Verarbeitung sie verantwortlich sind, an einen Dritten, damit dieser eigenständig (ohne Einfluss der übermittelnden Stellen) eine eigene KI trainiert. Jede beteiligte Partei bestimmt selbst die Zwecke und Mittel. zB ein Unternehmen erhebt personenbezogene Kundendaten und stellt diese (ggf. pseudonymisiert) einem Anbieter von KI-Diensten zur Verfügung, damit der Anbieter ein KI-Modell trainieren kann.	Mehrere Verantwortliche legen gemeinsam Zwecke und Mittel bei der Erhebung von Trainingsdaten fest. zB mehrere Verantwortliche erheben gemeinsam Kundendaten, um damit ein KI-Modell zu trainieren.	Eine Stelle erhebt im Auftrag und insbesondere weisungsgebunden für einen oder mehrere Verantwortliche Trainingsdaten.
Verarbeitung von Daten für das Training von KI	Die Stelle, die die KI trainiert, agiert in eigener Verantwortung und bestimmt Mittel und Zwecke der Verarbeitung personenbezogener Daten selbst (und bedient sich dabei ggf. eines Auftragsverarbeiters). zB ein Anbieter von KI möchte ein KI-Modell erstellen, das er selbst am Markt seinen eigenen Kunden anbietet.		Mehrere Stellen bestimmen gemeinsam Mittel und Zwecke des Trainings des KI-Modells. zB ein Anbieter von KI möchte ein KI-Modell erstellen, dass er selbst am Markt seinen Kunden anbietet, verwendet dafür Daten eines anderen Verantwortlichen und berücksichtigt im Gegenzug dessen besondere Anforderungen/ Zwecke beim Training.	Eine Stelle führt das Training von KI im Auftrag und insbesondere weisegebunden für einen oder mehrere Verantwortliche durch. zB ein IT-Dienstleister, der für seinen Auftraggeber auf Basis dessen Daten ein KI-Modell trainiert.
Bereitstellung von Anwendungen der KI[58]	Eine Stelle stellt seinen Kunden ein KI-Modell zur Verfügung, das – so wie es bereitgestellt wird – genutzt wird. zB ein Anbieter von KI stellt seinen Kunden einen eigenen ChatBot auf Basis eines KI-Modells zur Verfügung, oder ein Unternehmen setzt ein selbst entwickeltes KI-Model für eigene Zwecke ein.	Eine Stelle übermittelt ein KI-Modell an einen anderen Verantwortlichen, der dies eigenständig nutzt. zB eine Stelle erwirbt von einem Anbieter von KI eine Software, die ein KI-Modell enthält, in dem personenbezogene Daten enthalten sind und setzt dies unabhängig vom Anbieter von KI ein.	Mehrere Stellen bestimmen gemeinsam Mittel und Zwecke des bereitgestellten des KI-Modells. zB eine Stelle stellt ein KI-Modell bereit, eine andere Stelle kann durch Einstellung von Parametern bestimmen, wie dieses KI-Modell genutzt werden kann (zB Modifikation / Konfiguration des Prompts einen LLM)	Eine Stelle stellt im Auftrag und insbesondere weisungsgebunden für einen oder mehrere Verantwortliche eine KI-Anwendung bereit. zB ein IT-Dienstleister, der für seinen Auftraggeber weisungsgebunden ein KI-Model bereitstellt.
Nutzung von Anwendungen der KI – bei KI-Systemen, die keine	Eine Stelle nutzt ein bereitgestelltes KI-Modell ausschließlich für eigene Zwecke und bestimmt die Mittel der Verarbeitung selbst.	Eine Stelle nutzt eine KI-Anwendung, die ein Dritter (Anbieter von KI) bereitstellt, dabei legt jede Partei selbst die Zwecke und Mittel der Datenverarbeitung fest.	Mehrere Stellen bestimmen gemeinsam Mittel und Zwecke der bei der Nutzung eines KI-Modells anfallenden Daten. zB ein Unternehmen nutzt ein von einem Anbieter von KI bereitgestelltes KI-Modell für	Eine Stelle nutzt ein KI-Modell, dass eine andere Stelle in deren Auftrag und insbesondere weisungsgebunden bereitstellt.

[58]Nur relevant, soweit mit der Bereitstellung der KI-Anwendung eine Verarbeitung von personenbezogenen Daten einhergeht. Vgl. LfDI BW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz, V. 1.0, S. 3; s.o. Nr. 2. b).

Phasen der Verarbeitung	Alleinige Verantwortung einer Partei	Alleinige Verantwortung mehrerer Parteien (C2C)	Gemeinsame Verantwortung	Auftragsverarbeitung (C2P)
personenbezogene Daten beinhalten[59]	zB ein Unternehmen setzt ein selbst entwickeltes KI-Modell im Kundenservice ein.	zB ein Verantwortlicher nutzt ein KI-Modell (einen ChatBot), den ein anderes Unternehmen bereitstellt, indem er personenbezogene Daten eingibt.	Kundenservice, der Anbieter von KI generiert dabei Erkenntnisse und erhebt personenbezogene Daten, um das KI-Modell für seine Kunden zu verbessern und berücksichtigt im Gegenzug die besonderen Anforderungen des nutzenden Unternehmens.	zB Nutzung eines KI-Models, das ein IT-Dienstleister, der für seinen Auftraggeber weisungsgebunden tätig ist, bereitstellt.
Nutzung von Anwendungen der KI – bei KI-Systemen, die personenbezogene Daten beinhalten können[60]		Anwender und Anbieter von Anwendungen der KI sind eigenständige Verantwortliche, wenn die Entscheidungen des jeweils anderen keinen wesentlichen Einfluss auf die Verarbeitung personenbezogener Daten in dem KI-System haben. zB Nutzung eines von einem Anbieter von KI-Modellen bereitgestellten LLM, ohne dass die konkrete Nutzung Einfluss auf das Modell und die Verarbeitung personenbezogener Daten hat.	Anwender und Anbieter von Anwendungen der KI sind gemeinsam verantwortlich, wenn sie gemeinsamen Einfluss auf die Verarbeitung personenbezogener Daten in dem KI-System haben. zB Der Anbieter eines KI-Systems passt die Verarbeitung von personenbezogenen Daten in dem KI-System an die Nutzung und konkreten Abfragen des Anwenders an, damit für diesen noch bessere Ergebnisse erzielt werden.	
Nutzung von Ergebnissen der KI	Keine Besonderheiten durch die Nutzung von KI			

[59] S.o. 2. b)
[60] S.o. 2. b)

5. Zu Kapitel V (Rechtsgrundlagen für öffentliche und nicht-öffentliche Stellen) und VI (Rechtsgrundlagen für nicht-öffentliche Stellen).

