www.jusletter.ch

David Rosenthal / Livio Veraldi

Das Training von KI-Sprachmodellen mit fremden Inhalten und Daten aus rechtlicher Sicht

Grosse Sprachmodelle benötigen für ihr Training umfangreiches Sprachmaterial. Viel davon stammt aus dem Internet. Diese Daten können jedoch Personendaten und geistiges Eigentum Dritter enthalten. Unter welchen Voraussetzungen ist deren Nutzung für Trainingszwecke zulässig? Zur Beantwortung dieser Frage analysieren wir zunächst Zweck und Funktionsweise des Trainings grosser Sprachmodelle. Anschliessend betrachten wir die relevanten Trainingsdatenquellen aus rechtlicher Perspektive. Schliesslich erläutern wir den Umgang mit diesen Quellen nach Schweizer Urheber-, Datenschutz-, Lauterkeits- und Vertragsrecht. Wir behandeln zudem «Crawler-Verbote» und die Haftung von Anbietern.

Beitragsart: Beiträge
Rechtsgebiete: Informatik und Recht, Urheberrecht, Datenschutz, Wettbewerbsrecht

Zitiervorschlag: David Rosenthal / Livio Veraldi, Das Training von KI-Sprachmodellen mit fremden Inhalten und Daten aus rechtlicher Sicht, in: Jusletter 3. Februar 2025

Inhaltsübersicht

I. Sinn und Zweck des Trainings

[1] Das Training eines grossen Sprachmodells besteht darin, dass bestehende Texte Wort für Wort (bzw. Token für Token) von einem System analysiert werden, um daraus Hinweise abzuleiten, wie Sprache künstlich generiert werden kann. Es geht nicht darum, konkrete Texte zu konservieren, um sich später daran «erinnern» zu können oder sie gar als solche, d.h. in mehr oder weniger gleicher Form, wiederzugeben (sog. Memorisierung, die in diesem Fall eine wortwörtliche wäre).

[2] Eine Memorisierung lässt sich allerdings nicht vermeiden, wenn derselbe Text im Training genügend häufig gesehen wird, weil er dann als solcher (und nicht nur die daraus gewonnenen sprachlichen Erkenntnisse) Teil des Sprachwissens bildet. Dieses Sprachwissen wird in der Anwendung benutzt, um den Input eines Benutzers mit dazu passendem Text fortzusetzen. Dies ist dann der Output. Hat das Modell im Training einen Satz genügend oft gesehen (z.B. «Der Apfel fällt nicht weit vom Stamm»), dann wird es ihn im Output dementsprechend fortsetzen, wenn es im Input mit seinem Anfang gefüttert wird («Der Apfel fällt nicht . . . »), weil dies für das Modell die statistisch wahrscheinlichste Fortsetzung ist. So kann sich ein grosses Sprachmodell beispielsweise charakteristische Slogans merken und sie auf Knopfdruck wiedergeben, wenn es entsprechend aktiviert wird.

[3] Dasselbe gilt für Sachwissen: Es ist nicht das primäre Ziel des Trainings eines grossen Sprachmodells, dass sich das Modell das Sachwissen aus einem einzelnen Beitrag in eben dieser Form merkt. Wenn es aber bestimmte Sachinformationen im Training genügend häufig sieht (statistische Relevanz), werden diese zum Sachwissen des Modells, was gewollt ist. Dieses Wissen kann einerseits Allgemeinwissen sein (etwa wie unsere Welt funktioniert), andererseits aber auch Fachwissen umfassen, wie es sich aus der Fachliteratur insgesamt ergibt. In diesem Sachwissen ist auch der Grund dafür zu sehen, warum die grossen Sprachmodelle zum Beispiel Personendaten wie etwa den Geburtstag von bekannten Personen wie z.B. Donald Trump sehr gut kennen, während sie den genauen Geburtstag einer anderen Person des öffentlichen Lebens, die ihn kaum vollständig publik machte und der daher nicht ins Training einfliessen konnte, nur erraten können – was sie denn auch tun, weil es zu ihrem Auftrag gehört, sprachlich korrekte Sätze und damit auch vollständige Geburtstagsangaben zu produzieren.

[4] In der Praxis der grossen Sprachmodelle wird der wortwörtlichen Memorisierung entgegengewirkt, indem versucht wird, jedes Quelldokument nur einmal für das Training zu benutzen. Dafür wird eine sog. De-Duplizierung durchgeführt (teilweise sogar Zeile für Zeile). Das ist auch sachgerecht, denn das Sprachmodell wird nicht besser, wenn es denselben Inhalt mehrfach sieht – es wird lediglich stärker darauf fixiert, was letztlich seiner Performanz schadet. Das Ziel des Trainings eines grossen Sprachmodells ist, ein möglichst diverses und umfassendes Bild von Sprache zu erhalten, d.h. Regeln, Bedeutungen und weitere Assoziationen von Sprachelementen im Sprachmodell abzubilden, die universell sind, weil sie in vielen verschiedenen Inhalten vorkommen. Freilich können bestimmte Inhalte oder Textfragmente, die populär sind, in unterschiedlichen Dokumenten vorkommen. Wird etwa ein bestimmter Slogan in der Sprache häufig gebraucht, macht es auch Sinn, dass das Sprachmodell ihn im entsprechenden Kontext benutzt, weil es seine Texte nach denselben Prinzipien und Regeln wie in den Trainingsinhalten aufbauen will. Das ist der Sinn und Zweck eines grossen Sprachmodells.[1]

