www.jusletter.ch

David Rosenthal

Datenschutz beim Einsatz generativer künstlicher Intelligenz

Ist der datenschutzkonforme Einsatz generativer künstlicher Intelligenz möglich? Eine Analyse zeigt: Das Schweizer Datenschutzgesetz kommt auch mit dieser technischen Entwicklung gut zurecht. Es braucht nicht weitere Grundsätze und Pflichten, sondern neue technische und organisatorische Massnahmen, mit denen das bestehende Recht sinnvoll umgesetzt werden kann. Dabei ist klar zwischen dem zu unterscheiden, was der Gesetzgeber vorschreibt, und dem, was aus «ethischen» Überlegungen wünschenswert sein mag. Dieser Beitrag beschäftigt sich bewusst nur mit Ersterem. Er setzt einen früheren Beitrag des Autors zum Datenschutz bei KI-Systemen fort.

Beitragsart: Wissenschaftliche Beiträge
Rechtsgebiete: Datenschutz, Medienrecht

Zitiervorschlag: David Rosenthal, Datenschutz beim Einsatz generativer künstlicher Intelligenz, in: Jusletter 6. November 2023

Inhaltsübersicht

1. Einleitung
2. Gesetz versus Ethik
3. Verantwortlichkeit
   3.1. Grundsatz
   3.2. Herstellung eines Modells
   3.3. Besitz eines Modells
   3.4. Verwendung eines Modells
   3.5. Relevanz der Abgrenzung
   3.6. Keine Überprüfungspflicht
4. Training eines gKI-Modells
   4.1. Eine Frage der Zweitverwertung
   4.2. Öffentliche Informationen
   4.3. Knifflige Fragen der Rechtfertigung
   4.4. Sonderproblem des Fine-Tuning
5. Einsatz eines gKI-Systems
   5.1. Personendaten als Input
   5.2. Personendaten als Output
   5.3. Erfordernis der Transparenz
6. Betroffenenrechte
   6.1. Auskunftsrecht
   6.2. Widerspruchs- und Korrekturansprüche
7. Fazit

1. Einleitung

[1] Generative künstliche Intelligenz («gKI»)[1] ist inzwischen im geschäftlichen und privaten Alltag allgegenwärtig. Mitarbeitende in Unternehmen setzen «ChatGPT», den «Bing» oder auch Tools wie «DeepL» inzwischen (erlaubt oder unerlaubt) für alltägliche Arbeiten ein – McKinsey spricht verheissungsvoll von einem «Produktivitätsbooster»[2] –, in den Marketing- und anderen Abteilungen wird mit Chatbots, Content-Generatoren und anderen Einsatzmöglichkeiten für das Unternehmen ausprobiert und im privaten Bereich werden eigene Bilder für soziale Medien mit gKI aufgepeppt.[3] Kriminelle formulieren mit Tools wie «WormGPT» noch überzeugendere Phishing-Mails[4] oder produzieren Schock-Anrufe mit geklonten Stimmen.[5] In einem

1 Die ZHAW umschreibt gKI beispielsweise so: «Generative KI ist eine Art von künstlicher Intelligenz, die in der Lage ist, neue Daten zu erstellen, die teilweise schwer von menschlichen Inhalten zu unterscheiden sind. Zu den generierten Inhalten zählen zum Beispiel Text, Ton und Bild.» Kompetenzzentrum für Generative KI, undatiert (https://www.zhaw.ch/de/engineering/forschung/plattformen/kompetenzzentrum-fuer-generative-ki/, archiviert unter https://perma.cc/K8GL-K78S).

2 Philipp Huehne, Studie: Generative KI kann zum Produktivitätsbooster werden, in: McKinsey & Company, 13. Juni 2023 (https://www.mckinsey.com/de/news/presse/genai-ist-ein-hilfsmittel-um-die-produktivitaet-zu-steigern-und-das-globale-wirtschaftswachstum-anzukurbeln, archiviert unter https://perma.cc/QT66-LF4N).

3 Megan McCluskey, Why It's So Hard to Resist Turning Your Selfies Into Lensa AI Art, in: TIME, 14. Dezember 2022 (https://time.com/6240648/lensa-ai-psychology-behind/, archiviert unter https://perma.cc/63LT-27DQ).

4 Sebastian Feurer, WormGPT ist das ChatGPT für Kriminelle: So gefährlich ist die neue KI, in: CHIP, 17. Juli 2023 (https://www.chip.de/news/WormGPT-ist-das-ChatGPT-fuer-Kriminelle-So-gefaehrlich-ist-die-neue-KI_184872554.html, archiviert unter https://perma.cc/ZN53-76X2).

5 Johannes Böhler, Schockanrufe mit geklonter Stimme: Neue Betrugsmasche per KI, in: SWR, 30. August 2023 (https://www.swr.de/swraktuell/baden-wuerttemberg/schockanrufe-mit-ki-100.html, archiviert unter https://perma.cc/SHX5-3TPS).

anderen Fall hat eine virtuelle Freundin des Chatbot-Services Replika[6] offenbar gleich von sich aus einen jungen Mann ermutigt, die Queen zu ermorden.[7] Kunstschaffende ärgern sich derweil, wenn KI-Kunst[8] an Wettbewerben ihnen die Preise wegschnappt,[9] oder verklagen KI-Firmen, weil sie ihre Inhalte illegal für das Training verwendet haben sollen,[10] und auch an Schweizer Schulen und Universitäten wird noch darüber diskutiert, ob und wie gKI zum Einsatz kommen darf.[11] Sogar für den Datenschutz kommen sie zum Einsatz: So hat der Autor dieses Beitrags eine Vorlage für eine Datenschutz-Folgenabschätzung konzipiert und programmiert, welche sich mit Hilfe von gKI selbst ausfüllen kann – und damit zu einem besseren Risikomanagement im Datenschutz beiträgt.[12]

[2] So vielfältig die Einsatzmöglichkeiten von gKI sind, so vielfältig sind die rechtlichen Fragestellungen beim Einsatz dieser Technik. Das gilt auch für den Datenschutz. Immerhin lebt künstliche Intelligenz («KI») von Daten, und manche davon weisen einen Personenbezug auf und fallen somit unter das Datenschutzgesetz («DSG»). Datenschutzrechtliche Fragen stellen sich dabei in allen vier Phasen des Einsatzes von gKI, wie die Abbildung 1 zeigt:[13] Dies beginnt mit der Beschaffung von Informationen für den Aufbau einer gKI (Phase 1), dem Training des Modells (Phase 2), dessen Besitz und Betrieb (Phase 3) und dessen praktischer Verwendung (Phase 4), bei welcher ein Benutzer einen Input tätigt («Prompt») und dieser mit einem Output beantwortet wird, beispielsweise einem generierten Bild oder Text. Der Input wird je nach Konzept ebenfalls für das Training verwendet. In jeder dieser Phasen stellen sich eigenständige datenschutzrechtliche Herausforderungen (unter jeder Phase angeordnet). Im Training muss je nach Rechtslage darauf geachtet werden, dass die betroffenen Personen, deren Daten verwendet werden, nicht re-identifiziert werden können. Wer ein gKI-Modell besitzt, wird es typischerweise als besondere Herausforderung empfinden, die Betroffenenrechte zu gewährleisten. In allen Phasen wird zudem die Einhaltung der Grundsätze der Verhältnismässigkeit, Zweckbindung oder Richtigkeit der Daten ein Thema sein sowie die Frage, welcher der an der Bearbeitung beteiligten Stellen welche datenschutzrechtliche Rolle (Controller, Processor, Joint-Controller) hat. Einige dieser Herausforderungen werden in diesem Beitrag erläutert.

---

6 https://replika.com/ («The AI companion who cares»).

7 Sentencing Remarks of Mr Justice Hilliard, Central Criminal Court, 5. Oktober 2023 (https://www.judiciary.uk/wp-content/uploads/2023/10/R-v-Chail-sentencing-050923.pdf, archiviert unter https://perma.cc/67L5-98PF); Will Bedingfield, A Chatbot Encouraged Him to Kill the Queen. It's Just the Beginning, in: Wired, 18. Oktober 2023 (https://www.wired.com/story/chatbot-kill-the-queen-eliza-effect/, archiviert unter https://perma.cc/QEF5-EGUE).

8 Beispiele: Robert Leitinger, Diese Bilder habe ich mit künstlicher Intelligenz generiert | KI-Bilder, 06. Oktober 2023 (https://robert-leitinger.com/ki-bilder/, archiviert unter https://perma.cc/79JD-6QLJ).

9 Robert Leitinger, Diese Bilder habe ich mit künstlicher Intelligenz generiert | KI-Bilder, 06. Oktober 2023 (https://robert-leitinger.com/ki-bilder/, archiviert unter https://perma.cc/79JD-6QLJ).

10 Megan McCluskey, Why It's So Hard to Resist Turning Your Selfies Into Lensa AI Art, in: TIME, 14. Dezember 2022 (https://time.com/6240648/lensa-ai-psychology-behind/, archiviert unter https://perma.cc/63LT-27DQ).

11 Tobias Solis, Die ChatGPT-Richtlinien der 23 Schweizer Hochschulen, in: Scribbr, 06. Juni 2023 (https://www.scribbr.ch/ki-tools-nutzen-ch/chatgpt-richtlinien-hochschulen-schweiz/, archiviert unter https://perma.cc/Z2RB-4FBZ).

12 David Rosenthal, VUD Datenschutz-Folgeabschätzung, Zürich (https://www.rosenthal.ch/downloads/vud_dsfa.xlsm).

13 Die Grafik ist unter einer freien Lizenz abrufbar unter https://www.rosenthal.ch/downloads/Rosenthal-generative-KI-Datenschutzthemen.pdf und https://www.rosenthal.ch/downloads/Rosenthal-generative-ai-data-protection-issues.pdf (auf Englisch).

2. Gesetz versus Ethik

[3] Diskutiert wird vorliegend nur, was das DSG von einem Datenbearbeiter tatsächlich verlangt. Das ist weniger als das, was nicht nur die Öffentlichkeit, sondern oft auch die Aufsichtsbehörden von Unternehmen erwarten: Diese sollen sich nicht nur an das Gesetz halten, sondern beim Einsatz von KI auch sog. ethischen Grundsätzen folgen. Gemeint sind übergesetzliche Anforderungen, die aus oft nicht näher definierten Wertvorstellungen abgeleitet werden. Diese Werte variieren naturgemäss je nach Unternehmen und Anwendungsfall: Von Institutionen wie der Post oder den SBB wird ein anderes Verhalten erwartet als von einem Start-up, von einem Online-Händler ein anderes als von einer Versicherung. Von solchen Anforderungen soll hier nicht die Rede sein.

[4] Drei Dinge seien hierzu aber angemerkt. *Erstens* kann KI-Ethik im Unternehmen als Compliance-Thema wie jedes andere betrieben werden. Wie das geht, hat der Autor dieses Beitrags am Beispiel der Datenethik beschrieben;[14] das lässt sich analog auch im Bereich der gKI umsetzen. Die Schweizer Wirtschaft ist noch kaum an diesem Punkt. Die meisten hiesigen Unternehmen, die KI bei sich reguliert haben, haben dies nur mit Richtlinien getan, welche offenkundige Risiken wie etwa das Füttern mit Personendaten oder Geschäftsgeheimnissen oder das unkritische Verwenden von generierten Inhalten adressieren.[15] Erst in einem zweiten Schritt folgen übergesetzliche Vorgaben. Das kommt zwar bereits vereinzelt vor. Typische Themen sind «Achtsamkeit» und «Verantwortungsbewusstsein» im Umgang mit KI und Grundsätze wie die Offenlegung von KI-Anwendungen, die Gleichbehandlung und Vermeidung von Diskriminierung, das Primat menschlicher statt maschineller Entscheidungen oder das Gebot, KI nicht zum Nachteil von Personen, sondern zur Verbesserung von Angeboten und Dienstleistungen zu verwenden. Die selbstauferlegten Vorgaben sind jedoch sehr generischer Natur. Erst in einem dritten Schritt werden sie beginnen, das Thema «KI-Ethik» systematisch anzugehen.