In den Kapiteln V und VI macht der LfDI Ausführungen zu den einzelnen möglichen Rechtsgrundlagen für die mit dem Betrieb einer KI-Anwendung erfolgenden Datenverarbeitungen. Die Behörde behandelt Rechtsgrundlagen sowohl für öffentliche als auch nicht-öffentliche Stellen in Kapitel V und ausschließlich für nicht-öffentliche Stellen in Kapitel VI.[61]

Der LfDI unterscheidet in seinen Ausführungen zu den Rechtsgrundlagen nicht stringent zwischen den einzelnen Verarbeitungsphasen. Vorweg ist festzustellen, dass es förderlich wäre, die Rechtsgrundlagen anhand der in Kapitel III des Diskussionspapiers herausgearbeiteten Phasen zu bewerten, da es diesbezüglich erhebliche praxisrelevante Unterschiede gibt. ZB kommt die Einholung einer Einwilligung für die derzeit am Markt angebotenen generativen KI-Modelle, für deren Training umfangreiche Daten aus dem Internet verwendet werden, praktisch nicht in Betracht. Denn es ist nicht möglich, von allen betroffenen Personen Einwilligungen einzuholen. Dagegen ist es bei der Nutzung von KI-Modellen durch einen Anwender im Einzelfall sehr wohl möglich, eine Einwilligung als Rechtsgrundlage zu verwenden. Der LfDI unterscheidet in seinen Ausführungen zu den Rechtsgrundlagen nicht stringent zwischen den einzelnen Verarbeitungsphasen. Auch insofern wäre eine klare Differenzierung wünschenswert.

Als Übersicht haben wir eine vereinfachte tabellarische Darstellung zur Veranschaulichung der Besonderheiten bei der Anwendung der Rechtsgrundlagen orientiert an den einzelnen Phasen der Verarbeitung eingefügt (Tabelle 2). Im Folgenden gehen wir außerdem auf die einzelnen, vom LfDI aufgeworfenen Fragen zu den einzelnen Rechtsgrundlagen ein. Bei unsere Ausführungen beschränken wir uns auf die Rechtsgrundlagen, die für nicht-öffentliche Stellen in Betracht kommen.

[61] LfDI BW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz V. 1.0, S. 11f. und S. 16 f.

Sehr vereinfachte Darstellung zur Veranschaulichung der besonderen Herausforderungen bei der Anwendung von Rechtsgrundlagen (Tabelle 2):

	Fokus auf Anbieter der KI			Fokus auf Anwender der KI	
Rechtsgrundlagen	Erhebung von Trainingsdaten	Verarbeitung von Daten für das Training von KI	Bereitstellung von Anwendungen der KI, die personenbezogene Daten beinhaltet,[62]	Nutzung von Anwendungen der KI – bei KI-Systemen, die personenbezogene Daten beinhalten	Nutzung von Anwendungen der KI – bei KI-Systemen, die keine personenbezogene Daten beinhalten
				In Bezug auf personenbezogenen Daten in der KI	In Bezug auf personenbezogene Eingabedaten
Art. 6 Abs. 1 UAbs. 1 lit a DS-GVO	Einwilligung bezieht sich auf die Erhebung der Daten für Trainingszwecke Hervorzuhebende Herausforderung: Informiertheit darüber, was mit den Daten in der KI im Rahmen des Trainings geschieht	Einwilligung bezieht sich auf die Verarbeitung der Daten für das Training Hervorzuhebende Herausforderung: Informiertheit darüber, was mit den Daten in der KI im Rahmen des Trainings geschieht	Einwilligung bezieht sich auf die Bereitstellung der KI, in der die Daten verarbeitet werden Hervorzuhebende Herausforderungen: Löschung bei Widerruf kann eine Herausforderung sein Informiertheit über die Folgen der Bereitstellung von Daten in einer KI	Anwender müsste sich auf eine Einwilligung der betroffenen Person stützen, die die verarbeiteten Daten bereitgestellt hat und die sich auf seine konkrete Verwendung bezieht	Keine Besonderheiten, da KI nur ein besonderes Mittel der Datenverarbeitung ist, worauf es nicht ankommt (Technologieneutralität der DS-GVO).[63] Ausnahme, wenn Nutzung der KI selbst durch Einwilligung legitimiert werden soll: zB Abfrage von Informationen eines Bewerbers aus einer KI Einwilligung bezieht sich auf die Verarbeitung der Daten im Rahmen der Nutzung der KI Hervorzuhebende Herausforderung und Voraussetzung ist die Geeignetheit der Verarbeitung
	Nur in besonderen Konstellation praktisch denkbar, zB wenn die Stelle, die Daten für das Training für ein KI-Modell erhebt und/oder verwendet, mit allen betroffenen Personen in direktem Kontakt steht (zB bei medizinischen Forschungsprojekten denkbar). Kommt in vielen relevanten Fällen nicht in Betracht, zB bei der Erhebung von Trainingsdaten im Internet			Praktisch nur möglich, wenn der Anwender auch die Trainingsdaten bereitstellt und/oder die KI selbst erstellt	Praktisch relevant, zB ein Kunde hat in die Verarbeitung seiner Daten für Werbezwecke eingewilligt; die Verwendung von KI muss nicht zwingend Bestandteil der Einwilligung sein