[1] Der Vollständigkeit halber sei vermerkt, dass der Begriff der *Memorisierung* lediglich eine vereinfachte Darstellung der Vorgänge in einem Sprachmodell ist, um bestimmte rechtliche und tatsächliche Problemstellungen zu erörtern und einzuordnen. Denn es gibt keinen allgemeingültigen Punkt, ab welchem ein Inhalt als memorisiert gilt. Wenn wir etwa davon sprechen, dass ein Sprachmodell «weiss», dass Harry Potter ein Zauberlehrling ist, dann deshalb, weil sich «Harry Potter» und «Zauberlehrling» in einer der unzähligen Dimensionen, in welchen sich ein Sprachmodell die Bedeutung und Beziehung von Wörtern vorstellt, vergleichsweise nahe stehen und daher die Wahrscheinlichkeit, dass Harry Potter eben ein solcher Zauberlehrling ist, sehr hoch ist. Es bleibt aber eine Wahrscheinlichkeit. Zu anderen Eigenschaften oder anderen Wissenselementen wird die Beziehung bzw. Nähe weniger stark sein, aber sie ist da und kann mathematisch berechnet werden. Praktisch wird ein Inhalt dann als memorisiert gelten, wenn ihn das Modell auf inhaltlich vergleichbare Prompts und bei unterschiedlichen Parametern (wie etwa der «Temperatur») immer wieder genau so liefert, statt in seinen Antworten diesbezüglich zu variieren, weil andere Inhalte vergleichbar wahrscheinlich sind (sog. Halluzinationen). Eine andere ebenfalls relevante Frage ist, wie wahrscheinlich es ist, dass überhaupt ein entsprechender Prompt eingegeben wird und

II. Technischer Hintergrund

[5] Für das Training von grossen Sprachmodellen müssen die entsprechenden Trainingsinhalte zunächst zur Bereinigung für die Zwecke des Machine Learnings bereitgestellt und damit vervielfältigt werden; sie müssen zur Einhaltung des Datenschutzes teilweise geschwärzt werden; und sie müssen in kleine, von der Maschine verwendbare Bruchstücke (sog. Tokens) aufgeteilt werden. Es kommt also zu verschiedenen Vervielfältigungen und Bearbeitungsvorgängen der betreffenden Trainingsinhalte. Insbesondere erfolgen im Arbeitsspeicher der verwendeten Systeme und auch auf den sonstigen Speichersystemen desjenigen, der das Modell trainiert, kurzzeitige Kopien der Inhalte. Das Sprachmodell selbst speichert allerdings die Trainingsinhalte nicht als solche, sondern vielmehr den «Durchschnitt» der Erkenntnisse daraus. Im Modell selbst wird nicht eine Kopie der einzelnen Trainingsinhalte erstellt, sondern im Grunde «nur» eine sie betreffende Statistik.

[6] Wir haben andernorts viel detaillierter (für Nicht-Fachleute) beschrieben, wie grosse Sprachmodelle im Detail funktionieren und in den Grundzügen dargelegt, warum sie das können, was sie können.[2] Der Trainingsvorgang eines grossen Sprachmodells kann aber vereinfacht gesagt so beschrieben werden (hier ohne Unterscheidung der einzelnen Komponenten des Sprachmodells):

1. Das Modell startet gewissermassen «dumm» – seine Milliarden von Parametern sind nach dem Zufallsprinzip eingestellt. Wird dem Modell ein Input gegeben (was später der Prompt sein wird), kommt eine zufällige, bedeutungslose Zeichenfolge heraus.
2. Dem Modell werden nun zahlreiche Texte für das Basistraining vorgelegt. Das geschieht stückchenweise. Jedes Stückchen wird als sog. Token bezeichnet. Das Wort «stückchenweise» besteht beim beliebten Modell «gpt-4o» von OpenAI zum Beispiel aus den drei Tokens «stück» (Nr. 41979), «chen» (Nr. 3184) und «weise» (Nr. 17864).[3] Die Sprachmodelle arbeiten also mit Tokens, nicht mit Buchstaben oder Wörtern. Es geht dabei aber nichts verloren, d.h. die Umwandlung von Text in Tokens und zurück in Text ist verlustfrei möglich.[4]
3. Das Modell wird nun mit dem Anfang eines Textes gefüttert (quasi als Prompt) und es muss dann berechnen, wie der Text weitergeht. Das kann man sich im übertragenen Sinne und stark vereinfacht so vorstellen, wie wenn ein Tabellenkalkulationsblatt mit Milliarden von Zellen erstellt wird. Jede Zelle reicht die Werte, die sie erhält, basierend auf einfachen Formen und vorprogrammierten Parametern an jeweils weitere Zellen weiter. In der ersten Spalte wird der Input eingetragen (die Tokens des Inputs). In der letzten Spalte des Kalkulationsblatts hat es wiederum einen Wert für jeden Token. Der Token, bei welchem der höchste Wert angezeigt wird, wird benutzt.

---

wer dafür verantwortlich ist. Ausführlich dazu im Kontext des Datenschutzes: David Rosenthal, Teil 19: Sprachmodelle mit und ohne personenbezogene Daten, abrufbar unter: https://www.vischer.com/know-how/blog/teil-19-sprachmodelle-mit-und-ohne-personenbezogene-daten/; siehe weiter: David Rosenthal, Teil 21: Das Auskunftsrecht bei grossen Sprachmodellen, abrufbar unter: https://www.vischer.com/know-how/blog/teil-21-das-auskunftsrecht-bei-grossen-sprachmodellen/.

2 Mehr dazu, wie ein Sprachmodell technisch funktioniert: David Rosenthal, Teil 17: Was in einem KI-Modell steckt und wie es funktioniert, abrufbar unter: https://www.vischer.com/know-how/blog/teil-17-was-in-einem-ki-modell-steckt-und-wie-es-funktioniert/.

3 Zum selbst ausprobieren: https://tiktokenizer.vercel.app/.

4 Dies heisst jedoch nicht, dass Sprachmodelle alle Trainingstexte speichern. Das tun sie gerade nicht. Tokens sind lediglich eine effizientere Art der Darstellung von Text für die Zwecke des Trainings. Unterschiedliche Anbieter verwenden mitunter unterschiedliche Arten der Tokenisierung.