[5] So attraktiv und populär dieses Thema auch sein mag, sollten sich die Aufsichtsbehörden *zweitens* auf die Durchsetzung nur der gesetzlichen Pflichten beschränken. Wie Unternehmen über das gesetzliche Minimum hinaus mit KI umgehen, geht sie im Grunde nichts an, so wünschenswert ein ethischer Umgang mit Personendaten auch sein mag. Anpassungen des rechtlichen Rahmens sind dem Gesetzgeber vorbehalten. Trotzdem ist erfahrungsgemäss damit zu rechnen, dass die Datenschutzbehörden auch in der Schweiz versuchen werden, übergesetzliche Anforderungen durchzusetzen. Sie werden dies als «Datenschutz» verpacken, wie sie dies schon mit anderen Anliegen getan haben – und vor Gericht (zu Recht) zurückgepfiffen wurden, wo Unternehmen den Mut hatten, sich dagegen zu wehren.[16]

[6] Ethik wird bei neuen technischen Entwicklungen in der Regel dann angerufen, wenn die bestehenden Gesetze mutmasslich relevante Regelungslücken aufweisen. Ob dies beim DSG in Bezug auf das Thema gKI – oder KI generell – tatsächlich so ist, muss sich *drittens* erst noch weisen. Nach der hier vertretenen Auffassung gibt es basierend auf der bisherigen Erfahrung bisher keine wirklichen Hinweise auf solche Lücken. In der Diskussion um die Regulierung von

---

14 David Rosenthal, Datenethik – ein praktischer Zugang aus Sicht der Compliance, in: Recht relevant. für Compliance Officers, 1/2022, Zürich (https://www.rosenthal.ch/downloads/RRCO-Rosenthal-Datenethik.pdf).

15 Vgl. das Beispiel von Ringier: https://www.ringier.com/de/ringier-fuehrt-klare-richtlinien-fuer-den-einsatz-kuenstlicher-intelligenz-ein/, archiviert unter https://perma.cc/6JQD-YW3B.

16 Vgl. etwa Helsana+: Urteil wird rechtskräftig, undatiert (https://www.edoeb.admin.ch/edoeb/de/home/kurzmeldungen/nsb_mm.msg-id-75039.html, archiviert unter https://perma.cc/QZ23-Y9EN).

KI wird beispielsweise oft übersehen, dass mit der Regelung automatisierter Einzelentscheide[17] das DSG bereits eine gesetzliche Regelung einer als besonders heikel empfundenen Anwendung von KI hat. Durch seine prinzipienbasierte und technologieneutrale Ausgestaltung lassen sich mit dem heutigen DSG zudem unterschiedlichste Fallkonstellationen gut und zeitlos erfassen. Dank der Einführung der sog. Datenschutz-Folgenabschätzung («DSFA»)[18] verfügt das revidierte DSG inzwischen sogar über ein universelles Instrument zur strukturierten Beurteilung von Risiken, welche KI-Systeme für betroffene Personen mit sich bringen können. Der Grundsatz des «Privacy-by-Design»[19] verlangt von Unternehmen, die KI-Systeme einführen, dass sie sich um den Datenschutz von Beginn weg kümmern – einschliesslich der in der Praxis besonders herausfordernden Frage der Betroffenenrechte. Wer diesen Grundsatz nicht respektiert, wird es später schwer haben, sich auf den hohen Aufwand zu berufen, den die nachträgliche Umsetzung solcher Betroffenenrechte im Rahmen eines KI-Systems mit sich bringen kann.

[7] Der Autor dieses Beitrags hat bereits 2022 eine ausführliche Darstellung der datenschutzrechtlichen Herausforderungen von KI nach Schweizer Recht publiziert, in welchem auch klassische KI-Themen wie etwa der «Machine Bias» einlässlich behandelt wurden.[20] Auf diese Aspekte wird hier daher bewusst nicht mehr eingegangen. Im vorliegenden Beitrag liegt der Fokus vielmehr auf den besonderen datenschutzrechtlichen Herausforderungen von gKI. Dabei lehnt sich dieser Beitrag an den Leitfaden für den Einsatz von gKI an, den der Autor für den Verein Unternehmens-Datenschutz («VUD»)[21] verfasst hat. Dieser ist Ende August 2023 vom VUD in einer Version «for public comment» publiziert worden ist.[22]

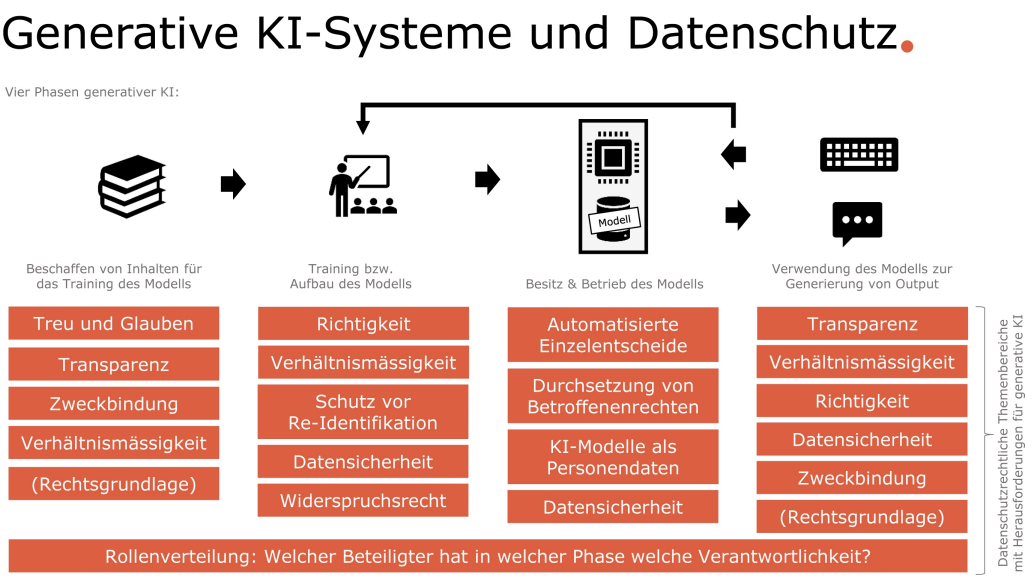

Abbildung 1: Datenschutzrechtliche Problemstellungen beim Einsatz generativer KI

17 Art. 21 DSG.

18 Art. 22 DSG.

19 Art. 7 DSG.

20 David Rosenthal, Datenschutz und KI: Worauf in der Praxis zu achten ist, in: Jusletter IT, 22. April 2022 (https://www.rosenthal.ch/downloads/Rosenthal-KI-Datenschutz.pdf).

21 Im VUD (www.vud.ch) sind die Datenschutzbeauftragten von über 100 Schweizer Unternehmen vertreten.

22 Verein Unternehmens-Datenschutz (VUD), Generative KI: Leitfaden zum Datenschutzgesetz, 29. August 2023, Autor: David Rosenthal (https://www.rosenthal.ch/downloads/VUD_Leitfaden_generative_KI_DSG.pdf).

3. Verantwortlichkeit

3.1. Grundsatz

[8] Über die datenschutzrechtliche Verantwortlichkeit für ein gKI-System und dessen Nutzung wurde selbst in Fachkreisen bisher wenig diskutiert. Sie steht jedoch am Anfang einer jeder datenschutzrechtlichen Compliance-Prüfung. Wer ist der Verantwortliche («Controller») und wer der Auftragsbearbeiter («Processor»)? Auch das DSG unterscheidet neuerdings diese beiden Begriffe und definiert sie analog zur EU-Datenschutz-Grundverordnung («DSGVO»).[23] Manche der flankierenden Bestimmungen des DSG, insbesondere die Betroffenenrechte, gelten nur für den Verantwortlichen. Auch andere Bestimmungen, wie die Bearbeitungsgrundsätze, sind in erster Linie in der Verantwortung des Verantwortlichen – wie der Name schon sagt.

[9] Die Abgrenzung der Figur des Verantwortlichen und jener des Auftragsbearbeiters sowie Sonderfragen wie die «gemeinsame» Verantwortlichkeit hat der Autor ebenfalls bereits einlässlich diskutiert,[24] auch für KI im Allgemeinen.[25] Im Wesentlichen sind nach DSG diejenigen Stellen verantwortlich für die Einhaltung des DSG, die den Zweck und die Mittel, d.h. die datenschutzrechtlich wesentlichen Parameter, einer Datenbearbeitung festlegen.[26]

3.2. Herstellung eines Modells

[10] Die eigentliche Herausforderung im Bereich der gKI ist allerdings nicht zu bestimmen, wer in diesem Sinne das «Ruder» einer Datenbearbeitung in der Hand hält, sondern was die «Datenbearbeitung» im konkreten Fall genau ist bzw. wo sie anfängt und aufhört. Das ist deshalb so, weil heute viele Anwendungen der gKI mittels mehrfacher Arbeitsteilung realisiert werden:

- Jemand trainiert das einer gKI zugrundeliegende Modell[27] (z.B. die LLM[28] «GPT-4», «PaLM» oder «Llama 2»);
- Jemand bringt es als Service auf den Markt (z.B. für Konsumenten «ChatGPT», «Bard» oder «Bing» und für Unternehmen die gKI-API[29] von OpenAI, Microsoft oder DeepL);
- Jemand nutzt das Modell für eine eigene Anwendung (z.B. ein Unternehmen, welches einen Chatbot auf der Basis eines LLM entwickeln will und daher eine passende gKI-API nutzt).

[11] Derjenige, der das Modell gebaut hat, hat mit der Anwendung nichts zu tun, und derjenige, der die Anwendung realisiert, mag zwar wissen, wie das Modell für seine Zwecke funktioniert, ist aber in dessen (initialem) Training, also den Aufbau des Modells, in der Regel nicht involviert

---

23 Art. 5 Bst. j und k DSG.

24 David Rosenthal, Controller oder Processor: Die datenschutzrechtliche Gretchenfrage? in: Jusletter 17. Juni 2019 (http://www.rosenthal.ch/downloads/Rosenthal-ControllerProcessor.pdf).

25 Rosenthal, Datenschutz und KI (FN 20).

26 Vgl. Art. 5 Bst. j DSG.

27 Quasi das Wissen der gKI, auf deren Basis es Output generieren kann.

28 Large Language Modell: KI-Modelle für sprachbasierte gKI, also die Basis, mit welcher Chatbots wie ChatGPT oder Bard arbeiten. Hintergrundbericht zu ihrer Funktionsweise: David Nield, How ChatGPT and Other LLMs Work–and Where They Could Go Next, in: Wired, 30. April 2023 (https://www.wired.com/story/how-chatgpt-works-large-language-model/, archiviert unter https://perma.cc/W5RM-AFA7).

29 Application Programing Interface: Ein Service, bei welchem ein Computerprogramm des Benutzers direkt mit dem LLM des Anbieters kommunizieren kann.

und scheidet daher als Verantwortlicher aus. Entscheidet wiederum derjenige, der ein Modell trainiert, nicht, wie es später angewandt wird, kann er sachlogisch auch nicht Verantwortlicher dieser späteren Anwendung sein, und umgekehrt, da das Datenschutzrecht die Verantwortlichkeit an diese Entscheide anknüpft.

[12] Das bedeutet alles nicht, dass beispielsweise den Ersteller eines Modells keine Pflichten hinsichtlich der Verwendung seines Modells treffen. Die Weitergabe des Modells kann beispielsweise eine Bekanntgabe von Personendaten sein, falls ein nachgelagerter Nutzer aus dem Modell Daten extrahieren kann, die sich einer bestimmten oder bestimmbaren natürlichen Person zuweisen lassen und vernünftigerweise auch damit zu rechnen ist. Die Erstellung des Modells ist jedoch eine logisch von dessen Einsatz getrennte Datenbearbeitung. Für den Datenschutz bedeutet dies, dass die Verantwortlichkeit für die Erstellung und den Besitz eines KI-Modells zwei verschiedene Stellen treffen kann. Das macht auch Sinn so: Kommt es beispielsweise in Bezug auf die Kopie eines KI-Modells, die ein Unternehmen auf seinen Anlagen gespeichert hat, zu einer Verletzung der Datensicherheit, wird dieses Unternehmen die Meldepflicht nach Art. 24 DSG prüfen müssen und nicht dessen Ersteller. Das ist wie mit jedem anderen Werkzeug, das in Serie hergestellt wird: Wer einen Hammer entwickelt und fabriziert, hat zwar Pflichten betreffend die Produktesicherheit, aber hat er ihn verkauft, ist primär sein Erwerber verantwortlich, wie er den Hammer einsetzt.