[62] S.o. 2. b)

[63] Das gilt in der Annahme, dass KI als Mittel eingesetzt wird, um den eigentlichen Zweck der Rechtgrundlage (zB der Einwilligung) zu erreichen (zB bezieht sich die Einwilligung darauf, dass die betroffene Person mit der Verwendung von Daten für Zwecke der Direktwerbung einverstanden ist, KI wird als Mittel der Steuerung von Kampagnen anstelle anderer „herkömmlicher" Datenverarbeitungssysteme eingesetzt). Das gilt nur, soweit bei der Nutzung nicht zugleich zB auch Daten für das Training erhoben und/oder verwendet werden, dafür sind dann zusätzliche Rechtsgrundlagen für die Zwecke des Trainings erforderlich. Auch der Grundsatz der Datensparsamkeit ist hierbei zu berücksichtigen, vgl. EuGH C-252/21 Rn. 109.

	Fokus auf Anbieter der KI			Fokus auf Anwender der KI	
Rechtsgrundlagen	Erhebung von Trainingsdaten	Verarbeitung von Daten für das Training von KI	Bereitstellung von Anwendungen der KI, die personenbezogene Daten beinhaltet,[62]	Nutzung von Anwendungen der KI – bei KI-Systemen, die personenbezogene Daten beinhalten	Nutzung von Anwendungen der KI – bei KI-Systemen, die keine personenbezogene Daten beinhalten
				In Bezug auf personenbezogenen Daten in der KI	In Bezug auf personenbezogene Eingabedaten
Art. 6 Abs. 1 UAbs. 1 lit b DS-GVO	Erhebung der Trainingsdaten muss objektiv erforderlicher Bestandteil des Vertrages sein	Verarbeitung der Trainingsdaten muss objektiv erforderlicher Bestandteil des Vertrages sein	Bereitstellung der personenbezogenen Daten im KI-Modell muss objektiv erforderlicher Bestandteil des Vertrages sein	Die Nutzung der personenbezogenen Daten, die im KI-Modell enthalten sind, müsste Bestandteil eines Vertrags mit dem Anwender sein	Keine Besonderheiten, da KI nur ein besonderes Mittel der Datenverarbeitung ist, worauf es nicht ankommt (Technologieneutralität der DS-GVO) Herausforderung und Voraussetzung ist die Geeignetheit der Verarbeitung
	Nur möglich, wenn der Anbieter der KI einen Vertrag mit der betroffenen Person hat, zB Nutzung eines Services verbunden mit der Vereinbarung, dass personenbezogene Daten für das Training der KI erhoben und verarbeitet werden			Praktisch nur möglich, wenn der Anwender Verträge mit den betroffenen Personen hat und selbst die die Trainingsdaten bereitstellt und/oder die KI selbst erstellt	Praktisch relevant, zB ein Verantwortlicher setzt KI sein, um vertragsrelevante Anfragen eines Kunden zu beantworten, ohne dass die Daten dabei auch zum Training des KI-Modells genutzt werden
Art. 6 Abs. 1 UAbs. 1 lit c DS-GVO	Gesetzliche Verpflichtung muss sich auf die Erhebung von Trainingsdaten beziehen	Gesetzliche Verpflichtung muss sich auf die Verarbeitung von Trainingsdaten beziehen	Gesetzliche Verpflichtung muss sich auf die Bereitstellung von Anwendungen der KI beziehen	Gesetzliche Verpflichtung muss sich gerade auf die Nutzung von personenbezogenen Daten, die im KI-Modell enthalten sind, beziehen	Keine Besonderheiten, da KI nur ein besonderes Mittel der Datenverarbeitung ist, worauf es nicht ankommt (Technologieneutralität der DS-GVO) Herausforderung und Voraussetzung ist die Geeignetheit der Verarbeitung
	Auch für die Erfüllung gesetzlicher Pflichten (zB der Daseinsfürsorge durch öffentliche Stellen oder Umsetzung gesetzlicher Pflichten durch nicht-öffentliche Stellen) kann die Verwendung moderner Datenverarbeitungssysteme einschließlich KI erforderlich sein				Praktisch relevant, zB ein Verantwortlicher setzt KI ein, um gesetzliche Aufgaben zu erfüllen, ohne dass die Daten dabei auch zum Training des KI-Modells genutzt werden
Art. 6 Abs. 1 UAbs. 1 lit d DS-GVO	Im Regelfall schwierig aufgrund der Eilbedürftigkeit einer Notsituation	Im Regelfall schwierig aufgrund der Eilbedürftigkeit einer Notsituation	Kommt in der Regel als Rechtsgrundlage nicht in Betracht	Keine Besonderheiten, da KI nur ein besonderes Mittel der Datenverarbeitung ist, worauf es nicht ankommt (Technologieneutralität der DS-GVO) Herausforderung und Voraussetzung ist die Geeignetheit	Keine Besonderheiten, da KI nur ein besonderes Mittel der Datenverarbeitung ist, worauf es nicht ankommt (Technologieneutralität der DS-GVO) Herausforderung und Voraussetzung ist die Geeignetheit der Verarbeitung