4. Beim Training wird nun geprüft, ob der Token mit dem höchsten Wert auch derjenige ist, der im eingespeisten Text an nächster Stelle kommen würde. Ist das nicht der Fall, wird quasi rückwärts berechnet, wie die Parameter der vorangehenden Zellen hätten sein müssen, damit auf der letzten Ebene der «richtige» Token den höchsten Wert erhalten hätte. Anhand dieser Berechnungen werden dann diese Parameter wie Schräubchen ein wenig gedreht, damit es beim nächsten Mal besser klappt.
5. Das wird unzählige Male wiederholt, aber nicht mehr mit demselben Text, sondern einem neuen. Dabei kommen auch Techniken gegen die Memorisierung zum Einsatz, bei welchen das Modell z.B. nur einen Teil der Wörter vorhersagen muss und nicht jeder einzelne Token als Trainingsziel verwendet wird. Je mehr Texte verarbeitet werden, desto besser kann das Modell allerdings die Texte ergänzen. Es ist dies ein massiver Rechenaufwand. Das ist auch der Grund, warum es erst jetzt so leistungsfähige grosse Sprachmodelle gibt: Früher gab es die Rechnerkapazitäten schlicht nicht dafür.
6. Damit ist das Modell allerdings noch nicht gebrauchstauglich. Damit es z.B. für einen Chatbot benutzt oder für andere Aufgaben eingesetzt werden kann, muss es auf solche Anwendungen getrimmt werden. Dies geschieht wiederum durch ein Training. Dieses besteht nun in der Vorlage von Beispielen, also beispielsweise von Fragen und passenden Antworten. Allerdings braucht es für dieses Training nur einen Bruchteil der Trainingsdaten, die für den Grundaufbau des Sprachmodells (auch «Pre-Training» genannt) benutzt worden sind. Wenn Firmen wie OpenAI daran interessiert sind, die Verwendung ihrer Chat-Anwendungen durch die Anwender für das Training ihrer Modelle zu nutzen, dann geht es typischerweise genau um diese Art des Trainings (und weniger um den Grundaufbau des Basiswissens der Modelle). Es macht aus Textkomplettierungsautomaten nützliche Werkzeuge.
7. Schliesslich kann dann mit dem Prompt noch genauer gesteuert werden, was ein Sprachmodell tut. Dienste wie «ChatGPT» arbeiten aber nicht nur mit dem Prompt, den der Benutzer oder die Benutzerin eingibt. Sie verfügen auch über sog. Systemprompts, also unterschiedlich lange Anweisungen, die dem Modell zusammen mit der Eingabe des Benutzers mit auf den Weg gegeben werden, wenn es eine Antwort entwerfen soll.[5]

[7] Wichtig für die vorliegende Diskussion ist, dass das Modell in allen Fällen die Trainingsinhalte nicht als solche abspeichert. Im Ergebnis findet eine Analyse bestehender Texte statt, um daraus übergeordnete Hinweise abzuleiten, wie Sprache künstlich generiert werden kann. Diese Hinweise werden – auf einer Meta-Ebene betrachtet – in Form von Assoziationen zwischen Sprachelementen und Mustern in Milliarden von Parametern im Modell abgelegt und sind für den Menschen nicht ohne weiteres lesbar; er sieht nur die Parameter. Es braucht spezielle Techniken, um die Assoziationen erkenntlich und damit sichtbar zu machen, welche Begriffe einander z.B. in einem bestimmten Sinne nahestehen oder nicht. Dabei trägt jeder neue Text durch die Abbildung der in ihm enthaltenen Hinweise zur Sprachbildung zur Justierung der Parameter und damit zur Sprachbildung im Modell bei – jedenfalls ein klein wenig. Die sprachlichen Hinweise sind im Grunde wie einzelne Sandkörner an einem Sandstrand; ihre Masse macht den Strand aus.

5 Die System-Prompts behalten die Anbieter normalerweise unter Verschluss, aber immer wieder gelingt es, diese den Chat-Diensten zu entlocken, wie z.B. hier: https://patmcguinness.substack.com/p/gpt-4-system-prompt-revealed.

Denn das Ziel des Trainings ist, dass das Modell stets besser darin wird, neue Texte zu generieren – nicht Trainingsinhalte wiederzugeben. Schliesslich werden die Trainingsdatensätze regelmässig für spätere Nachtrainings aufbewahrt. Bestimmte Inhalte werden ausserdem für Tests ausgesondert, d.h. in einem nachgelagerten, ähnlichen Verfahren benutzt.

III. Arten von Quellen für das Training

[8] Quellen für das Training von grossen Sprachmodellen gibt es heute zahlreiche:

- **Öffentlich zugängliche Websites:** Es werden Texte von öffentlich zugänglichen Websites und anderen Internet-Ressourcen eingelesen und verarbeitet. Dieser Vorgang nennt sich Webscraping. Ein Programm ruft eine Webseite auf, wie sie jeder Benutzer aufrufen könnte, und liest den Text und den Code der Seite und extrahiert daraus den eigentlichen Inhalt, den es in einer Datenbank zur weiteren Verarbeitung ablegt. Das geschieht entweder direkt, indem eine Organisation entsprechende Suchroboter (sog. Crawler) selbst einsetzt, oder es wird auf bereits aufbereitete Datensätze zurückgegriffen, die aus früheren Crawling-Suchläufen stammen. So wurde z.B. 2007 die US-Stiftung Common Crawl gegründet, die seither über 250 Milliarden Internetseiten erfasst und in aufbereiteter Form zur Verfügung stellt, so auch für das Training von Sprachmodellen. Jeden Monat kommen 3–5 Milliarden Seiten hinzu. Es gibt auch andere solche Datensätze (z.B. auf der bekannten Machine-Learning-Plattform Hugging Face). Diesen Datensätzen ist gemeinsam, dass sie aus frei zugänglichen Inhalten bestehen. Frei zugänglich bedeutet allerdings nicht, dass an diesen Inhalten keine Drittrechte bestehen können. So mögen beispielsweise die Texte eines Online-Magazins frei zugänglich sein, urheberrechtlich geschützt sind sie trotzdem. Ebenso können sie personenbezogene Daten enthalten.