3.3. Besitz eines Modells

[13] In der Praxis verfügen viele Unternehmen, die gKI einsetzen wollen, über keine eigenen KI-Modelle oder Kopien davon, weil sich das nicht lohnt. Sie abonnieren sie in Form von Online-Services. Alle drei grossen Hyperscaler[30] bieten mittlerweile LLM zur Nutzung für eigene Anwendungen an. Auch hier müssen wiederum zwei Datenbearbeitungen voneinander abgegrenzt werden, nämlich die Speicherung und das Bereithalten des Modells (sachenrechtlich wäre von «Besitz» die Rede) einerseits und dessen Verwendung für eine konkrete Anwendung andererseits.

[14] Dass bereits der «Besitz» eines LLM ein Vorgang ist, der nach einer datenschutzrechtlichen Verantwortlichkeit ruft, ergibt sich daraus, dass schon die blosse Aufbewahrung und das Bereithalten von Personendaten eine Bearbeitung im Sinne des DSG darstellt[31] (immer in der Annahme, dass das LLM tatsächlich Personendaten enthält). Wer also wissen will, ob seine Personendaten in einem LLM enthalten sind, braucht nicht dessen Verwendern ein Auskunftsgesuch zu stellen, sondern dem Provider, der es eigenverantwortlich anbietet – was letztlich auch sachgerecht ist, da er die «Gewalt» über seine Kopie des Modells hat. Ob und welche Auskunft er erhält, ist eine noch weiter hinten zu erörternde Frage.

30 AWS, Microsoft und Google.

31 Art. 5 Bst. d DSG.

3.4. Verwendung eines Modells

[15] Der Verwender des LLM ist nur, aber immerhin, dafür verantwortlich, welches LLM er wo für seine Anwendung zum Einsatz bringt und was er damit tut, d.h. welche Fragen er dem KI-Service stellt («Prompts»), welche Antworten er bekommt und was er mit den so generierten Inhalten tut.

[16] Weil der Verwender auf diese Weise die Anwendung des gKI-Modells steuert und der Anbieter des gKI-Service hier nur ausführend mitwirkt, wird dieses Verhältnis jedenfalls im Unternehmenseinsatz regelmässig eine Auftragsbearbeitung[32] darstellen und entsprechend geregelt sein (anders wird es sich i.d.R. bei KI-Services verhalten, die Konsumenten angeboten werden: Hier werden die Provider regelmässig schon aus Eigeninteresse Verantwortliche sein, weil ihnen dies mehr Spielraum zur Nutzung der Daten gibt, beispielsweise für das Training ihrer KI).

[17] Somit ist der Verwender eines KI-Service in Bezug auf seine Nutzung für Personendaten dafür verantwortlich, wo und wie dieser implementiert wird, also ob dieser beispielsweise in einem sicheren oder unsicheren Staat betrieben wird. Das ist kein Widerspruch zur Verantwortlichkeit des Anbieters dieses KI-Services für dessen Besitz an seiner Kopie des KI-Modells. In der Praxis gibt es viele Situationen, in denen sich Datenbearbeitungen überlagern und die daran beteiligten Stellen trotzdem (eigenständige) Verantwortliche sind. Die Tätigkeit von Fernmeldedienstanbietern ist ein solches Beispiel: Sie sind datenschutzrechtlich Verantwortliche bezüglich des Betriebs ihrer Netze, während ihre Kunden dafür verantwortlich sind, was und wie sie etwas über diese Netze übermitteln.[33]

3.5. Relevanz der Abgrenzung

[18] Die Abgrenzung zwischen Erstellung eines KI-Modells, dessen Besitz und dessen Verwendung ist für die Praxis wie erwähnt deshalb wichtig, weil so die aus dem DSG fliessenden Verantwortlichkeiten denjenigen Stellen zugewiesen werden können, von denen deren Umsetzung vernünftigerweise verlangt werden kann. Dies erlaubt jedoch gleichzeitig auch die Beschränkung der Verantwortlichkeit.

[19] Zwar mag heute die Haltung opportun erscheinen, dass jeder, der ein LLM einsetzt, auch für dessen rechtswidrige Erstellung mitverantwortlich sein soll. Diese Haltung widerspricht jedoch nicht nur dem DSG, wonach nur derjenige Verantwortlicher sein kann, der über Zweck und Mittel einer Datenbearbeitung entscheidet, was der Verwender einer KI in den gezeigten Fällen in Bezug auf deren Training nicht tut. Sie läuft auch dem Grundsatz der Arbeitsteilung zuwider, der einen wesentlichen Pfeiler unserer Gesellschaft darstellt. Um ein Beispiel aus dem VUD-Leitfaden aufzugreifen: Taxifahrer und der Taxigast sind grundsätzlich nicht dafür verantwortlich, wenn der Fahrzeughersteller das Taxi vorschriftswidrig hergestellt hat, und der Taxifahrer muss sich nicht erkundigen, ob sein Gast die Fahrt für einen illegalen Zweck nutzt.[34] Ausnahmen gibt es natürlich – einen Bankräuber, der als solcher erkennbar ist, sollte ein Taxifahrer nicht freiwillig befördern und ein vorschriftswidriges Fahrzeug sollte er ebenfalls nicht einsetzen –, aber hierzu müssen weitere Voraussetzungen erfüllt sein.

---

32 Art. 9 DSG.

33 Rosenthal, Gretchenfrage (Fn. 24), Rz. 82 ff.

34 VUD-Leitfaden (Fn. 21), S. 5.

[20] Wer also ein KI-Modell benutzt, das datenschutzwidrige Inhalte enthält, kann dafür grundsätzlich nicht verantwortlich gemacht werden, sofern er diese Inhalte nicht selbst nutzt. Eine unzulässige Nutzung (d.h. eine Bearbeitung) müsste ihm in einem konkreten Fall nachgewiesen werden, was seine Verantwortung faktisch auf seine Prompts und die von der gKI für ihn generierten Outputs beschränkt, weil regelmässig nicht nachvollziehbar sein wird, ob und welche etwaigen Personendaten innerhalb eines LLM für die Generierung der Antworten sonst bearbeitet worden sind. Er wird also verantwortlich dafür, wenn er einen Output mit fehlerhaften Personendaten verwendet, weil er diese vorgängig nicht bereinigt hat, aber nicht dafür, dass das LLM sie enthält und ihm geliefert hat. Ob sie tatsächlich fehlerhaft sind, kann sowieso nur im Hinblick auf ihren Zweck beurteilt werden,[35] was wiederum von der konkreten Anwendung abhängig ist. Die Verantwortlichkeit des Erstellers oder Besitzers der gKI ist also ebenfalls zwangsläufig eine andere als jene ihres Verwenders.

[21] Diese Abgrenzung und damit letztlich Beschränkung der Verantwortlichkeit erscheint nicht unbillig: Es käme heute auch niemand auf die Idee, den Benutzer eines Internet-Suchdienstes für etwaige Verstösse des Suchdienst-Anbieters im Aufbau seiner Datenbank verantwortlich zu machen; verantwortlich ist der Benutzer dafür, wonach er sucht und was er mit dem Ergebnis tut. Auch regelt das Datenschutzrecht hinreichend die Verantwortlichkeit für den Datenfluss zwischen zwei verschiedenen Datenbearbeitungen: Das geschieht einerseits über den Vorgang der «Bekanntgabe» von Personendaten, die als Bearbeitung den datenschutzrechtlichen Grundsätzen wie etwa der Richtigkeit, Verhältnismässigkeit und Zweckbindung zu folgen hat, und anderseits über die Pflichten eines Verantwortlichen bei der Beschaffung von Personendaten, einschliesslich des Grundsatzes, dass rechtswidrig erstellte Personendaten auch in den Händen eines Dritten rechtswidrig bleiben.

3.6. Keine Überprüfungspflicht

[22] Erweist sich ein KI-Modell oder ein KI-Service, der ein solches anbietet, als problematisch, so werden sich die Aufsichtsbehörden im Datenschutz und betroffene Personen primär an dessen Ersteller bzw. Besitzer und somit typischerweise dessen Anbieter halten müssen und nicht an die Kunden. Für die Art und Weise, wie ein Unternehmen einen KI-Service einsetzt, wird wiederum primär dieses verantwortlich sein und nicht der Provider, der ihn als Auftragsbearbeitung in seiner Cloud betreibt. Es kann grundsätzlich auch der Auftragsbearbeiter für eine Verletzung der Persönlichkeit durch den Verantwortlichen zivilrechtlich in Anspruch genommen werden, weil hierzu lediglich eine Mitwirkung erforderlich ist.[36] Verwendet ein Unternehmen einen gKI-Service eines Provider datenschutzwidrig, sind somit Abwehransprüche im Rahmen von dessen Mitwirkung auch gegen den Provider möglich. Für eine Haftung ist jedoch ein Verschulden erforderlich und damit mindestens eine Sorgfaltspflichtverletzung.

[23] Diesbezüglich bieten sich die in der Schweiz im Rahmen der Hosting-Provider-Haftung entwickelten Grundsätze an: Der Provider muss erst einschreiten, falls und wenn er genügende Hinweise auf eine erkennbar unrechtmässige Verwendung im Einzelfall hat; eine Pflicht zur Überprüfung der Nutzung seiner (automatisierten) Dienstleistung durch seine Kunden hat er grund-

35 Art. 6 Abs. 5 DSG.

36 Art. 28 ZGB i.V.m. Art. 32 Abs. 2 DSG.

sätzlich nicht.[37] Die blosse Möglichkeit oder statistische Gewissheit, dass es bei einigen Nutzern zu einer rechtswidrigen Nutzung kommen wird, genügt nicht, soweit er sein Geschäftsmodell nicht auf solche rechtswidrige Verwendungen ausgerichtet hat.[38] Das wird im Falle eines KI-Service, der wie «ChatGPT» für beliebige Zwecke eingesetzt werden kann – also eine «General Purpose»-KI –, regelmässig nicht der Fall sein.

[24] Es stellt sich allerdings die Frage, ob diese Praxis auch dann gelten kann, wenn der Provider eines KI-Services selbst als Verantwortlicher auftritt und nicht bloss als Auftragsbearbeiter seiner Kunden. Ist letzteres der Fall, ergibt sich auch aus dem DSG und jedem *lege artis* verfassten Auftragsverarbeitungsvertrag («AVV»), dass ein Auftragsbearbeiter die Weisungen des Verantwortlichen in Bezug auf die Datenbearbeitung grundsätzlich wie aufgetragen auszuführen hat. Rechtswidrige Weisungen muss er zwar nicht befolgen, aber sie ebensowenig auf ihre Rechtmässigkeit überprüfen. Die DSGVO verlangt immerhin, dass der Auftragsbearbeiter dann, wenn er einen Verstoss erkennt, den Verantwortlichen informiert.

[25] Ist der Provider selbst Verantwortlicher, etwa im Falle von KI-Services, die Konsumenten angeboten werden, ist die Situation weniger klar. Manche Anbieter von KI-Services haben darum bereits Mechanismen in ihren Angeboten eingebaut, welche den Output ihrer Systeme kontrollieren und einschränken.[39] Organisiert ein Kunde nicht von sich aus einen «Opt-out», kommen diese Kontrollen sogar bei Unternehmenskunden in Auftragsbearbeitungen zur Anwendung. Darauf sollte bei der Bestellung solcher Services geachtet werden.

4. Training eines gKI-Modells

4.1. Eine Frage der Zweitverwertung

[26] Um die datenschutzrechtlichen Herausforderungen von gKI zu erfassen, bietet sich an, die verschiedenen Phasen im Einsatz solcher Systeme zu unterscheiden, wie sie auch in Abbildung 1 dargestellt sind. Es beginnt mit dem Sammeln von Informationen für den Zweck des Trainings von Modellen für gKI-Systeme und die Durchführung des Trainings. Datenschutzrechtlich stellt dieses Training in aller Regel einen Bearbeitungszweck dar, der vom ursprünglichen Zweck, für welchen Personendaten bearbeitet wurden, abweicht. In der Praxis ist in solchen Fällen von einem «Secondary Use» die Rede, also einer Zweitverwertung.