<table>
<tr><td></td><td colspan="3">Fokus auf Anbieter der KI</td><td colspan="2">Fokus auf Anwender der KI</td></tr>
<tr><td rowspan="2">Rechtsgrundlagen</td><td rowspan="2">Erhebung von Trainingsdaten</td><td rowspan="2">Verarbeitung von Daten für das Training von KI</td><td rowspan="2">Bereitstellung von Anwendungen der KI, die personenbezogene Daten beinhaltet,[62]</td><td>Nutzung von Anwendungen der KI – bei KI-Systemen, die personenbezogene Daten beinhalten</td><td>Nutzung von Anwendungen der KI – bei KI-Systemen, die keine personenbezogene Daten beinhalten</td></tr>
<tr><td>In Bezug auf personenbezogenen Daten in der KI</td><td colspan="2">In Bezug auf personenbezogene Eingabedaten</td></tr>
<tr><td rowspan="2">Art. 6 Abs. 1 UAbs. 1 lit f DS-GVO</td><td>Hervorzuhebende Herausforderung: Intervenierbarkeit - Umsetzung des Widerspruchs – zB Etablierung einer Kennzeichnung wie „robots.txt“ für Suchmaschinen-Crawler[64]</td><td>Hervorzuhebende Herausforderung: Intervenierbarkeit - Umsetzung des Widerspruchs
Vernünftige Erwartung der betroffenen Person – insbesondere, wenn diese keine direkte Verbindung zum Anbieter der KI hat</td><td>Hervorzuhebende Herausforderung: Intervenierbarkeit - Umsetzung des Widerspruchs und Löschung
Vernünftige Erwartung der betroffenen Person – insbesondere, wenn diese keine direkte Verbindung zum Anbieter der KI hat</td><td>Hervorzuhebende Herausforderung liegt in der Interessenabwägung im engeren Sinne. Aufgrund der vielen Einsatzmöglichkeiten durch sehr viele unterschiedliche Anwender sind die vernünftigen Erwartung der betroffenen Person problematisch – insbesondere, wenn diese keine direkte Verbindung zum Anwender der KI hat</td><td colspan="2">Keine Besonderheiten, da KI nur ein besonderes Mittel der Datenverarbeitung ist, worauf es nicht ankommt (Technologieneutralität der DS-GVO)

Herausforderung und Voraussetzung ist die Geeignetheit</td></tr>
<tr><td colspan="3">Sehr praxisrelevant, in vielen Fällen (zB bei der Erhebung von personenbezogenen Daten aus dem Internet) die einzig in Betracht kommende Rechtsgrundlage; nicht anwendbar auf die Verarbeitung besonderer Kategorien personenbezogener Daten</td><td colspan="2">Praktisch relevant, wenn ein Verantwortlicher KI einsetzt, um Zwecke auf Grundlage der Interessenabwägung zu erreichen, ohne dass die Daten dabei auch zum Training des KI-Modells genutzt werden, zB um Direktwerbung auf Grundlage der Interessenabwägung durchzuführen und mittels eines KI-Modells zu steuern</td></tr>
<tr><td>§ 26 BDSG</td><td colspan="5">s.o. Ausführungen zu Art. 6 Abs. 1 UAbs. 1 lit b DS-GVO</td></tr>
</table>

[64] Zur Umsetzung bei OpenAI: Welser, GRUR-Prax 2023, 516 (519)

a) Art. 6 Abs. 1 UAbs. 1 lit. f DS-GVO

Im Rahmen der Ausführung zu Art. 6 Abs. 1 UAbs. 1 lit. f DS-GVO geht der LfDI auf die drei Tatbestandmerkmale berechtigtes Interesse (Stufe 1), Erforderlichkeit (Stufe 2) und Abwägung der Interessen (Stufe 3) ein.[65]

aa) Erforderlichkeit

Im Zusammenhang mit den Ausführungen zum Tatbestandsmerkmal der „Erforderlichkeit" führt der LfDI aus, dass die Verarbeitung der Anwendung von KI-Systemen mit anderen Verfahren bereits auf der Ebene der Erforderlichkeit verglichen werden muss. Mit seinem Beispiel für die Anwendungsphase der KI zu Art. 6 Abs. 1 UAbs. 1 lit. f DS-GVO unterstreicht der LfDI seine Annahme, nämlich: „Bedarf es des konkreten Verfahrens überhaupt oder können die verfolgten Zwecke nicht auch auf sonstige Weise erreicht werden? (So wäre bspw. für einen einfachen Blick vor die Tür mittels einer Klingelkamera eine KI-basierte Gesichtserkennung nicht erforderlich.)"[66]

Hierzu ist anzumerken, dass das Beispiel in mehrfacher Hinsicht Probleme aufweist und somit ungeeignet ist. Zudem wird in den oben genannten Textpassagen das Kriterium der Geeignetheit als Element der Prüfung der Erforderlichkeit des Art. 6 Abs. 1 UAbs. 1 lit. f DS-GVO zu wenig betont.

(a) Kritik an dem oben genannten Beispiel (Anwendungsphase):

Mit dem Beispiel wird zunächst kein schlüssiger Fall geschildert. Der Zweck der beispielhaften Verarbeitung hätte genauer definiert sein müssen und wird unpräzise als ein „Einfaches-vor-die-Tür-Schauen" beschrieben. Gemeint ist wohl ein Nachprüfen, welche Person sich vor der Tür befindet. Wenn man davon ausgeht, dass bei einer KI-Anwendung zur Gesichtserkennung mehr Daten verarbeitet werden als beim bloßen Benutzen einer Videoklingel, dann ist der Funktionsumfang der KI-Anwendung umfangreicher. Eine KI-Anwendung ist dazu in der Lage die erfassten Personen mit einer Gesichtsdatenbank abzugleichen und somit den Anwender stärker zu unterstützen.

Auch unter der Beachtung der Systematik der DS-GVO ist das Beispiel als Referenzfall nicht gut geeignet. Das Beispiel wird im Zusammenhang mit der Rechtsgrundlage der Interessenabwägung geschildert. Es wird im Rahmen der Verhältnismäßigkeitsprüfung eine Klingelkamera mit einer KI-Anwendung zur Gesichtserkennung verglichen. Bei einer KI zur Gesichtserkennung ist von einer Verarbeitung biometrischer Daten im Sinne des Art. 9 Abs. 1 DS-GVO auszugehen. Art. 6 Abs. 1 UAbs. 1 lit. f DS-GVO kann nach dem Wortlaut des Art. 9 Abs. 1 DS-GVO gerade keine Verarbeitung von besonderen Kategorien personenbezogener Daten rechtfertigen.[67] Somit ist das Beispiel im Rahmen von Art. 6 Abs. 1 UAbs. 1 lit. f DS-GVO ungeeignet.