  Eine Sonderkategorie von frei zugänglichen Inhalten sind **gestohlene Inhalte** (z.B. aus Ransomware-Angriffen), die auf mehr oder weniger einfach zugänglichen Plattformen veröffentlicht werden. Bekanntes Beispiel ist der «books3» Datensatz aus der Machine-Learning-Datensatz-Sammlung «The Pile». Es ist ein Datensatz, welcher 196'640 Bücher von Autoren wie Stephen King, Margaret Atwood und Zadie Smith enthalten soll und 2020 erstellt wurde, jedoch als Raubkopie gilt und daher nach einer Beschwerde im August 2023 aus «The Pile» entfernt wurde (er ist allerdings immer noch im Internet verfügbar).
- **Open-Access-Plattformen:** Es gibt verschiedene Plattformen, die benutzt werden können, um Aufsätze, Bücher und andere Inhalte für die Allgemeinheit zugänglich zu machen und zur Nutzung über den blossen Konsum hinaus zur Verfügung zu stellen, wobei unterschiedliche Open-Access-Lizenzen (etwa von CreativeCommons) benutzt werden. Ein Beispiel für wissenschaftliche Publikationen (ohne Peer-Review) ist arXiv, mit rund 2.4 Mio. Aufsätzen. Zu den anderen Inhalten gehört auch Software-Code, der unter Open-Source-Lizenzen angeboten wird. Ein bekanntes Beispiel sind die öffentlich zugänglichen Angebote auf Github. Auf anderen Plattformen werden Inhalte angeboten, die als Public Domain gelten, d.h. das Urheberrecht wurde aufgegeben oder der Schutz ist abgelaufen. Ein Beispiel ist das Project Gutenberg. Auch auf Hugging Face sind Datensätze mit solchen Inhalten verfügbar. In Bibliotheken zugängliche Bücher bzw. deren Inhalte gehören auch in diese Kategorie, jedenfalls soweit sie von den Bibliotheken gekauft wurden (nicht hingegen Online-Dienste, die mit einem Abo der Bibliothek genutzt werden können).

- **User-Generated-Content-Plattformen:** Gemeint sind einerseits «freie» Plattformen wie Wikipedia, aber auch kommerzielle Angebote wie YouTube oder Reddit. Multimediale Inhalte können für das Training von Sprachmodellen ebenfalls benutzt werden, indem gesprochene Sprache (z.B. in Videos mit Anleitungen) in einem ersten Schritt transkribiert werden. Es gibt bereits Anbieter, die entsprechende Datensätze anbieten (z.B. Youtube-Commons). Auch wenn bei diesen Plattformen die Inhalte von deren Benutzern stammen, unterliegen sie jeweils Lizenzbedingungen, die entweder auf der Plattform formuliert sind oder vom Benutzer selbst festgelegt werden. Mitunter werden Rechte auch an die Plattform-Betreiber abgetreten, die sie dann selbst nutzen können. Die Inhalte auf diesen Plattformen können auch Drittrechte tangieren, z.B. wenn ein Benutzer unerlaubt fremde Inhalte heraufladt.
- **Vertraglich zugängliche Inhalte:** Gemeint sind Inhalte, die nicht frei zugänglich sind, bei denen also eine Registrierung oder sonstige Vereinbarung erforderlich ist. Das sind z.B. die Inhalte von Massenmedien, für welche ein bezahltes Abo oder auch schon nur eine kostenlose Registrierung erforderlich ist. Darunter fallen auch Inhalte von nur ihren registrierten Benutzern zugänglichen Online-Plattformen und Archiven. Es kann sich dabei um öffentliche oder nicht-öffentliche Inhalte handeln. «Öffentlich» ist ein Inhalt dann, wenn er für eine unbestimmte Anzahl an Personen zugänglich ist, auch wenn hierfür z.B. eine Bezahlschranke besteht. Entscheidend ist die Kontrolle des Zugangs, die nur denjenigen gewährt wird, die sich vertraglich zur Einhaltung von bestimmten Bedingungen verpflichten.
- **Öffentliche Bibliotheken:** Soweit eine Bibliothek öffentlich frei zugänglich ist, d.h. ohne Abschluss eines Vertrags, fällt sie nicht unter die vorstehende Kategorie. Physische Bibliotheken unterscheiden sich von einer Website, einem Online-Medium oder den erwähnten Open-Access-Plattformen dahingehend, dass nicht Online-Kopien von Werken, sondern physische Werkexemplare zugänglich gemacht werden (z.B. gedruckte Bücher und Zeitschriften). Das ist urheberrechtlich relevant, weil der Rechteinhaber eines Werks, der zugestimmt hat, dass ein Exemplar dieses Werks veräussert wird, nicht verhindern kann, dass dieses Exemplar weiterveräussert oder sonst wie verbreitet oder eben in einer Bibliothek bereitgestellt und ausgeliehen wird (Grundsatz der Erschöpfung). Die Erschöpfung beschränkt sich jedoch auf das Verbreitungsrecht. Damit kann eine weitergehende Nutzung (z.B. das Scannen) von Inhalten öffentlicher Bibliotheken urheberrechtlich relevant sein. Demgegenüber hat der Rechteinhaber bei einer Publikation auf einer Open-Access-Plattform üblicherweise in eine weitergehende Nutzung eingewilligt. Das Werk selbst ist hier typischerweise mit einer freien Lizenz versehen, d.h. es kann unabhängig von der Open-Access-Plattform entsprechend den Lizenzbedingungen genutzt werden (oder eben nicht).
- **Synthetische Inhalte:** Sie sind eine Sonderkategorie der bisher erwähnten Inhalte, weil sie nicht von Menschen geschaffen wurden und daher in vielen Rechtsordnungen – so auch in der Schweiz – keinen Urheberrechtsschutz geniessen.[6] Gemeint sind beispielsweise die von einem bestehenden KI-Modell wie gpt-4o generierten Outputs. Auch sie werden mitunter für das Training von Sprachmodellen verwendet, auch wenn dies Sonderprobleme punkto Datenqualität mit sich bringt. Die Nutzung synthetischer Daten kann auch Probleme