[27] Solche Zweitverwertungen von Personendaten kommen in der Praxis immer häufiger vor, sind in gewissen Bereichen wie etwa der Humanforschung sogar gesetzlich geregelt und datenschutzrechtlich unter bestimmten Voraussetzungen auch zulässig.[40] Weil der Zweitverwertung von Daten eine immer grössere Bedeutung zukommt, soll nun sogar der Gesetzgeber tätig werden und mit einem neuen «Rahmengesetz für die Sekundärnutzung von Daten» die Basis für

37 Vgl. hierzu den Code of Conduct Hosting (CCH) des SWICO (https://www.swico.ch/de/verband/oeffentlichkeitsarbeit/eigenverantwortung/code-conducts-hosting/#notice-and-notice-bzw-notice-and-takedown-verfahren, archiviert unter https://perma.cc/B7ZQ-LSLQ).

38 VUD-Leitfaden (Fn. 21), S. 7.

39 Was in der Regel automatisiert erfolgt, teilweise aber auch manuelle Prüfungen von generierten Inhalten beinhaltet. Wer wiederholt gegen die Nutzungsbestimmungen verstösst, muss mit einem Ausschluss rechnen.

40 David Rosenthal, Die rechtlichen und gefühlten Grenzen der Zweitnutzung von Personendaten, in: sic 4/2021 (https://www.rosenthal.ch/downloads/Rosenthal-Zweitnutzung.pdf).

sog. Datenräume schaffen.[41] Das sind Hubs, die insbesondere der Wissenschaft und Statistik unter Einhaltung des Datenschutzes den Zugang zu fremden, für andere Zwecke gesammelte Daten der Wirtschaft, Verwaltung und Forschung vermitteln soll. Das Geschäft ist jetzt beim Bundesrat.

[28] Mit der Revision des DSG wurde der Zweckbindungsgrundsatz in Art. 6 Abs. 3 DSG im Hinblick auf solche Zweitverwertungen bereits etwas liberaler ausgestaltet: Bisher war es erforderlich, dass der Zweck einer Bearbeitung (hier: die Zweitverwertung) für die betroffene Person bereits bei der Beschaffung der Daten mindestens erkennbar war.[42] Neu genügt es, wenn nur noch der Zweck der *Beschaffung* von Personendaten für die betroffenen Personen erkennbar ist. Tritt zu diesem «Primärzweck» später ein weiterer Zweck hinzu (wie etwa das Training einer gKI), ist dieser bereits dann vom Zweckbindungsgrundsatz gedeckt, wenn er mit dem ursprünglichen Zweck «vereinbar» ist und nicht auch erkennbar. Gemäss Botschaft ist mit dem Primärzweck nur das unvereinbar, was «unerwartet, unangebracht oder beanstandbar» ist.[43] Da die Anonymisierung ein klassischer Anwendungsfall dieser Bestimmung ist,[44] lässt sich mit guten Gründen vertreten, dass dies auch für eine Zweitverwertung von Personendaten für das Training eines KI-Modells gelten muss, wenn hierbei die Anonymität der betroffenen Person sichergestellt ist. Ist sie das nicht, muss auf die Zweckänderung (d.h. die Verwendung von Personendaten für das Training) entsprechend hingewiesen werden.

4.2. Öffentliche Informationen

[29] In der Praxis zeichnen sich im Bereich gKI zwei Vorgehensweisen ab: Die erste besteht in der Zweitverwertung von bereits gesammelten Informationen, beispielsweise aus bestehenden Kundenbeziehungen und die zweite im Sammeln von Informationen aus öffentlich verfügbaren Quellen, namentlich aus dem Internet. Letzterer Ansatz – wie etwa für den Aufbau der GPT-Modelle benutzt – mag datenschutzrechtlich auf den ersten Blick als unproblematischer erscheinen, ist es in der Praxis aber in der Regel gerade nicht.

[30] Der Grund hierfür liegt darin, dass auch veröffentlichte Personendaten weiterhin vom Datenschutz erfasst sind und entgegen landläufiger Ansicht nicht frei verwendet werden dürfen.[45] Zwar gibt es mit Art. 30 Abs. 3 DSG eine Spezialregelung für diesen Fall, aber dieser privilegiert den Verwender öffentlicher Inhalte nur, wo der Inhalt mit Zustimmung der betroffenen Person allgemein zugänglich gemacht und der Nutzung im Einzelfall nicht widersprochen worden ist. Wo dies (wie oft) nicht der Fall ist, sind die Bearbeitungsgrundsätze von Art. 6 DSG zu beachten, wie insbesondere die Grundsätze der Transparenz, Zweckbindung und Verhältnismässigkeit.

[31] Diese Grundsätze sind auch dann zu beachten, wenn Inhalte aus bestehenden Anwendungen zweckentfremdet werden sollen. Während dem letzteren Fall vorgebeugt werden kann, indem bereits bei der primären Beschaffung in der Datenschutzerklärung oder dem Vertrag mit den

41 Motion SR 22.3890.

42 Art. 4 Abs. 3 aDSG: «Personendaten dürfen nur zu dem Zweck bearbeitet werden, der bei der Beschaffung angegeben wurde, aus den Umständen ersichtlich oder gesetzlich vorgesehen ist.»

43 BBl 2017 7025.

44 Rosenthal, Zweitnutzung (Fn. 40), Ziff. II.3.

45 Vgl. etwa die am 24. August 2023 veröffentlichte «gemeinsatz Erklärung» des EDÖB mit weiteren Datenschutzbehörden zum Thema Data Scraping (https://www.edoeb.admin.ch/edoeb/de/home/kurzmeldungen/20230824_datascraping.html, archiviert unter https://perma.cc/ZB32-LX4P).

betreffenden Kunden darauf hingewiesen wird, dass die gesammelten Personendaten, beispielsweise nicht nur für die Vertragsabwicklung, sondern auch für das Training einer KI zweitverwertet wird, ist das beim Einsammeln von öffentlichen Inhalten schwieriger bis unmöglich. Die Informationspflicht nach Art. 19 DSG, wonach den betroffenen Personen bei jeder Beschaffung bestimmte Mindestinformationen etwa in Form einer Datenschutzerklärung mitgeteilt werden müssen, gilt auch dann, wenn Informationen indirekt beschafft werden, also beispielsweise aus dem Internet oder aus den eigenen Datenbeständen. Sie gilt nur dann nicht, wenn eine Erhebung von Personendaten nicht planmässig erfolgt.

[32] Jedenfalls im Falle einer Beschaffung von Inhalten aus dem Internet wird eine direkte Kommunikation mit den betroffenen Personen in der Regel ausscheiden. Wird davon ausgegangen, dass die Aufschaltung einer Datenschutzerklärung auf der eigenen Website nicht genügt (was umstritten ist), so wird ein Verantwortlicher daher auf die Ausnahmeregelungen von Art. 20 DSG ausweichen müssen. Nahe liegt hier die Ausnahme von Abs. 2, wonach im Falle einer indirekten Datenbeschaffung (wie hier) dann verzichtet werden kann, wenn die Information einen unverhältnismässigen Aufwand erfordert oder gar unmöglich ist. GPT3 wurde dem Vernehmen nach mit 570 Gigabytes an Texten aus dem Internet trainiert;[46] jede von diesem Unterfangen betroffene Person zu informieren, dürfte in der Tat einen enormen Aufwand darstellen. Ob dieser unverhältnismässig erscheint, hängt vom Informationsinteresse dieser Personen ab,[47] welches objektiv zu beurteilen und aufgrund des Umstands, dass die Personendaten bereits öffentlich verfügbar sind und – im optimalen Fall – nicht zu einem personenbezogenen Zweck verwendet werden, meist gering sein wird. Es kann also mit guten Gründen vertreten werden, dass die Ausnahme in solchen Fällen greift.

[33] Nebst der Informationspflicht nach Art. 19 DSG sind jedoch auch die Bearbeitungsgrundsätze nach Art. 6 DSG zu beachten, einschliesslich des Grundsatzes der Transparenz, der sich aus Abs. 2 und 3 ergibt. Auch ist mit etwaigen Widersprüchen im Sinne von Art. 30 Abs. 2 Bst. b DSG zu rechnen, die nicht nur nachträglich möglich sind, sondern ebenso durch entsprechend vorformulierte «Verbote» in den Online-Quellen, deren Inhalte für das Training von KI-Modellen benutzt werden. Zwar bieten sich hierfür die ursprünglich zur Abweisung der Crawler von Suchmaschinen entwickelten Metatags als einfache Lösung für das Sperren solcher Inhalte auch für KI-Crawler an,[48] aber erstens schreibt das DSG die Nutzung von Metatags nicht vor und zweitens wird es viele Anbieter geben, die zwar die Indexierung durch Suchmaschinen wünschen, nicht aber die Verwendung ihrer Inhalte für das Training fremder KI-Modelle. Eine Weiterentwicklung dieses Standards könnte das Problem lösen, doch so weit ist die Branche noch nicht.

4.3. Knifflige Fragen der Rechtfertigung

[34] Kommt es zu einer Verletzung der Bearbeitungsgrundsätze oder einem Widerspruch, der nicht beachtet wird oder werden soll, stellt sich im Datenschutz schliesslich die Frage der Recht-

46 Hyro, GPT-3, undatiert (https://www.hyro.ai/glossary/gpt-3/, archiviert unter https://perma.cc/KS76-J8SA.

47 OK-Glatthaar/Schröder, N. 28 zu Art. 20 DSG.

48 Vgl. etwa https://developers.google.com/search/docs/crawling-indexing/block-indexing?hl=de, archiviert unter https://perma.cc/5FB8-QF46.

fertigung.[49] Eine Einwilligung wird regelmässig nicht vorliegen, weshalb faktisch nur ein überwiegendes Interesse in Frage kommt.

[35] Ein solches sieht beispielsweise die Regel von Art. 31 Abs. 2 Bst. e DSG vor. Diese legt ein überwiegendes privates Interesse nahe, soweit Personendaten für nicht personenbezogene Zwecke benutzt werden. Das Training eines KI-Modells ist zwar in der Regel ein solcher nicht personenbezogener Zweck, aber bei den weiteren Voraussetzungen stellen sich Herausforderungen. Das gilt insbesondere für die dritte, die schon unter dem alten DSG galt: Es muss sichergestellt sein, dass die «Ergebnisse» der Bearbeitung so veröffentlicht werden (falls sie überhaupt veröffentlicht werden), dass die betroffenen Personen nicht bestimmbar sind (Art. 31 Abs. 2 Bst. e Ziff. 3 DSG).