Für die Praxis wäre es hilfreich, wenn Beispiele aufgeführt würden, die ausführlicher sind und bei denen eine bessere Vergleichbarkeit der gegenübergestellten Alternativen gegeben ist. Besonders praxisrelevant dürften für das Training einer KI die Erhebung und Verarbeitung von Daten aus dem Internet oder bei der Anwendung von KI-Systemen eine Verwendung von KI-

[65] LfDI BW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz V. 1.0, S. 16f.
[66] LfDI BW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz V. 1.0, S. 18f.
[67] Taeger/Gabel, DS-GVO/Taeger, 4. Aufl. 2022, Art. 6 Rn. 123

Modellen zur Durchführung von Direktmarketing jeweils auf Basis einer Interessenabwägung sein.

(b) Plädoyer für eine angemessene Auslegung des Kriteriums der Erforderlichkeit:

Die Erforderlichkeit für die Verarbeitung personenbezogener Daten, sowohl in der Trainings- als auch der Anwendungsphase liegt vor, wenn „das berechtigte Interesse an der Verarbeitung der Daten nicht in zumutbarer Weise ebenso wirksam mit anderen Mitteln erreicht werden kann, die weniger stark in die Grundrechte und Grundfreiheiten der betroffenen Personen“ eingreifen[68]. Daraus ist abzuleiten, dass eine Prüfung der Geeignetheit und der Wirksamkeit unerlässlich ist.

Die Geeignetheit umfasst, dass die Verarbeitung den Zweck auch erfüllen kann. Eine wesentliche Herausforderung bei der Feststellung der Erforderlichkeit in der Anwendungsphase von KI besteht unserer Einschätzung nach vor allem im Nachweis der Geeignetheit. Die Geeignetheit kann zB in Frage stehen, wenn die Richtigkeit der Ergebnisse der Verarbeitung nicht festgestellt werden kann. Dies muss ggf. bei der Einführung von KI-Systemen berücksichtigt werden, zB durch eine Verifizierung der Ergebnisse. Kann dadurch (oder durch andere Maßnahmen) die Geeignetheit nicht hergestellt werden, ist die Verarbeitung im Ergebnis nicht erforderlich. Ist die Verarbeitung von personenbezogenen Daten durch ein KI-System jedoch ausreichend wirksam zur Erreichung des Zwecks, bestehen unserer Ansicht nach bei der Nutzung von KI-Systemen keine besonderen Anforderungen in Bezug auf die Festlegung der Rechtsgrundlage im Vergleich zu anderen automatisierten Datenverarbeitungen.[69] Die Geeignetheit ist in der Anwendungsphase von KI zB nicht gegeben, wenn eine LLM-gestützte KI-Anwendung für die automatisierte Beantwortung von Kunden-E-Mails aufgrund falscher Antworten nicht für den gewollten Zweck eingesetzt werden kann. In diesem Fall ist im Rahmen der Erforderlichkeitsprüfung unter dem Aspekt der fehlenden Geeignetheit eine Rechtfertigung gem. Art. 6 Abs. 1 UAbs. 1 lit. f DS-GVO zu verneinen.

Darüber hinaus muss die Wirksamkeit einer Verarbeitung im Rahmen der Erforderlichkeitsprüfung in allen Phasen betrachtet werden.[70] Die Wirksamkeit umfasst den Grad der Effektivität zur Zweckerreichung. Sie ist eher weit auszulegen, sodass die Wirtschaftlichkeit und Effizienz des Mittels als Faktoren angeführt werden können.[71] Ließe man diese Faktoren außer Acht, so würde der technische Fortschritt stark beschränkt werden. Bei Nichtbeachtung müsste man die Papierakte gegenüber der digitalen Akte, das Festnetztelefon gegenüber dem Videokonferenzdienst als auch den Brief gegenüber der E-Mail bevorzugen. Eine solche Auslegung wäre auch nicht mit der Technologieneutralität der DS-GVO und dem Recht auf unternehmerische Freiheit aus Art. 16 GRCh zu vereinen.[72] Eine enge Auslegung der Erforderlichkeit und der berechtigten Interessen, die eine Effizienzsteigerung und die Wirtschaftlichkeit für den Verantwortlichen

[68] EuGH Urt. v. 4.7.23 – C-252/21, Rn. 108 – Meta Platforms
[69] Dies gilt für KI-Systeme, die im Modell keine personenbezogene Daten enthalten. S.o. 2. b)
[70] So auch der LfDI selbst Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz V. 1.0, S. 17; EuGH Urt. v. 4.7.23 – C-252/21, Rn. 108 – Meta Platforms
[71] Taeger/Gabel, DS-GVO/Taeger, 4. Aufl. 2022, Art. 6 Rn. 57; a.A. Kühling/Buchner, DS-GVO/Buchner/Petri, 4. Aufl. 2024, Art. 6 Rn. 147c
[72] Zur Notwendigkeit der Abwägung von Art. 8 GRCh mit Art. 16 GRCh Kühling/Buchner, DS-GVO/Kühling/Raab, 4. Aufl. 2024, A. Einführung Rn. 31; zur Notwendigkeit von Abwägung zu anderen Grundrechten allgemein ErwG 4 S. 1 DS-GVO

nicht für die Erforderlichkeit genügen lässt[73], ist auch nicht geboten, um den Betroffenen vor der Unterminierung seiner Rechte zu schützen. Die Belange des Betroffenen werden gerade im Rahmen der Interessenabwägung im engeren Sinne betrachtet. [74]

Das folgende Beispiel illustriert ein Fehlschlagen der Erforderlichkeitsprüfung. In der Trainingsphase ist die Erforderlichkeit für das Training einer KI-Anwendung mit personenbezogenen Daten im Ergebnis zB zu verneinen, wenn das Training mit anonymen oder synthetischen Daten ebenso wirksam möglich ist.