6 Sie können jedoch deshalb von urheberrechtlicher Relevanz sein, weil sie in den Schutzbereich bestehender Werke fallen und damit eine Urheberrechtsverletzung darstellen können.

mit dem Datenschutz vermeiden, etwa wenn die synthetischen Daten benutzt werden, um «echte» Personendaten durch vom Computer nach dem Zufallsprinzip «erfundene» Daten zu ersetzen, die gleich aussehen bzw. aufgebaut sind und «wie echt» erscheinen, aber sich nicht auf eine tatsächlich lebende Person beziehen. Das hat den Vorteil, dass das Sprachmodell lernt, wie diese Daten aussehen und in der Sprache benutzt werden, es aber keine reale betroffene Person gibt. Aus diesem Grund werden synthetische Daten auch für Testzwecke benutzt.

- **Kombinierte Inhalte:** Es gibt eine wachsende Anzahl an kuratierten Datensätzen, die direkt für das Training von grossen Sprachmodellen verwendet werden können und mehrere der genannten Quellen in sich vereinen. Ein Beispiel ist OpenHermes, ein aus diversen Quellen kombinierter Datensatz für das Finetuning, bestehend aus (vorwiegend synthetisch generierten) Frage-Antwort-Sequenzen.

IV. Durch das Training tangierte Rechtsgebiete

[9] Insbesondere folgende Rechtsgebiete werden durch das Training eines grossen Sprachmodells tangiert:[7]

- **Urheberrecht:** Soweit ein Inhalt gemäss Urheberrechtsgesetz (**URG**) urheberrechtlich geschützt ist, darf er grundsätzlich nur verwendet werden, wenn der Rechteinhaber entweder eingewilligt hat oder eine gesetzliche Schrankenbestimmung (z.B. das Zitatrecht) dies erlaubt. Es stellt sich allerdings die Frage, ob ein bestimmter Inhalt überhaupt urheberrechtlich geschützt ist, und ob die Nutzung von urheberrechtlich geschützten Inhalten für das Training eines Sprachmodells überhaupt eine urheberrechtlich relevante Verwendung darstellt. Urheberrechtlichen Schutz geniessen in der Schweiz *a priori* nur geistige Schöpfungen mit individuellem Charakter (wobei letzterer bei Fotografien nicht erforderlich ist), sofern die Schutzfrist des Urheberrechts nicht abgelaufen ist (i.d.R. mit 70 Jahren nach dem Tod des Urhebers) oder es von Gesetzes wegen nicht besteht (z.B. Gesetze, Entscheidungen und Berichte von Behörden und öffentlichen Verwaltungen).
- **Datenschutz:** Soweit sich ein Inhalt auf eine identifizierte oder identifizierbare Person bezieht, handelt es sich um personenbezogene Daten. Die Verarbeitung solcher Daten setzt nach Datenschutzgesetz (**DSG**) und EU-Datenschutz-Grundverordnung (**DSGVO**) zunächst voraus, dass eine Reihe von Grundsätzen eingehalten wird (Transparenz, Treu und Glauben, Zweckbindung, Verhältnismässigkeit inklusive Datenminimierung und Speicherbegrenzung, Richtigkeit im Hinblick auf den Zweck, Datensicherheit). Nach DSGVO und in der Schweiz für öffentliche Organe auch nach DSG ist ferner ein hinreichender Rechtsgrund erforderlich (z.B. eine Einwilligung, ein berechtigtes Interesse oder eine gesetzliche Bestimmung). Weiter ist im privaten Bereich ein Rechtfertigungsgrund erforderlich, wenn eine betroffene Person der Bearbeitung ihrer Daten widerspricht (d.h. nicht will, dass ihre Daten verarbeitet werden und dies einem Datenbearbeiter konkret sagt). Im DSG ist dies zudem der Fall, wenn besonders schützenswerte personenbezogene Daten weitergegeben werden und unter der DSGVO, wenn besonders schützenswerte personenbezogene Daten verarbei-

[7] In diesem Sinne auch: Matthias Städeli/Lisa Mary, Künstliche Intelligenz und Urheberrecht, in: SJZ 120/2024, S. 244 ff., S. 244.

tet werden. Im DSG können private Organisationen die Verletzung der oben genannten Grundsätze immerhin durch den Nachweis eines überwiegenden Interesses rechtfertigen.

An dieser Stelle sei auch noch Art. 28 Zivilgesetzbuch (**ZGB**) erwähnt, welcher im weiteren Sinne ebenfalls zur Diskussion stehen kann, weil die Bestimmung ebenfalls die Persönlichkeit schützt – anders als das DSG im Übrigen auch die Persönlichkeit juristischer Personen. Soweit also letztere vom Training eines Modells betroffen sind, könnte diese Bestimmung von Relevanz sein. Es sind hierzu allerdings keine Fälle bekannt.