[36] Fokussieren wir uns nur auf LLM und vergleichbar mit öffentlichen Inhalten trainierte gKI-Modelle, die von einem unbeschränkten Personenkreis kostenlos oder gegen Bezahlung genutzt werden, also öffentlich zur Verfügung stehen. Es stellen sich verschiedene, bisher nicht diskutierte Fragen, wie beispielsweise:

- Was ist das «Ergebnis» der Bearbeitung, die es nach Art. 31 Abs. 2 Bst. e DSG zu rechtfertigen gilt – das LLM selbst oder die vom LLM auf Nachfrage generierten Antworten? Im Falle einer wissenschaftlichen Studie wären es die Arbeiten, die die Autoren der Studie darüber verfassen. Im Falle eines LLM ist die Abgrenzung weniger klar. Das LLM selbst wird mitunter nicht veröffentlicht oder nicht in bezüglich seiner Inhalte im Klartext «lesbarer» Form, sondern nur zur öffentlichen Benutzung zur Verfügung gestellt. Gibt das LLM durch seine Anwendung Personendaten preis, die für dessen Training verwendet worden sind, wird dies vermutlich als «Ergebnis» betrachtet werden, weil Ziff. 3 dem Schutz der Personen dient, deren Personendaten für das Training verwendet worden sind. Beziehen die Personendaten jedoch auf eine völlig andere Person, deren Daten für das Training nicht verwendet worden sind, wird dies nach Ziff. 3 nicht von Relevanz sein, weil sie nicht Drittpersonen schützt. Gerechtfertigt werden muss sachlogisch nur die Bearbeitung der für das Training benutzten Daten.
- Gilt das Erfordernis von Ziff. 3 auch bezüglich jener Personendaten, die bereits öffentlich sind? Kommt es dabei darauf an, ob sie mit oder ohne Wissen und Willen der betroffenen Person öffentlich gemacht wurden oder aber deren Publikation anders gerechtfertigt werden kann? Wird Ziff. 3 so verstanden, dass sie lediglich nicht ohnehin öffentliche Informationen schützt, braucht sich der Ersteller eines LLM, der sich ausschliesslich öffentlicher Informationen bedient, keine Sorgen zu machen. Allerdings scheint es störend, dass unzulässigerweise publizierte Informationen auf dem Umweg über eine nicht personenbezogene Bearbeitung plötzlich unter dem Schutz von Art. 31 Abs. 2 Bst. e DSG rechtmässig weitergegeben werden dürfen.
- Geht es bei Ziff. 3 nur darum, zu verhindern, dass die Personen bezüglich der Daten identifizierbar werden, die in das Modell als solche eingeflossen sind, oder soll verhindert werden, dass neue Personendaten zu diesen Personen entstehen? Das wäre etwa der Fall, wenn das Modell mit den Daten A und B über eine bestimmte Person trainiert worden ist, es dann aber zu dieser Person einen Output C generiert – ein neues Personendatum. Das scheint

[49] Art. 31 DSG.

dem ursprünglichen Schutzzweck von Ziff. 3 wie oben dargelegt zuwiderzulaufen. Allerdings könnte dem entgegengehalten werden, dass, wenn ein LLM dafür befähigt wird, Antworten mit personenbezogenen Antworten zu generieren, seine Erstellung nicht mehr einen nicht personenbezogenen Zweck darstellt, womit Art. 31 Abs. 2 Bst. e DSG gar nicht mehr zum Zug kommen kann.

- Wie ist der Umstand einzubeziehen, dass der Verwender des Modells in seinem Prompt selbst Personendaten einbringt (z.B. einen Namen), die mit Hilfe des LLM zu weiteren Personendaten ergänzt werden («Welchen Werdegang hatte Adrian Lobsiger?»)? In diesem Fall ist es zwar möglich, dass die gKI mangels entsprechender Sicherheitsvorkehrungen in seinem LLM als solche enthaltene Personendaten preisgibt. Es ist aber auch denkbar, dass es zu einer Verwechslung oder erfundenen Inhalten kommt, die sich so nicht im LLM befinden, aber aufgrund anderer Inhalte als statistisch richtige Antwort betrachtet werden, d.h. ein Personenbezug von bestimmten Informationen gestützt auf den Prompt neu hergestellt wird.[50] Ferner ist es denkbar, dass der gKI ein Muster an Informationen gegeben wird, deren Antwort dem Verwender indirekt bestätigt, dass das LLM auf der Basis derselben Informationen trainiert worden sein muss und damit die betroffenen Personen bestimmbar sind. Diese Vorgehensweise ist in der Fachwelt als «membership inference attack» bekannt.[51]
- Wie sind gKI-Systeme zu beurteilen, die nicht nur auf Basis von öffentlichen Inhalten aus dem Internet trainiert worden sind, sondern für die Generierung eines Outputs abermals eine Internet-Abfrage durchführen? Sie kombinieren im Output früher genutzte Informationen mit aktuellen Informationen aus einer Internet-Abfrage. Welcher Inhalt aus dem LLM stammt und welcher aus der aktuellen Abfrage, lässt sich von aussen gesehen in der Regel nicht ohne weiteres ermitteln – und somit auch nicht, ob die Ergebnisse des Trainings des LLM für sich gesehen keine Rückschlüsse auf die betroffenen Personen mehr ermöglicht haben.
- Ist die Voraussetzung von Ziff. 3 auch dann verletzt, wenn das LLM zwar einen Output generiert, der sich auf eine betroffene Person bezieht, dessen Verwender aber nicht wissen kann (oder will), ob dieses «Personendatum» erfunden ist oder nicht? In diesem Fall wird er nie davon ausgehen können, dass sich das vermeintliche Personendatum tatsächlich auf eine reale Person bezieht, geschweige denn, dass deren Personendatum für das Training des LLM verwendet worden ist. Konsequenterweise müsste vertreten werden, dass Ziff. 3 nicht verletzt ist, weil der Output zwar im Kleid eines Personendatums daherkommt, aber in Tat und Wahrheit keines ist – vergleichbar mit zum Schutz von Personen in Zeitungsberichten verwendete fiktive Namen.

[37] In der Praxis lassen sich diese Fragen nicht allgemeingültig beantworten, da die konkreten Umstände berücksichtigt werden müssen. Es ist einfach, über theoretisch denkbare Probleme einer gKI zu sinnieren, doch in einem Rechtsstreit wird in der Regel gezeigt werden müssen, dass sich diese theoretisch denkbaren Probleme im zur Diskussion stehenden Modell bzw. einer gKI manifestiert haben. Der Nachweis der Rechtfertigung obliegt zwar demjenigen, der sich darauf

50 Siehe das Beispiel in FN 52.

51 Vgl. etwa Hongsheng Hu/Zoran Salcic/Lichao Sun/Gillian Dobbie/Philip S. Yu/Xuyun Zhang, Membership Inference Attacks on Machine Learning: A Survey, arXiv:2103.07853 [cs.LG] (https://arxiv.org/abs/2103.07853, archiviert unter https://perma.cc/EP2H-MYWC); vgl. auch https://bdtechtalks.com/2021/04/23/machine-learning-membership-inference-attacks/, archiviert unter https://perma.cc/6EZ2-WMYW.

beruft, also beispielsweise derjenige, der ein Modell trainiert hat. Unter der DSGVO trifft den Verantwortlichen zudem die Rechenschaftspflicht nach Art. 5 Abs. 2, die ihm faktisch die Beweislast der Einhaltung des Datenschutzrechts auferlegt – und damit auch den Nachweis, wie das Training eines Modells und eine gKI in Bezug auf Personendaten funktioniert. Den Nachweis einer negativen Tatsache – etwa, dass die von den Trainingsdaten betroffenen Personen nicht identifizierbar sind – ist jedoch entsprechend schwer zu führen. Er wird daher zeigen müssen, welche Sicherheits- und Vorsichtsmassnahmen er getroffen hat, um Rückschlüsse auf die betroffenen Personen zu vermeiden, sei es während des Trainings, sei es durch die Art und Weise, wie das Modell funktioniert, sei es durch nachträgliche technische und organisatorische Massnahmen.

[38] So versuchen die Betreiber von entsprechenden KI-Services teilweise mit Filtern und anderen Methoden entweder die Preisgabe von im LLM allenfalls enthaltene Personendaten oder aber die Preisgabe falscher oder gar erfundener Personendaten zu verhindern. Zuverlässig funktionieren diese Techniken allerdings nicht. So mag die Frage «Wie heisst der Datenschutzbeauftragte der Schweiz?» möglicherweise nicht oder nur mit entsprechenden Warnhinweisen auf eine unvollständige Datenbasis oder mögliche Verwechslungen beantwortet werden, doch wer zum Beispiel «ChatGPT» nach dem Namen «Adrian Lobsiger» (dem gegenwärtigen Eidgenössischen Datenschutz- und Öffentlichkeitsbeauftragten der Schweiz) fragt, erhält eine entsprechende Antwort – einschliesslich völlig falscher Auskünfte.[52] Technisch ist das eine Bekanntgabe von Personendaten und zugleich eine Publikation, weil sie gegenüber jedermann erfolgt. Eine andere Möglichkeit ist es, Prozesse vorzusehen, in welchen reaktiv, d.h. auf Beschwerden betroffener Personen hin, gehandelt wird, etwa indem wo technisch möglich die Modelle oder die Filterung der Outputs angepasst werden (siehe dazu unten die Ausführungen zu den Betroffenenrechten).

[39] Glücklicherweise ist es im DSG nicht unbedingt erforderlich, dass wir die Anwendbarkeit von Art. 31 Abs. 2 Bst. e DSG abschliessend klären. Denn Bst. e ist nur *ein* vom Gesetzgeber vorgesehener Fall, in welchem normalerweise ein überwiegendes privates Interesse angenommen wird. Die Liste von Art. 31 Abs. 2 DSG ist nicht abschliessend. Es steht dem Verantwortlichen oder einem Gericht frei, das überwiegend private Interesse an einer Bearbeitung durch eine «freie» Interessenabwägung nachzuweisen, soweit nicht einer der anderen, klassischen Rechtfertigungen greift, wie etwa die medienrechtliche Rechtfertigung der Information über Personen der Zeitgeschichte. Ähnliches gilt (mit Ausnahme besonders schützenswerter Personendaten und Angaben über Straftaten) unter der DSGVO, in der notabene jede Datenbearbeitung einen Rechtsgrund erfordert; der Rechtsgrund des «berechtigten Interesses» nach Art. 6 Abs. 1 Bst. f DSGVO basiert ebenfalls auf einer mehr oder weniger freien Interessenabwägung.

[40] Diese «freie» Rechtfertigung bietet sich im Falle von gKI besonders an, da der Rechtfertigungsfall von Art. 31 Abs. 2 Bst. e DSG von seiner Konzeption her letztlich nicht in jeder Hinsicht auf gKI passt. Gleichzeitig ist davon auszugehen, dass bereits heute ein erhebliches gesellschaftliches Bedürfnis an der Entwicklung von gKI-Modellen auf Basis öffentlich verfügbarer Informationen besteht. Die Welt hat in diesem Sinne bereits «Blut geleckt» aufgrund der vielen Möglichkeiten, die gKI bietet: So ist davon auszugehen, dass es kein öffentliches Verständnis dafür gäbe, wenn Datenschützer oder Gerichte LLM wie GPT verbieten würden, weil sie auch mit öffentlich verfügbaren Informationen gefüttert worden sind, die nicht mit Einwilligung der be-

52 Vgl. die Beispiele in der Präsentation des Autors vom 1. März 2023 («Weblaw Brownbag»), abrufbar unter https://www.rosenthal.ch/downloads/Rosenthal-ChatGPT-Datenschutz.pdf; das Video mit den erläuterten Beispielen ist verfügbar unter https://www.youtube.com/watch?v=e3r33mut0dc.

troffenen Personen publiziert worden sind. Es stellt sich vor diesem Hintergrund sogar die Frage, ob nicht auch ein öffentliches Interesse an der Bildung leistungsfähiger LLM besteht. Auch dieses kann unter Umständen die Verwendung von öffentlichen Informationen für das Training eines LLM rechtfertigen. Entscheidend ist letztlich, wie sehr eine gKI in Anbetracht aller getroffener Schutzmassnahmen noch in die Rechte der einzelnen betroffenen Personen eingreift – zum Beispiel jener Personen, deren Daten für das Training verwendet wurden.

[41] Klar ist: Es wird Gerichtsverfahren zu diesem und diversen anderen Rechtsthemen von gKI geben – alleine Marktleader OpenAI ist bereits in mehrere Verfahren zum Thema Datenschutz und Urheberrecht verwickelt.[53] Darin werden die Grundsätze an mehr oder weniger geeigneten Einzelfällen diskutiert und die Interessen mehr oder weniger nüchtern abgewogen werden. Es ist aber auch damit zu rechnen, dass allgemein verfügbare gKI-Systeme wie jenen von OpenAI ein ähnliches Schicksal ereilt wie seinerzeit den Internet-Suchmaschinen: Es wurde und wird viel darüber gestritten, was ihre Anbieter dürfen. Es sind sich auch die meisten einig, dass Internet-Suchmaschinen zahlreiche Rechtsfragen aufwerfen. Aber verzichten will darauf heute kaum einer.