Insgesamt sieht die Behörde an vielen Stellen Unwägbarkeiten beim Einsatz von KI[75], und geht dabei teilweise nicht auf konkrete datenschutzrechtliche Risiken ein, die bei der Abwägung zu berücksichtigen sind. Für die Praxis wäre es hilfreicher, wenn solche Herausforderungen konkret benannt würden, um diesen beim Einsatz von KI begegnen zu können, anstatt den Anwendern als Lösung den Einsatz vermeintlich datensparsamerer Alternativen aufzuzeigen.

Nach unserem Verständnis bedarf es aber einer gesonderten Prüfung der Interessenabwägung, wenn in der Phase der Anwendung von KI-Modellen die personenbezogenen Eingabedaten auch für Trainingszwecke verwendet werden sollen (s.o. Ausführungen zur Trainingsphase).

(c) KI und Datenminimierung

Eine Differenzierung nach den Phasen der Verarbeitungen bei KI-Anwendungen erfolgt durch die Behörde nicht. Sie stellt fest, dass die Prüfung der Erforderlichkeit vom Grundsatz der Datenminimierung bestimmt ist und im Zusammenhang mit KI-Systemen nur das Nötigste verwendet werden soll. Aus dieser richtigen Überlegung schlussfolgert die Behörde jedoch zu verkürzt, dass ein Vorgehen nach dem Prinzip „Viel hilft viel!“ nicht der Prüfung der Erforderlichkeit standhalten würde.[76] Diese vereinfachte Darstellung wird den Spezifika der KI-Entwicklung nicht gerecht. Bei der KI-Entwicklung und insbesondere im Rahmen der Trainingsphase ist zu beachten, dass für die spätere Qualität der Ergebnisse von KI-Anwendungen eine große und möglichst umfassende Verarbeitung von Datenmengen entscheidend ist. Zum einen ist die Quantität der Trainingsdaten entscheidend, um bestimmte KI-Anwendungen überhaupt entwickeln zu können.[77] Außerdem bestimmt die Menge der Daten bis zu einem gewissen Grad die Qualität der Ergebnisse von KI-Anwendungen im Bereich des Machine Learnings.[78] Zum anderen können bestimmte negative Effekte bei der Funktionalität von KI-Systemen nur verhindert werden, wenn eine möglichst umfassende Datenbasis vorliegt. Diese Überlegung findet sich auch in den Entwürfen des Rats als auch des EU-Parlaments zur KI-Verordnung wieder. In den Entwürfen des Art. 10 Abs. 5 wurde sogar eine neue Rechtsgrundlage zur Verarbeitung von

[73] Kühling/Buchner, DS-GVO/Buchner/Petri 4. Aufl. 2024, Art. 6 Rn. 147c
[74] Gola/Heckmann, DS-GVO/Schulz, 3. Aufl. 2022, Art. 6 Rn. 61; auch für geringe Filterwirkung der Erforderlichkeitsprüfung: Sydow/Marsch, DS-GVO/BDSG/Reimer, 3. Aufl. 2022, Art. 6 Rn. 81
[75] LfDI BW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz V. 1.0, S. 13 f, 19
[76] LfDI BW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz V. 1.0, S. 17
[77] Begleitforschung Mittelstand-Digital WIK GmbH, Künstliche Intelligenz im Mittelstand Relevanz, Anwendungen, Transfer, S. 10, online abrufbar unter: https://www.mittelstand-digital.de/MD/Redaktion/DE/Publikationen/kuenstliche-intelligenz-im-mittelstand.pdf?__blob=publicationFile&v=5
[78] Alzubaidi et al., A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. J Big Data 10, 46 (2023), S. 13f; Keber/Maslewski, RDV 2023, 273, (274)

besonderen Kategorien personenbezogener Daten im Sinne von Art. 9,10 DS-GVO für die Erkennung und Beseitigung von negativen Verzerrungseffekten bei Hochrisiko-KI-Anwendungen vorgesehen.[79]

Der kurze Abschnitt in dem Papier der Behörde wird dem Thema nicht gerecht und kann aufgrund seiner Knappheit beim Leser zu falschen Annahmen führen, die rechtlich nicht haltbar sind, sowie zu unnötigen Verkomplizierungen von KI-Systemen und könnte auch dem Willen des europäischen Gesetzgebers widersprechen. Es ist eine genaue Betrachtung der konkreten Entwicklung unter der Einbeziehung der technischen Gegebenheiten notwendig.[80]

bb) Interessenabwägung

Im Rahmen der Prüfung der Interessenabwägung im engeren Sinne wäre eine stärkere Differenzierung nach den unterschiedlichen Phasen ebenfalls wünschenswert. Insbesondere eine Auseinandersetzung mit der Phase der Nutzung der KI-Anwendung, bspw. durch einen anderen als den KI-Ersteller, findet in dem Abschnitt keine Erwähnung. Diese Konstellation ist insofern besonders praxisrelevant, als das davon auszugehen ist, dass in den generativen LLM von großen Anbietern (beispielsweise GPT4) personenbezogene Daten von „unbeteiligten" betroffenen Personen enthalten sind, die im Rahmen von Web-Crawling gesammelt wurden.[81]

Dabei muss zwischen zwei Varianten differenziert werden: zum einen den personenbezogenen Daten von Dritten, die bereits in dem Modell enthalten sind und möglicherweise im Hintergrund einer Anfrage an die KI oder unmittelbar als Ergebnis verarbeitet werden; zum anderen personenbezogenen Daten von Dritten, welche aufgrund der Eingabe des Nutzers verarbeitet werden.