Im vorliegenden Beitrag gehen wir nicht vertieft auf die DSGVO ein. Sie findet auf ein Training von Sprachmodellen in der Schweiz durch Stellen mit Sitz in der Schweiz grundsätzlich keine Anwendung, weil keine der beiden Voraussetzungen von Art. 3 Abs. 2 DSGVO erfüllt ist.

- **Lauterkeitsrecht:** Das Lauterkeitsrecht komplementiert je nach Rechtsordnung das Urheberrecht dahingehend, dass es Inhalte aus einem anderen Blickwinkel schützt (und zwar mitunter auch solche, die nicht urheberrechtlich geschützt sind). Es muss daher separat geprüft werden. In der Schweiz verbietet das Bundesgesetz gegen den unlauteren Wettbewerb (**UWG**) die Verwertung von anvertrauten oder unbefugt zugänglich gemachten Arbeitsergebnissen ohne Erlaubnis desjenigen, dem das Arbeitsergebnis gehört. Weiter untersagt das UWG die Übernahme und Verwertung eines marktreifen Arbeitsergebnisses eines anderen als solches ohne angemessenen eigenen Aufwand durch technische Reproduktionsverfahren.
- **Geheimnisschutz:** In einigen Rechtsordnungen wird die Verletzung von Geheimhaltungspflichten oder die Verwertung von Inhalten, die jemand aufgrund einer Verletzung einer Geheimhaltungspflicht erfahren hat, unter Strafe gestellt. In der Schweiz tun dies das UWG und das Strafgesetzbuch (**StGB**). Der Geheimnisschutz kann sich aus verschiedenen Umständen ergeben, so etwa aus einer Vertragsbeziehung oder aus gesetzlichen Regelungen wie dem Datenschutzrecht oder Bestimmungen zum Amtsgeheimnis.
- **Strafrecht:** Bei bestimmten, nicht gewollten Inhalten verbietet der Gesetzgeber den Umgang damit oder gar den Besitz ganz oder teilweise, so z.B. bei Kinderpornographie oder bei öffentlichem Aufruf zu Hass oder zu Verbrechen oder Gewalttätigkeit.
- **Spezialgesetzliche Schutzbestimmungen:** In einzelnen Rechtsordnungen sind bestimmte Inhalte spezialgesetzlich vor der Verwendung ohne Erlaubnis geschützt. In der EU sind beispielsweise Datenbanken, für welche erhebliche Investitionen getätigt worden sind, während 15 Jahren vor der Übernahme des vollständigen Inhalts oder eines erheblichen Teils geschützt.
- **EU AI Act:** Mit der Verordnung über künstliche Intelligenz (**AI Act**) hat die EU eine auf Produktesicherheit im Bereich KI ausgerichtete Regulierung eingeführt. Sie trat am 1. August 2024 in Kraft und findet seither über einen Zeitraum von 36 Monaten schrittweise Anwendung. Allzweck-KI-Modelle, wozu auch grosse Sprachmodelle gehören, werden damit ebenfalls reguliert. Der AI Act schreibt allerdings nicht vor, mit welchen Inhalten solche Modelle zu trainieren sind. Er definiert in Art. 53(1) AI Act vielmehr diverse Dokumentationspflichten, einschliesslich in Bezug auf das Training und das Testen eines Modells. Verwendern dieser Modelle sollen bestimmte Mindestinformationen zur Verfügung gestellt werden. Art. 53(1) AI Act enthält allerdings auch eine Regel, die das Urheberrecht betrifft und potenziell vorliegend relevant sein könnte. Die Bestimmung schreibt für Allzweck-

KI-Modelle unter anderem vor, dass eine «Strategie zur Einhaltung des Urheberrechts der Union» bestehen und eine «Zusammenfassung der für das Training des KI-Modells [...] verwendeten Inhalte» publiziert werden muss (vgl. auch Erwägungsgrund 105). Die Regelung erwähnt des Weiteren eine Schrankenbestimmung des EU-Urheberrechts, die sog. Text-und-Data-Mining-Ausnahme (**TDM**), welche eine Verwendung von Inhalten für das Training von Sprachmodellen erlaubt, wobei der Rechteinhaber ein Opt-out-Recht hat in Bezug auf Anwendungen ausserhalb der wissenschaftlichen Forschung.[8] Der AI Act verlangt, dass jeweils ermittelt wird, ob dieses Opt-out-Recht geltend gemacht worden ist. Es stellt sich allerdings die Frage, ob die Vorgaben des AI Act auch dann zur Anwendung gelangen, wenn das EU-Urheberrecht selbst gar nicht auf einen bestimmten Sachverhalt angewandt werden will, weil sich nach dem sog. Schutzlandprinzip ergibt, dass ein Training eines Sprachmodells ausschliesslich in der Schweiz oder etwa in den USA stattfindet. Es muss dann im Rahmen der erwähnten Strategie nur – aber immerhin – sichergestellt werden, dass das Modell *erstens* nicht in der EU trainiert wird und *zweitens* das EU-Urheberrecht bei einem späteren Vertrieb des Sprachmodells in der EU durch dessen Output nicht verletzt wird.[9] Wir sind der Ansicht, dass dieser einschränkende Ansatz zutreffend ist, was sich einerseits aus dem Wortlaut ergibt und andererseits daraus, dass der AI Act das bestehende Urheberrecht nicht ändern, d.h. auch seine Anwendbarkeit nicht ausweiten will.[10] Dies steht zwar in einem gewissen Widerspruch zu den Erwägungen, wonach die Regelung beabsichtige, «gleiche Wettbewerbsbedingungen» für alle Anbieter zu schaffen, was so verstanden werden könnte, dass für alle Organisationen, die Modelle in der EU anbieten wollen, dieselben urheberrechtlichen Regeln für das Training gelten sollen, ganz gleich, wo es stattfindet.[11] Ein solcher Gedanke allein in einer Erwägung, entstanden aus dem Unwissen oder Missverständnis derjenigen, die die Bestimmung entworfen haben, vermag jedoch nicht die tatsächliche Regelung des AI Act und des Urheberrechts zu übersteuern: Der AI Act verlangt nur – aber immerhin –, dass eine «Policy» besteht, wie das EU-Urheberrecht eingehalten wird, und einzuhalten ist das EU-Urheberrecht nur – aber immerhin – dann, wenn es selbst angewandt werden will – was wie gezeigt in der Regel nicht der Fall ist, soweit es um das Training eines Modells im Ausland geht. Mit dieser Interpretation kann auch der Widerspruch zur Erwägung aufgelöst werden. Die Logik gilt übrigens auch umgekehrt, wenn Schweizer Inhalte für das Training im Ausland verwendet werden. Sie dürfte allerdings nicht ohne Widerspruch akzeptiert werden. So können wir uns durchaus vorstellen, dass jedenfalls dann, wenn Inhalte in der Schweiz oder sonst einem Land gesammelt werden, ein hinreichender Bezug zu diesem Land besteht, um dessen Urheberrecht auch für nachgelagerte Vorgänge (wie etwa die Nutzung für das Training) zur Anwendung zu bringen. Das entspricht allerdings nicht dem traditionellen Verständnis des zur Bestimmung des anwendbaren Rechts geltenden Schutzlandprinzips.