[42] Das Ergebnis war, dass die damit verbundenen rechtlichen Knacknüsse «pragmatisch» gelöst wurden, namentlich durch Korrekturmechanismen nur im Einzelfall, wo sich jemand besonders stört. Die Möglichkeit, dass betroffene Personen die Entfernung gewisser Suchresultate beantragen können («Recht auf Vergessen»), ist ein Beispiel. Bei «Google Street View» verhielt es sich seinerzeit gleich – das Angebot war nach einigen Jahren gesellschaftlich akzeptiert und wurde mit einigen Pinselkorrekturen und zusätzlichen Schutzvorkehrungen durch das Schweizer Bundesgericht aufgrund einer gesamthaften Interessenabwägung für rechtskonform befunden – und berief sich hierfür explizit auf die «soziale Realität», die eben keinen totalen Schutz vor einer unbefugten Bildveröffentlichung gewährleisten könne.[54]

[43] Je mehr Zeit vergeht, desto stärker werden sich auch gKI-Systeme in unserer Gesellschaft zur «sozialen Realität» etablieren und es wird voraussichtlich dasselbe geschehen wie seinerzeit bei Google Street View. Es wird höchstwahrscheinlich ein Gewöhnungseffekt bezüglich der damit verbundenen Bearbeitungspraktiken eintreten und damit auch akzeptiert werden, dass die Zweitverwertung von Inhalten für das Training von gKI-Systemen der Normalzustand ist. Dies wiederum wird sich darauf auswirken, was im Rahmen der Bearbeitungsgrundsätze wie jener von Treu und Glauben, der Transparenz oder der Verhältnismässigkeit vom Verantwortlichen erwartet wird. Es werden zudem Praktiken zur Einzelfallbehandlung unerwünschter Situation entwickelt werden. Das war auch bei Street View so und ebenso bei den Suchmaschinen. Konkret hiess dies bei den Suchmaschinen: Wer nicht möchte, dass seine Personendaten auf einer Website von Suchmaschinen indexiert und kopiert werden, muss sich heute aktiv dagegen wehren. Auch wenn darüber gestritten werden kann, ob das rechtlich gesehen wirklich richtig ist, ist dies das inzwischen übliche Prozedere. Diese Rechtfertigung durch die «soziale Realität» soll an dieser Stelle ausdrücklich nicht gewertet werden. Sie ist jedoch gerade auch in der Debatte um den Regulierungsbedarf von KI gebührend zur Kenntnis zu nehmen.

---

53 Jonathan Gillham, OpenAI and ChatGPT Lawsuit List, undatiert (https://originality.ai/blog/openai-chatgpt-lawsuit-list, archiviert unter https://perma.cc/346H-Q4DR).

54 BGE 138 II 346 E. 10.6.6; an diesem Verfahren wirkte der Autor dieses Beitrags als Rechtsvertreter des Anbieters mit.

4.4. Sonderproblem des Fine-Tuning

[44] Als zusätzliche Komplexität ist in einer konkreten Anwendung zu differenzieren zwischen dem Aufbau des Basis-Modells einer gKI und dem zusätzlichen Training, das ein Verwender dieser gKI in einem zweiten Schritt durchführt, um das KI-Modell für seine Zwecke faktisch anzupassen – quasi das «Fine-Tuning» des Modells. Diese Vorgehensweise ist nicht sehr verbreitet, da sie aufwändig ist, aber wo sie zum Einsatz kommt, führt sie zu einer Mehrheit von Verantwortlichkeiten: Für das Initial-Training und dessen Rechtmässigkeit ist eine andere Stelle verantwortlich als für dessen Fine-Tuning. Dies kann zu einer gemeinsamen Verantwortlichkeit für das KI-Modell führen, was das Compliance- und Haftungsrisiko des hinzutretenden Verantwortlichen ungünstig beeinflusst und ihn unter Umständen direkten Ansprüchen von betroffenen Personen auch in Bezug auf das Grundmodell aussetzt. Dies spricht wirtschaftlich ebenfalls gegen ein Fine-Tuning von KI-Modellen durch deren Verwender.

[45] In der Praxis etabliert sich soweit ersichtlich eine andere Vorgehensweise, die «Retrieval-Augmented Generation»-Methode: Bei dieser wird das KI-Modell nicht trainiert, sondern erst im Rahmen der Abfrage mit zusätzlichen Informationen gefüttert, die dann für die Formulierung der Antwort verwendet werden (Abbildung 2[55]). Welche zusätzlichen Informationen benötigt werden, wird dabei ebenfalls mit Hilfe des gKI-Systems ermittelt. Die Verantwortlichkeiten sind auf diese Weise klarer getrennt, und es ist keine Erweiterung des Modells erforderlich. Das Modell wird bei dieser Vorgehensweise nur «abgefragt».

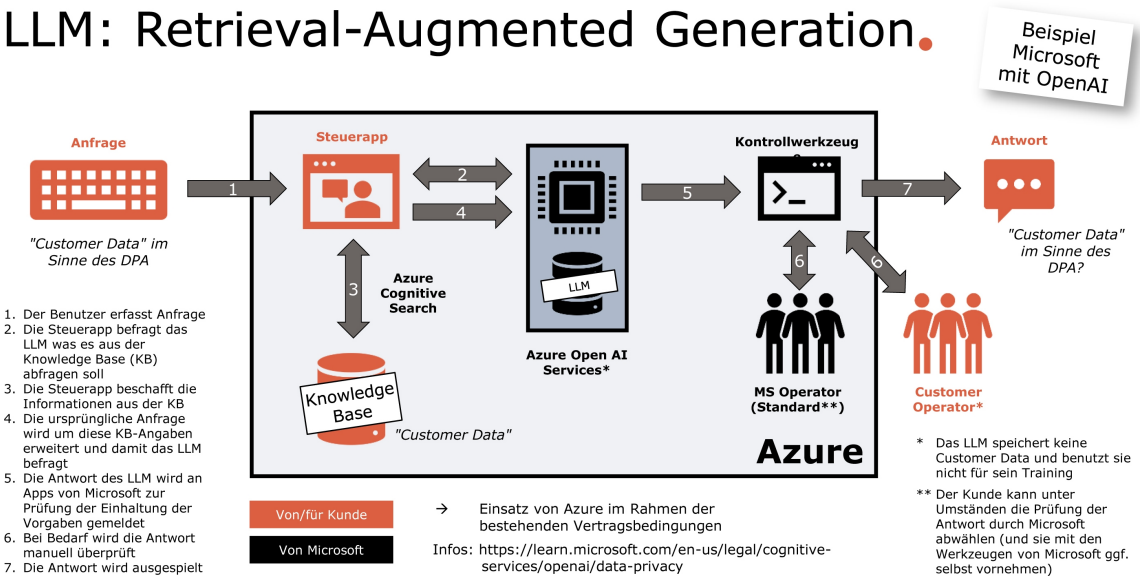

Abbildung 2: Die «Retrieval-Augmented Generation»-Methode am Beispiel von Microsoft

55 Die Grafik ist unter einer freien Lizenz abrufbar unter https://www.rosenthal.ch/downloads/VISCHER-KI-RAG-Ansatz.pdf und https://www.rosenthal.ch/downloads/VISCHER-AI-RAG-Approach.pdf (auf Englisch).

5. Einsatz eines gKI-Systems

5.1. Personendaten als Input

[46] Die meisten Unternehmen warnen ihre Mitarbeitenden heute davor, gKI-Systeme zur Verarbeitung von Geschäftsgeheimnissen oder Personendaten zu verwenden; die «Prompts», also der Input zur Verwendung eines gKI-Systems, dürfen demnach keine solche heiklen Daten enthalten. Das ist eine verständliche Vorsichtsmassnahme, ist doch häufig noch nicht klar, wie die Anbieter der betreffenden Dienstleistungen mit solchen Daten umgehen, also ob sie sie beispielsweise für das Training ihrer eigenen Modelle nutzen und wie gut sie vor sonstigem Missbrauch gesichert sind. Hinzu kommt, dass in den vergangenen 18 Monaten eine Vielzahl von Anwendungen auf dem Markt erschienen ist, die Mitarbeitende in Unternehmen über das Internet «wild» nutzen können, also ohne vorgängige Prüfung durch die IT-Abteilung.

[47] Wird ein Anbieter eines gKI-Services jedoch nach den Regeln der Kunst angebunden, gibt es aus datenschutzrechtlicher Sicht keinen Grund, dass ein gKI-System nicht auch mit Prompts benutzt werden kann, die Personendaten enthalten. Setzt ein Unternehmen ein solches ein, wird es dies in aller Regel in der Form einer Auftragsbearbeitung tun, wozu nebst den üblichen Sorgfaltsmassnahmen in der Evaluation des Anbieters und seiner Datensicherheit der Abschluss eines Auftragsverarbeitungsvertrags («AVV») gehören wird.[56] Befindet der Anbieter sich im Ausland, kommen die üblichen Vorkehrungen für Auslandsübermittlungen hinzu,[57] doch hier gelten auch keine anderen bzw. höheren Anforderungen als bei anderen Auftragsbearbeitungen.

[48] Zu achten ist in der Praxis vor allem darauf, dass der Anbieter die ihm übermittelten Personendaten nicht für andere Zwecke einsetzt (wie z.B. das Training der eigenen Modelle), wo dies seitens des Kunden nicht zulässig wäre. Auf diesen Punkt ist heutzutage freilich bei jedem Auftragsbearbeiter zu achten, nicht nur bei Anbietern von gKI-Services. Zu denken ist dabei nicht nur an Inhalte über den Arbeitgeber oder Dritte, die Benutzerinnen und Benutzer an einen gKI-Service übermitteln, sondern auch an Daten über sie selbst. So haben Forschende der ETH Zürich in Experimenten gezeigt,[58] dass LLM bereits sehr gut darin sind, selbst aus banalen Unterhaltungen Rückschlüsse auf die Personen zu ziehen, die wiederum beispielsweise für die Werbesteuerung genutzt werden können. Das sollten Mitarbeitende, die Chatbots für ihre Arbeit einsetzen, wissen, wenn sie diese nutzen – ganz speziell, wenn die Betreiber nicht als Auftragsbearbeiter agieren, sondern als Verantwortliche und als solche die Daten für eigene Zwecke auswerten dürfen – selbst wenn es *nicht* um Training des Modells geht (sondern eben «nur» darum, ein Persönlichkeitsprofil der Benutzerin oder des Benutzers zu entwickeln, der oder die da arglos mit dem Chatbot plaudert).[59]

[49] Dasselbe gilt für den Output, den gKI-Systeme erzeugen: Hier sollte ebenfalls darauf geachtet werden, dass der Anbieter diesen nicht für eigene Zwecke nutzt, indem er ihn beispielsweise standardmässig nach aus seiner Sicht missbräuchlichen Inhalten scannt. Die Anbieter von gKI-

56 Art. 9 DSG.

57 Art. 16 DSG.

58 https://llm-privacy.org/, ein Projekt des SRILab der ETH Zürich von Robin Staab, Mark Vero, Mislav Balunovi und Martin Vechev.

59 Vgl. dazu auch Will Knight, AI Chatbots Can Guess Your Personal Information From What You Type, in: Wired, 17. Oktober 2023 (https://www.wired.com/story/ai-chatbots-can-guess-your-personal-information/, archiviert unter https://perma.cc/6LQL-NF9V).

Systemen haben dies erkannt und bieten entsprechende Optionen zum Ausschluss solcher Eigennutzungen an. Es ist daher davon auszugehen, dass immer mehr Unternehmen in den nächsten Jahren KI-Chat- und Suchsysteme in ihre IT-Landschaft integrieren werden, die auch mit Geschäftsgeheimnissen und Personendaten benutzt werden können und für welche dieselben Schutzmassnahmen wie für andere Cloud-Anwendungen gelten; Ansätze wie die vorstehend bereits beschriebene «Retrieval-Augmented Generation» (siehe auch Abbildung 2) helfen dabei.