Hinsichtlich der ersten Variante bleibt für die Anwender solcher KI-Systeme oft nur Art. 6 Abs. 1 UAbs. 1 lit. f DS-GVO als mögliche Rechtsgrundlage. Für den Fall, dass bestimmte personenbezogene Daten nur im Hintergrund des LLM verarbeitet und bei der Beantwortung der Anfrage dem Anwender nicht angezeigt werden, ist dies im Rahmen der Interessenabwägung zugunsten des Anwenders zu berücksichtigen. Für die bereits im Modell enthaltenen Daten, sowohl für die im Hintergrund verarbeiteten als auch im Ergebnis angezeigten, stellt die Prüfung der vernünftigen Erwartungen der Verarbeitung (ErwG 47 S.1, 3, 4 DS-GVO) eine Herausforderung im Rahmen der Interessenabwägung im engeren Sinne dar.

Im Hinblick auf die Betroffenen, deren Daten gerade nur durch die Eingabe der Anfrage verarbeitet werden, kann unter anderem Art. 6 Abs. 1 UAbs. 1 lit. f DS-GVO als Rechtsgrundlage dienen. In einem solchen Fall ist eine Gruppierung von möglichen Personen anhand der Verbindung zum Anwender und der konkreten Verarbeitung sinnvoll. Mögliche Gruppen könnten dabei sein: Geschäftspartner und Bewerber des Anwenders sowie mit dem Anwender nicht in

[79] https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf
; Eine plastische Demonstration von der Qualität des Voice Clonings anhand von 30 Minuten Audiomitschnitten und knapp 8 Stunden von Audiomitschnitten https://www.youtube.com/watch?v=EAowXV_hWks
[80] Keber/Maslewski, RDV 2023, 273, (279)
[81] https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed; Bei Tests mit ChatGPT durch Google-Wissenschaftler wurde nachgewiesen, dass das dahinterliegende Modell sich an ihre Trainingsdaten teils erinnern kann und diese Trainingsdaten unter anderem öffentlich zugängliche Blogartikel und Webseiten enthalten, https://not-just-memorization.github.io/extracting-training-data-from-chatgpt.html?ref=404media.co

Verbindung stehende Dritte. Eine solche Gruppierung ermöglicht es, die unüberschaubare Anzahl von Betroffenen zu überschaubaren Gruppen zusammenzufassen und dabei ihre Interessen auf Grund der Differenzierungen nach ihrem Bezug zu dem Anwender zu würdigen.

b) Zu Kapitel V Nr. 1: Einwilligung, Art. 6 Abs. 1 lit. a DS-GVO

Auch bezüglich der Einwilligung ist es unserer Einschätzung nach hilfreich, zwischen den Phasen der Verarbeitung zu unterscheiden. Der LfDI spricht insoweit eine relevante Herausforderung an, nämlich dass im Fall eines Widerrufs personenbezogene Daten, die auf Basis einer Einwilligung verarbeitet werden, ggf. gelöscht werden müssen (Art. 17 Abs. 1 lit. b DS-GVO). Dabei ergeben sich aber Besonderheiten in den unterschiedlichen Phasen der Verarbeitung:

In der Phase der Erhebung von Trainingsdaten ergeben sich keine spezifischen Herausforderungen. Ein Widerruf, der beim Verantwortlichen eingeht, kann berücksichtigt werden und die erhobenen Trainingsdaten können gelöscht werden.

In der Phase der Verarbeitung von Daten für das Training der KI und der Entwicklung der KI ist dieser Aspekt jedoch besonders praxisrelevant, da etwaige hierbei verarbeitete personenbezogene Daten ggf. nicht direkt und mit einfachen Mitteln erkennbar sind[82] und ihre datenschutzkonforme Löschung eine erhebliche Herausforderung sein kann.

In der Phase des Bereitstellens kann dies im Einzelfall ebenfalls relevant sein, da – wenn in dem KI-Modell selbst Daten enthalten sind – auch die Bereitstellung von der Einwilligung umfasst sein muss und daher auch hier die vom LfDI aufgezeigte Problematik der Löschung nach Widerruf relevant sein kann. In diesem Fall können die Erkennung und Löschung von personenbezogenen ebenfalls eine erhebliche Herausforderung – ggf. sogar unmöglich – sein.

Dahingegen gibt es bei der Nutzung von KI und der Nutzung von Ergebnissen der KI unserer Ansicht nach keine Besonderheiten im Vergleich zum Einsatz anderer Technologien – es sei denn die personenbezogenen Daten werden zugleich auch zum Training der KI verwendet (zB wenn eine betroffene Person gem. Art. 6 Abs. 1 lit. a DS-GVO eingewilligt hat, dass ihre Daten für Zwecke des Direktmarketings durch den Verantwortlichen genutzt werden, und der Verantwortliche eine KI einsetzt, um die Daten der betroffenen Person zu verarbeiten und ihr Produkte anzubieten, für die sie sich wahrscheinlich interessiert, ohne dass dabei ihre Daten im KI-System selbst gespeichert werden).

c) Zu Kapitel V Nr. 2: Art. 6 Abs. 1 Buchst. b DS-GVO

Auch zur Rechtsgrundlage des Art. 6 Abs. 1 lit. b DS-GVO gelten die obigen Ausführungen.

Darüber hinaus geht der LfDI auf die folgenden praxisrelevanten Fragestellungen nicht ein:

aa) Er führt aus, dass „Allein durch die Aufnahme in eine Nutzungsvereinbarung […] eine Datenverarbeitung insoweit nicht bereits rechtmäßig sein [kann]“. Dabei geht er aber nicht auf naheliegende, praxisrelevante Gestaltungen ein, denen datenschutzrechtliche Regelungen nicht per se entgegenstehen: nämlich Modelle, bei denen der Nutzer mit der Bereitstellung seiner Daten „bezahlt“.[83] Aufgrund der Bedeutung solcher Modelle für die (Weiter-)

[82] S.o. 2., b)
[83] Vgl. Hacker, ZfPW 2019, 148 ff.