8 Wir gehen auf diese Thematik im Kapitel V zum Urheberrecht noch näher ein.

9 Siehe dazu den Exkurs in Kapitel V.F.

10 David Rosenthal, Der EU AI Act – Verordnung über künstliche Intelligenz, in: Jusletter vom 5. August 2024, N 65, m.w.H.

11 Erwägung 106 AI Act: «[...] Any provider placing a general-purpose AI model on the Union market should comply with this obligation, regardless of the jurisdiction in which the copyright-relevant acts underpinning the training of those general-purpose AI models take place. This is necessary to ensure a level playing field among providers of general-purpose AI models where no provider should be able to gain a competitive advantage in the Union market by applying lower copyright standards than those provided in the Union».

- **Vertragsrecht:** Als rechtliche, wenn auch nicht gesetzliche Vorgaben gelten auch Vereinbarungen, die bei der Beschaffung von Inhalten für das Training getroffen wurden. Hierzu gehören nebst den Geheimhaltungsvereinbarungen (vgl. oben «Geheimnisschutz») insbesondere auch Lizenzvereinbarungen und Vereinbarungen, welche die Verwendung von Inhalten selbst dann beschränken können, wenn sie gesetzlich nicht geschützt sind. Solche Verträge kommen in unterschiedlichster Form vor, z.B. als Nutzungsbedingungen bei der Registrierung auf einer Online-Plattform, bei der Anmeldung für einen Datenbankzugang oder im Rahmen einer Geschäftsbeziehung, in welcher Daten anfallen. Noch sind Klauseln, welche die Verwendung von Inhalten spezifisch für maschinelles Lernen einschränken, eher rar. Sie werden jedoch immer häufiger vorkommen. Ist maschinelles Lernen nicht spezifisch adressiert, muss jeweils ausgelegt werden, ob z.B. eine Regelung, welche die Verwendung von Inhalten «nur für interne Zwecke» erlaubt, durch den Einsatz für das Training von Sprachmodellen verletzt wird. Wird die Verwendung von Inhalten auf die «Zwecke der Vertragsabwicklung» beschränkt, was insbesondere in Geheimhaltungsklauseln heute Standard ist, so dürfte beispielsweise die Verwendung für nicht damit zusammenhängende Trainings von Sprachmodellen erschwert sein. Nutzt ein Vertragspartner die Inhalte trotzdem dafür, verletzt er den Vertrag und kann zu Schadenersatz verpflichtet werden. Diesfalls kann der Vertrag regelmässig auch gekündigt werden oder es kann verlangt werden, dass die vertragswidrige Nutzung aufhört. Eine Vertragsverletzung ist für sich jedoch noch nicht rechtswidrig; es müssen hier weitere Dinge hinzukommen, wie etwa der Verrat von Geschäftsgeheimnissen, Urheber- oder Datenschutzverletzungen. Letztere drohen natürlich jeweils dort, wo der Vertrag durch die Einräumung einer Nutzungslizenz überhaupt erst die Basis für die (z.B. nach Urheberrecht) rechtskonforme Nutzung von Inhalten geschaffen hat.

[10] In der Praxis muss somit jemand, der sein Sprachmodell mit bestimmten Inhalten trainieren möchte, sämtliche dieser Punkte in Bezug auf diese Inhalte prüfen. Sie gelten parallel, d.h. eine Verwendung von Inhalten kann urheberrechtlich unproblematisch sein, aber gegen den Datenschutz verstossen und umgekehrt. Hierzu haben wir am Ende ein Prüfschema.

[11] In der Folge beschreiben wir für einige der oben genannten rechtlichen Vorgaben Ansätze, wie diese eingehalten werden können. Worauf wir in diesem Beitrag nicht eingehen, ist die Frage, welche Vorkehrungen getroffen werden müssen, um den Qualitätserwartungen an das Training eines KI-Modells gerecht zu werden, die teilweise auch rechtlicher Natur sein können. Der EU AI Act verlangt beispielsweise bei Hoch-Risiko-KI-Systemen, dass im Rahmen der Trainings, der Validierung und des Testens von KI-Modellen eine bestimmte Data Governance durchgeführt worden ist, um bestimmten Qualitätsanforderungen zu entsprechen. Ebenfalls nicht abgehandelt werden hier die weiteren nach dem EU AI Act bestehenden Pflichten für Anbieter von Allzweck-KI-Modellen, etwa im Bereich der Dokumentation. Sie gelten ab dem 2. August 2025.