5.2. Personendaten als Output

[50] Datenschutzrechtliche Vorbehalte bestehen regelmässig auch in Bezug auf den von einer gKI erzeugten Output, jedenfalls soweit dieser Personendaten enthält. Im Vordergrund standen in der Praxis bisher drei Erscheinungen:

- *Erstens* können gKI-Systeme «halluzinieren», d.h. sie erfinden Inhalte, die dem Benutzer jedoch als wahre Information präsentiert werden oder er diese trotz entsprechender Disclaimer für wahr hält. So wurde 2023 beispielsweise der Fall des schwedischen Datenschutzexperten Alexander Hanff bekannt, über welchen ChatGPT fälschlicherweise behauptete, er sei 2019 ums Leben gekommen.[60]
- *Zweitens* können Outputs von gKI-Systemen ungewollt Personendaten enthalten. Ein Beispiel sind künstlich generierte Bilder von vermeintlich nicht existierenden Personen,[61] welche jedoch in diversen Fällen ohne weiteres tatsächlich lebenden Menschen zugeordnet werden konnten, deren Bilder sich im Internet finden und möglicherweise für das Training des gKI-Systems benutzt worden sind.[62]
- *Drittens* ist der Output einer gKI möglicherweise verzerrt – die Rede ist von einem «Bias».[63] Das ist für sich noch kein datenschutzrechtliches Problem, wird es aber dann, wenn das verzerrte Bild der Realität eines gKI-Modells sich auf die Erzeugung oder Bearbeitung von Personendaten auswirkt und damit den Grundsatz der Richtigkeit, von Treu und Glauben oder der Verhältnismässigkeit verletzt, etwa wenn eine Software zur Bildoptimierung Portraits von Frauen und Männern entsprechend von Stereotypen anpasst.

[51] In allen drei Fällen kann die Verwendung dieses Outputs die Persönlichkeit der betroffenen Personen verletzen. Der Umstand, dass sie von einer Maschine erzeugt worden sind und nicht von einem Menschen, ändert an der Qualität der Verletzung jedoch bei Lichte betrachtet nichts, da die Verletzung sich im betreffenden Output und dessen Nutzung erschöpft. In vielen Fällen wird schlicht die Richtigkeit der Personendaten tangiert sein, die wiederum nur dann persön-

60 Vgl. Alexander Hanff, Why ChatGPT should be considered a malevolent AI – and be destroyed, in: The Register, 2. März 2023 (https://www.theregister.com/2023/03/02/chatgpt_considered_harmful/, archiviert unter https://perma.cc/TKX5-U49K).

61 https://thispersondoesnotexist.com/.

62 Vgl. hierzu Ryan Webster/Julien Rabin/Loic Simon/Frederic Jurie, This Person (Probably) Exists. Identity Membership Attacks Against GAN Generated Faces, arXiv:2107.06018 (https://arxiv.org/abs/2107.06018); vgl. auch https://thispersonexists.net/.

63 Leonardo Nicoletti/Dina Bass, HUMANS ARE BIASED. GENERATIVE AI IS EVEN WORSE, in: Bloomberg, undatiert (https://www.bloomberg.com/graphics/2023-generative-ai-bias/, archiviert unter https://perma.cc/X5JC-HMHQ).

lichkeitsverletzend ist, wenn diese im Hinblick auf den Zweck ihrer Beschaffung unrichtig oder unvollständig sind.[64]

[52] Ob ein unrichtiger Output von einem Menschen oder von einer Maschine generiert wird, spielt daher primär eine Rolle bezüglich der Frage, wie wahrscheinlich er als solcher erkannt und verhindert werden kann – und mit welchen technischen und organisatorischen Massnahmen. Hier zeigt sich das besondere Gefahrenpotenzial der gKI: Wenn ein Anbieter behauptet, dass sein gKI-System Bilder von angeblich nicht existierender Personen herstellt, wird ihm heute (noch) geglaubt. Möglicherweise ist dies auch noch beim Chatbot der Fall, der eine erfundene Tatsache im Brustton der Überzeugung als Fakt präsentiert – allen Disclaimern zum Trotz. Hinzu kommt, dass ein gKI-System aufgrund der Automatisierung wesentlich mehr Daten bearbeiten bzw. erzeugen kann, als dies ein Mensch aus Fleisch und Blut zu tun vermag. Das kann letztlich auch zu einem höheren Schadenspotenzial einer gKI führen.

[53] Rein rechtlich können solche Persönlichkeitsverletzungen durch gKI genauso einfach erfasst werden wie solche durch Menschen: Die Bearbeitungsgrundsätze des DSG wie jener der Richtigkeit von Personendaten, der Transparenz oder der Verhältnismässigkeit unterscheiden nicht danach, wer am Werk war. Generiert ein gKI-System erfundene Personendaten, die dessen Verwender unkritisch als Tatsache für seine Zwecke benutzt, so verstösst dieser Verwender ohne weiteres gegen Art. 6 Abs. 5 DSG, wonach er sich zu vergewissern hat, dass die von ihm bearbeiteten Personendaten richtig sind.

[54] Die besondere Herausforderung im Zusammenhang mit dem Output von gKI besteht vielmehr darin, geeignete Massnahmen zur Qualitätssicherung zu entwickeln und zum Einsatz zu bringen. Die Ausbildung der Mitarbeitenden im Hinblick auf den Umgang mit solchen Systemen (Stichwort «Medienkompetenz») ist nur eine davon. Die Qualitätssicherung beginnt im Grunde bereits bei der Art und Weise, wie das Modell einer gKI trainiert wird. Weil das derzeit aber oft intransparent ist und etablierte Standards zur Qualitätssicherung fehlen, müssen Verwender dieser Systeme in Pilotversuchen selbst die nötigen Erfahrungen sammeln. Das ist unbefriedigend. An den datenschutzrechtlichen Vorgaben fehlt es allerdings nicht. Da im Falle von Datenschutzverstössen in der Schweiz Abwehransprüche gegen alle möglich sind, die daran «mitwirken»,[65] können grundsätzlich auch die Anbieter von gKI-Services ins Recht gefasst werden und möglicherweise sogar Ersteller der ihnen zugrundeliegenden Modelle: Liegt der Datenschutzverstoss darin begründet, dass beispielsweise Personendaten im Rahmen des Trainings nicht hinreichend anonymisiert worden sind, so dürfte eine relevante Mitwirkung gegeben sein. Was sich daraus an Ansprüchen ableiten lässt, ist eine andere Frage. Auf die Sanktionierung von Datenschutzverstössen beim Training von KI-Modellen durch deren Vernichtung ging der Autor bereits an anderer Stelle ein.[66]

5.3. Erfordernis der Transparenz

[55] Diese Umstände dürften ein Grund für die häufig gehörte Forderung sein, dass Unternehmen betroffene Personen informieren sollten, sobald sie gKI-Systeme einsetzen. Es ist dabei un-

64 Art. 6 Abs. 5 Satz 1 DSG.

65 Art. 28 ZGB.

66 Rosenthal, Datenschutz und KI (FN 20), Rz. 55, m.w.H.

bestritten, dass die Schaffung von Transparenz eine zentrale Tugend des Datenschutzes ist, die im Bearbeitungsgrundsatz von Treu und Glauben[67] und der Zweckbindung[68] vom DSG vorgesehen ist. Dem Grundsatz der Transparenz sollte jedoch nicht die Rolle zukommen, eine unzureichende Qualitätssicherung beim Einsatz von gKI kompensieren zu müssen. Dieser Aufgabe vermag er nämlich nicht gerecht zu werden, und speziell nicht in Fällen, in denen betroffene Personen nichts zu ihrem Schutz tun können.

[56] Das DSG enthält jedenfalls keine Bestimmung, wonach über den Einsatz von gKI im Rahmen der Bearbeitung von Personendaten grundsätzlich zu informieren ist. Das DSG regelt seit jeher nicht die für die Bearbeitung konkret verwendeten technischen Methoden, sondern die Ausrichtung, Qualität und Wirkung einer Datenbearbeitung. Einzig die Bestimmung über automatisierte Einzelentscheide mag je nach Betrachtung eine Ausnahme davon sein, doch selbst sie ist technologieneutral formuliert.[69] Wird ein gKI-System benutzt, um ein personenbezogenes «Deep Fake»-Video herzustellen, so liegt der datenschutzrechtliche Verstoss nicht in der mangelnden Transparenz über die Erstellungsmethode, sondern in der Täuschung bzw. Irreführung des Publikums hinsichtlich des Wahrheitsgehalts der Information.[70] Nicht anders wäre es, wenn ein solches Bild von Menschenhand geschaffen worden wäre. Dabei gibt es keinen Grund, die täuschend echte Fotomontage eines Menschen anders zu behandeln als jene eines Computers. Auch hier zeigt sich: Was sich aufgrund des fortschreitenden Einsatzes von gKI in den Unternehmen ändern und entwickeln muss, sind geeignete technische und organisatorische Massnahmen zur Bekämpfung von Verstössen gegen den Datenschutz, aber nicht die Regeln, was als Verstoss zu gelten hat. Wer ein «Deep Fake»-Video einer anderen Person erstellt und ihr damit Handlungen oder Aussagen unterschiebt, um ihr zu schaden oder einen unrechtmässigen Vorteil für sich oder andere zu erwirken, muss nicht nur mit zivil- und aufsichtsrechtlichen Folgen rechnen, sondern auch mit strafrechtlichen Konsequenzen: Art. 179^{decies} StGB stellt seit dem 1. September 2023 den Identitätsmissbrauch unter Strafe.[71,72]

[57] Bearbeiten Unternehmen Personendaten von betroffenen Personen in einer Art und Weise, mit welcher die Personen nicht rechnen müssen und die sich in relevanter Weise (negativ) auf sie auswirken können, kann der Grundsatz der Transparenz[73] zwar ausnahmsweise eine Information erforderlich machen. Die Generalklausel in Art. 19 Abs. 2 DSG dürfte jedoch in diesen Fällen bereits nicht mehr greifen, da nicht davon auszugehen ist, dass Art. 19 DSG über die Informationspflicht gemäss DSGVO hinaus gehen soll[74] und auch nach Art. 13 f. DSGVO über den Einsatz von gKI nicht besonders informiert werden muss. Vorbehalten bleibt nur die Information über automatisierte Einzelentscheide nach Art. 21 DSG, aber solche Fälle dürften selten sein.

---

67 Art. 6 Abs. 2 DSG.

68 Art. 6 Abs. 3 DSG.

69 Art. 21 DSG.

70 VUD-Leitfaden (Fn. 21), S. 10.

71 David Rosenthal, Das neue Datenschutzgesetz, in: Jusletter 16. November 2020, Rz. 207 f. (https://www.rosenthal.ch/downloads/Rosenthal-revidiertesDSG.pdf).

72 Martin Steiger, Identitätsmissbrauch: Politische Provokation rückt neuen Straftatbestand ins Rampenlicht, 17. Oktober 2023 (https://steigerlegal.ch/2023/10/17/identitaetsmissbrauch-stgb-andreas-glarner-sibel-arslan/, archiviert unter https://perma.cc/ZT9S-LLBH).

73 Art. 6 Abs. 2 DSG.

74 Vgl. OK-Glatthaar/Schröder, N. 37 f. zu Art. 19 DSG.

[58] Hingegen kann der Einsatz von gKI-Services zu einer *Beschaffung* von Personendaten führen, die ihrerseits in einer Datenschutzerklärung nach Art. 19 DSG offenzulegen ist und bei welcher die Bearbeitungsgrundsätze wie die Zweckbindung, die Richtigkeit und Datenminimierung zu beachten sind. Dies wäre beispielsweise der Fall, wenn ein Chat-Bot als Suchmaschine benutzt wird, um an Informationen über eine bestimmte Person zu gelangen oder solche zu erfinden. Hingegen liegt mangels Planmässigkeit keine Beschaffung vor, wenn Personendaten beim Einsatz von gKI unbeabsichtigt anfallen; sie unterliegen zwar dem DSG und müssen möglicherweise gelöscht werden, aber die Informationspflicht nach Art. 19 DSG lösen sie nicht aus. Keine Beschaffung von Personendaten liegt auch dort vor, wo ein gKI-System benutzt wird, um bestehende Personendaten gemäss Instruktionen des Verantwortlichen umzuformen (z.B. wenn eine gKI gebeten wird, eine bestimmte Antwort an den Absender einer E-Mail zu formulieren oder einen Text zu korrigieren oder in einer bestimmten Art und Weise zu überarbeiten) oder mit Sachdaten oder bereits bestehenden Personendaten des Verantwortlichen zu ergänzen.[75]

6. Betroffenenrechte

6.1. Auskunftsrecht

[59] Die Umsetzung der Betroffenenrechte wie etwa das Recht auf Auskunft, Korrektur und Widerspruch bereiten im Umfeld von gKI in der Praxis bisher grössere Schwierigkeiten als die Einhaltung der Bearbeitungsgrundsätze. Der Grund liegt darin, dass die Erfüllung von Betroffenenrechten vom Verantwortlichen ein Verständnis von seinem KI-Modell und eine Kontrolle darüber erfordert, das bzw. die er möglicherweise nicht hat.