Entwicklung von KI, wären hierzu Ausführungen hilfreich, die insoweit den aus Sicht des LfDI zulässigen Rahmen und konkrete Ausgestaltungen abstecken.

bb) Soweit der LfDI in Bezug auf medizinische Behandlungen Art. 6 Abs. 1 lit. b DS-GVO iVm § 630a Abs. 1 BGB heranzieht, ist dem grundsätzlich zuzustimmen. Soweit er aber darüber hinaus verlangt, dass „der Einsatz des KI-Systems aus der Perspektive der betroffenen Personen vernünftigerweise zu erwarten sein [müsste] und die betroffenen Personen […] über die Funktionalität des KI-Systems informiert werden [müsste]", bedarf dies weiterer Ausführungen, inwieweit es sich dabei gerade um datenschutzrechtliche Anforderungen handelt, denn der Umfang der allgemeinen Informationspflichten gem. Art. 13, 14 DS-GVO umfasst gerade keine Ausführungen zu den eingesetzten Technologien.[84]

d) Zu Kapitel V Nr. 3: Art. 6 Abs. 1 UAbs. 1 lit. c DS-GVO

In ihren Ausführungen zu dieser Rechtsgrundlage differenziert die Behörde nicht zwischen den unterschiedlichen Verarbeitungsschritten des Lebenszyklus einer KI. Ebenfalls benennt die Behörde keine Beispiele für Verarbeitungen mit KI-Anwendungen auf Grundlage des Erlaubnistatbestandes und beschränkt sich auf eine Wiedergabe der Tatbestandsvoraussetzungen des Art. 6 Abs. 1 UAbs. 1 lit. c DS-GVO.

Insbesondere in Bezug auf die Phase des Trainings wären Ausführungen zur Zulässigkeit der Verarbeitung von personenbezogenen Daten unserer Einschätzung nach hilfreich. Denn auch für die Erfüllung gesetzlicher Pflichten (zB der Daseinsfürsorge für öffentliche Stellen oder Umsetzung gesetzlicher Pflichten durch nicht-öffentliche Stellen) kann die Verwendung moderner Datenverarbeitungssysteme einschließlich KI erforderlich sein. Die vom EuGH insoweit vorgenommene Beschränkung auf das „unbedingt Notwendige"[85] ist kein geeignetes Kriterium, um herauszuarbeiten, für die Erfüllung welcher gesetzlichen Anforderungen ein Einsatz von KI zulässigerweise erfolgen kann.

Ein Beispiel für die Rechtfertigung der Datenverarbeitung bei der Anwendung einer KI zur Erreichung einer gesetzlichen Pflicht ist in folgendem Fall denkbar: Liegt bereits eine KI vor, die zum Beispiel mit Hilfe eines Videoterminals dazu geeignet ist, die gem. § 5 PAuswG notwendigen Daten aufzunehmen, dann wäre ihr Einsatz gem. Art. 6 Abs. 1 UAbs. 1 lit. c DSGVO i.V.m. §§ 9, 5 PAuswG gerechtfertigt. Ein Unterschied zwischen der Datenverarbeitung durch einen Menschen und einem mit KI-Software ausgestatteten Terminal ist aufgrund der Technologieneutralität der DSGVO nicht gegeben.

e) Zu Kapitel V Nr. 3: Art. 6 Abs. 1 UAbs. 1 lit. d DS-GVO

Hinsichtlich des Art. 6 Abs. 1 UAbs. 1 lit. d DS-GVO ist der Behörde im Regelfall im Hinblick auf das Training von KI-Systemen zuzustimmen. Es ist aber hinzuzufügen, dass die Grenzen des KI-Trainings und der KI-Anwendung in bestimmten Fällen ineinanderfließen können. So ist es möglich, dass eine KI im Gesundheitsbereich zunächst mit den Daten des Patienten in Notsituationen angelernt werden muss, um anschließend zuverlässige Ergebnisse in der Analyse liefern zu können.[86] In einem solchen Fall ist Art. 6 Abs. 1 UAbs. 1 lit. d DS-GVO auch für das Trainieren der KI eine mögliche Rechtsgrundlage.

---

[84] Vgl. Art. 29 Gruppe, WP 260 rev.01WP 260, Anhang

[85] EuGH, Urteil vom 4. Juli 2023, Meta Platforms, C-252/21 Rn. 138

[86] Ein solches kurzfristiges Antrainieren von KI-Systemen ist bereits jetzt im Bereich des Voice Clonings im Einsatz, https://www.heise.de/news/Synthetische-Stimme-verschafft-Zugang-zu-Bankdaten-7527666.html

Nach Beendigung des Notfalls müsste die antrainierte KI gelöscht werden aufgrund des Zweckfortfalls. Findet sich jedoch eine andere Rechtsgrundlage zum Erhalt der personenbezogenen Daten, beispielsweise eine Einwilligung, können die antrainierten Daten der KI für weitere Behandlungen des Patienten verarbeitet werden.

f) Zu Kapitel VI Nr. 2: Beschäftigtendatenschutz, § 26 BDSG

In Bezug auf die Verarbeitung von Beschäftigtendaten in KI verweist der LfDI zunächst auf die Frage der Anwendbarkeit der nationalen Vorschrift (§ 26 Abs. 1 S. 1 BDSG) aufgrund der Rechtsprechung des EuGH ist dies zumindest umstritten.[87] Zutreffend führt der LfDI aus, dass aber auf die Rechtsgrundlage von Art 6 Abs. 1 UAbs. 1 lit. b DS-GVO zurückgegriffen werden kann,[88] insoweit verweisen wir auf die obenstehenden Ausführungen hierzu.

[87] Vgl. EuGH, Urteil vom 30. März 2023, C 34/21.
[88] LfDI BW, Diskussionspapier: Rechtsgrundlagen im Datenschutz beim Einsatz von Künstlicher Intelligenz V. 1.0, IV. Nr. 2, S. 19