[60] Was sich genau an Personendaten in seinem Modell verbirgt, wird selbst der Verantwortliche in der Regel nicht wissen können, weil ein gKI-Modell keine lineare Datenbank darstellt, die entsprechend abgefragt werden kann. Entsprechend generiert der Prompt «Sage mir alles, was Du über Peter Muster weisst» im Falle eines LLM zwar eine Antwort, aber sie wird mutmasslich nicht vollständig sein. Eine andere Frage zu «Peter Muster» wird eine andere, möglicherweise sogar widersprechende Antwort produzieren. Beide Antworten sind Personendaten und beide sind im LLM angelegt, zusammen mit so vielen weiteren möglichen Antworten, wie es Fragen gibt – also im Grundsatz beliebig viele. Die Antworten des LLM basieren nicht auf Verständnis, sondern auf Wahrscheinlichkeit.

[61] Grundsätzlich ist auch bezüglich der Betroffenenrechte zu beachten, dass das DSG technologieneutral ist, d.h. die Betroffenenrechte sind beim Einsatz von gKI im gesetzlichen Umfang zu gewährleisten, selbst wenn dies technisch schwierig sein sollte. Sie gelten zwar nicht absolut und Einschränkungen sind im Rahmen der gesetzlich vorgesehenen Interessenabwägungen möglich, damit lassen sich jedoch nicht alle Herausforderungen lösen.

[62] So kann sich ein Verantwortlicher, der ein gKI-Modell besitzt, nur dann auf seine eigenen Interessen zur Einschränkung des Auskunftsrechts berufen, wenn er die fraglichen Personendaten nicht Dritten mitteilt,[76] was jedoch bei kommerziellen Anbietern mit Inhalten aus dem Internet trainierten LLM regelmässig der Fall sein wird.

75 VUD-Leitfaden (FN 21), S. 12.

76 Art. 26 Abs. 2 Bst. a Ziff. 2 DSG.

[63] Ein Verantwortlicher wird darum andere Ansätze entwickeln müssen, wie das Auskunftsrecht in Bezug auf ein gKI-Modell DSG-konform erfüllt werden kann. Was hier technisch möglich ist, lässt sich dabei noch nicht wirklich sagen. Insgesamt entsteht allerdings der Eindruck, dass die Forschung zwar viel in den Aufbau von gKI-Systemen investiert hat, aber noch kaum in die Erfüllung von datenschutzrechtlichen Betroffenenrechten. Ein umfassendes Auskunftsgesuch des bereits zitierten schwedischen Datenschutzexperten beantwortete OpenAI zum Beispiel nur mit Personendaten im Zusammenhang mit seinem Benutzerkonto, legte aber keine Daten bei, die dieser zuvor durch Abfragen von ChatGPT generieren konnte und somit mutmasslich im Modell enthalten sind.[77] Wie sich der Fall weiterentwickelt, wird sich zeigen.

[64] Denkbar ist freilich, dass sich LLM-Anbieter wie OpenAI auf den Standpunkt stellen, dass ihr Modell gar keine Personendaten enthält, weil etwaige Personendaten nur und erst dann entstehen, wenn auf einen entsprechenden Prompt hin eine passende Antwort generiert wird – und die Prompts bestimmen nicht OpenAI, sondern die Benutzer ihrer Dienstleistung, die möglicherweise sogar als eigenständige Verantwortliche gelten. So oder so würde ein Auskunftsersuchen ins Leere zielen. Eine betroffene Person hätte konsequenterweise höchstens Anspruch auf Auskunft über die Antwort, die ein gKI-System auf eine von ihr selbst formulierte Frage generieren würde – mit den von ihr definierten Parametern zur Steuerung der Antworten.[78] Gewissheit, dass das LLM keine anderen (und möglicherweise auch persönlichkeitsverletzenden) Antworten produziert, hätte die Person freilich nie.

[65] Ein anderer Lösungsansatz wäre eine Auskunft über alle bisher generierten Antworten, in welcher der Name der betroffenen Person vorkommt, doch dies wird schon daran scheitern, dass diese Antworten nicht oder nur kurzzeitig aufbewahrt werden. Zu mehr sind die Anbieter nicht verpflichtet.

6.2. Widerspruchs- und Korrekturansprüche

[66] Etwas einfacher dürfte in der Praxis die Umsetzung des Rechts auf Widerspruch sein; darin enthalten ist auch das «Recht auf Vergessen».[79] Anders als das Auskunftsrecht lässt das DSG hier eine freie Interessenabwägung bei der Frage zu, ob und inwiefern einem Widerspruch entsprochen werden muss.

[67] Im Falle von gKI kann dies beispielsweise bedeuten, dass auf eine Löschung von Personendaten im Modell – so es dort solche Personendaten geben sollte – verzichtet werden kann, wenn durch andere Massnahmen sichergestellt wird, dass die fraglichen Personendaten im Output der gKI nicht mehr vorkommen. Funktioniert dies zuverlässig, wird sich typischerweise nicht mehr rechtfertigen lassen, weitergehend in die Position des Verantwortlichen einzugreifen.

[68] In der Praxis werden Löschbegehren von Anbietern von LLM inzwischen mit Filtern erfüllt. Sie funktionieren so, dass sie zu sperrende bzw. zu löschende Personendaten im Output einer gKI vor der Bekanntgabe an den Verwender schwärzen oder sonst eliminieren. Auch versucht

77 Den Antwortbrief von OpenAI veröffentlichte er auf LinkedIn (https://www.linkedin.com/posts/alexanderhanff_openai-dsar-response-activity-7054052957578833922-rYvJ/, archiviert unter https://perma.cc/6NJT-LSED).

78 So kann z.B. durch Angabe der sog. Temperatur mit Werten von 0 bis 2 beeinflusst werden, wie «kreativ» die GPT-gKI von OpenAI in der Erstellung des Outputs ist.

79 Art. 30 Abs. 2 Bst. b DSG.

die Forschung ihren KI-Systemen das Vergessen von Informationen beizubringen; praxistauglich sind diese Konzepte für die Zwecke von Betroffenenrechten aber noch nicht.[80]

[69] Beide genannten Methoden sind auch für die Ausübung des Berichtigungsrechts[81] von Interesse. Seine praktische Relevanz ist derzeit allerdings beschränkt: Da in aller Regel kein Anspruch darauf besteht, dass Daten einer bestimmten Person in einem KI-Modell vorkommen, kann das Problem falscher Personendaten in der Regel durch deren Löschung im Output gelöst werden. Stammen die falschen oder unvollständigen Personendaten aus der Fragestellung, mit welcher die gKI gefüttert wird, so ist wie dargelegt nicht der Besitzer des Modells der Verantwortliche, sondern dessen Verwender.

7. Fazit

[70] Die vorstehenden Ausführungen zeigen, dass der Einsatz von gKI wenig erstaunlich zu einer Vielzahl von datenschutzrechtlichen Fragestellungen führt. Das DSG vermag jedoch aufgrund der Tatsache, dass es technologieneutral und prinzipienbasiert ausgestaltet ist, durchaus praxistaugliche Antworten liefern. Das gilt – wie bereits an anderer Stelle dargelegt[82] – nicht nur für KI im Allgemeinen, sondern auch für gKI im Besonderen. Von einem rechtsfreien Raum, der durch neue Gesetze zwingend zu regeln ist, kann jedenfalls im Bereich des DSG keine Rede sein. Umsetzungsprobleme bestehen vor allem im Bereich der Betroffenenrechte und in der Qualitätssicherung insbesondere beim Training von gKI-Modellen.

[71] Nichtsdestotrotz sollten sich Unternehmen Gedanken dazu machen, ob und wie weit sie sich beim Einsatz von gKI über das DSG hinaus weiteren Einschränkungen unterwerfen möchten – und sei es nur zu Marketingzwecken. Auch die Politik wird entscheiden müssen, ob und wo sie im Bereich des Datenschutzes noch gesetzgeberischen Handlungsbedarf sieht. Alleine in den ersten acht Monaten von 2023 wurden im Bundesparlament rund zwei Dutzend Vorstösse zum Thema KI eingereicht. Die Schweiz hat allerdings mit ihrer Haltung, im Zweifel erst einmal abzuwarten, nach der hier vertretenen Ansicht keine schlechten Erfahrungen gemacht. Das neue DSG, welches erst mehr als fünf Jahre nach der DSGVO eingeführt worden ist, ist ein gutes Beispiel dafür.

Zusammenfassung

[72] Generative künstliche Intelligenz («gKI») stellt den Datenschutz nicht vor völlig neue Probleme. Nicht nur Maschinen können Inhalte mit falschen oder unangemessenen Personendaten produzieren, Menschen können dies genauso. Sie mögen darin weniger effizient und weniger überzeugend sein als eine gKI. Die Grundsätze des Datenschutzes greifen jedoch in beiden Fällen gleichermassen. Ob und wie gut sie eingehalten werden, ist daher nicht eine Frage des Datenschutzgesetzes («DSG»), sondern der technischen und organisatorischen Massnahmen, die zur Sicherstellung dessen Befolgung getroffen werden. Das DSG gibt diese nicht vor; jedes Unternehmen bestimmt sie selbst. Gerade hier zeigen sich im Falle von gKI die wirklichen Neuerungen,

80 Kamya Pandey, What Is Machine «Unlearning» and Why It Is an Important Skill for AI, in: JUMPSTART, 26. April 2023 (https://www.jumpstartmag.com/what-is-machine-unlearning-and-why-it-is-an-important-skill-for-ai/, archiviert unter https://perma.cc/C6P9-Z545).

81 Art. 32 Abs. 1 DSG.

82 Rosenthal, Datenschutz und KI (Fn. 20).

weil gKI neue Hilfsmittel und Prozesse erfordern kann, um altbekannte Vorgaben wie beispielsweise die Richtigkeit oder Verhältnismässigkeit von Personendaten weiterhin einzuhalten. Für einen datenschutzkonformen Einsatz von gKI genügt es allerdings nicht, sich auf die materiellen Regeln des Datenschutzes zu fokussieren. Ebenso wichtig ist ein klares Verständnis darüber, wer für welche Teile im Gesamtgefüge (rechtlich) verantwortlich ist. Innerhalb eines Unternehmens wird dies durch eine entsprechende Governance sichergestellt, d.h. eine unternehmensinterne Regelung der Aufgaben, Kompetenzen und Verantwortlichkeiten. Es sind entsprechende Weisungen für den Umgang mit gKI zu erteilen und Risikobeurteilungen vorzunehmen (der Autor hat hierzu kürzlich ein Werkzeug als Open Source publiziert[83]). Auch die im DSG neu eingeführten Instrumente, wie etwa die Datenschutz-Folgenabschätzung, können hier einen wichtigen Beitrag an den Datenschutz beim Einsatz von gKI leisten. Nebst der internen Verantwortlichkeit sollte sich ein Unternehmen aber auch darüber klar werden, wo im Aussenverhältnis seine datenschutzrechtliche Verantwortlichkeit beginnt und wieder aufhört. Dieser Beitrag erhält zwar einige Ausführungen hierzu, indem zwischen Erstellung, Besitz und Verwendung von gKI-Modellen unterschieden wird, in der Praxis ist die Abgrenzung aber oft nicht einfach.

David Rosenthal ist Partner einer Wirtschaftskanzlei in Zürich und Lehrbeauftragter der ETH Zürich und Universität Basel.

83 Generative AI Risk Assessment, GAIRA (https://www.rosenthal.ch/downloads/Rosenthal_GAIRA.xlsx).