V. Urheberrecht

A. Vorbemerkungen

[12] Beim Training von grossen Sprachmodellen und der Frage der Zulässigkeit eben jenes unter immaterialgüterrechtlichen Aspekten ist primär das Urheberrecht von Relevanz. Dieses vermittelt dem Rechteinhaber das ausschliessliche Recht darüber zu bestimmen, ob, wann und wie sein

Werk verwendet wird (Art. 10 Abs. 1 URG). Selbst wenn für das Training von Sprachmodellen auch gemeinfreie Inhalte verwendet werden, an denen kein Urheberrecht besteht, werden viele Trainings-Datensätze urheberrechtlich geschützte Inhalte umfassen.[12] Das gilt insbesondere dann, wenn eine grössere Menge an Daten aus dem öffentlichen Internet benutzt wird. Wer damit Sprachmodelle trainieren will, muss also aufzeigen können, warum er diese Rechte nicht verletzt. Hier bieten sich drei Ansätze an:

- Das Training von Sprachmodellen mit urheberrechtlich geschützten Inhalten ist keine urheberrechtlich relevante Handlung;
- die Verwendung der Inhalte ist vom Gesetzgeber erlaubt, d.h. es gibt eine anwendbare Schrankenbestimmung; oder
- die Verwendung der Inhalte ist vom Rechteinhaber erlaubt.

[13] Wir werden nachfolgend auf alle drei Ansätze eingehen. Das folgende Diagramm bietet hierzu einen Überblick (Abbildung 1):

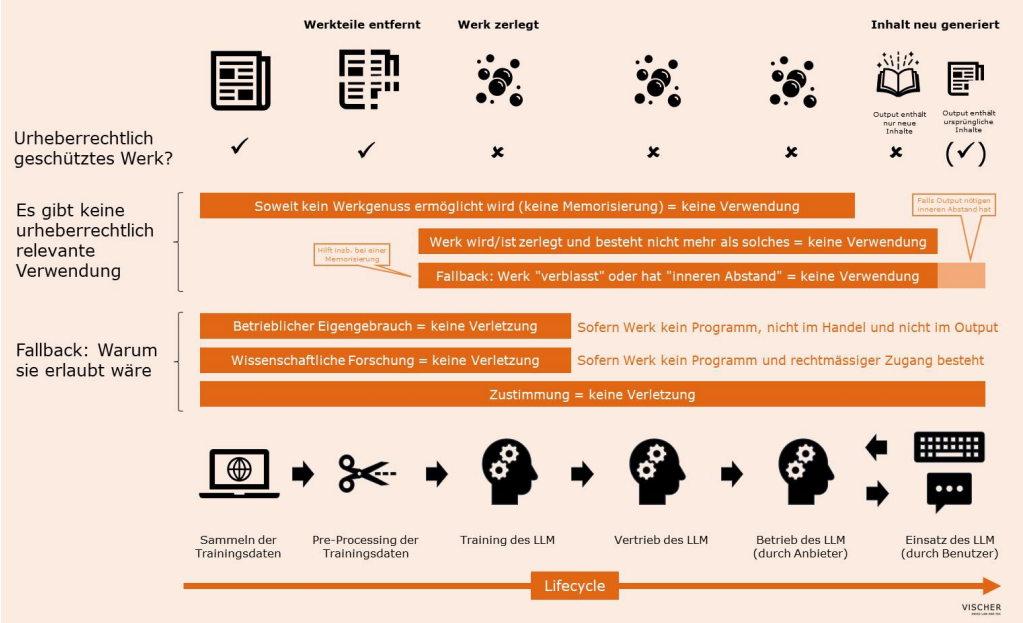

Abbildung 1: Training eines Sprachmodells aus Sicht des Schweizer Urheberrechts

PDF der Abbildung 1

[14] Bei den ersten beiden Ansätzen betreten wir, jedenfalls was das Training von Sprachmodellen betrifft, weitgehend Neuland.

[15] Bevor wir jedoch auf die rechtlichen Überlegungen zu sprechen kommen, sind gewisse tatsächliche Punkte zu klären, namentlich das «Problem» der Memorisierung.

12 Die Trainings-Datensätze als solche geniessen in der Schweiz keinen besonderen urheberrechtlichen Schutz (anders in der EU). Urheberrechtlicher Schutz setzt hier voraus, dass die Gestaltung oder die Auswahlkriterien der Datensammlung individuellen Charakter aufweist. Siehe hierzu auch: Urteil des Obergerichts Zürich vom 1. September 1992, in: SMI 1993, S. 331 ff., S. 334; Städeli/Mary (Fn. 7), S. 247 f.

B. Findet eine Memorisierung statt oder nicht?

[16] Im Zusammenhang mit der urheberrechtlichen Zulässigkeit des Trainings eines grossen Sprachmodells sind bei genauer Betrachtung zwei Konstellationen voneinander zu unterscheiden: *Erstens* die Konstellation, in der das Training des Modells zu keiner Memorisierung der Trainingsinhalte führt, und *zweitens* die Konstellation, in der im Modell selbst eine Memorisierung der Trainingsinhalte stattfindet. Eine solche Unterscheidung ist bereits deshalb von Bedeutung, weil die Memorisierung der Erkenntnisse aus einer Serie von Trainingsinhalten im Modell selbst (wir nennen dies nachfolgend auch die maschinenlesbare Repräsentation der Wissensbilanz von Trainingsinhalten oder kurz *maschinenlesbare Wissensbilanz*) als urheberrechtlich relevante Vervielfältigung und die Weitergabe des Modells folglich als Verbreitung bzw. als Zugänglichmachung qualifiziert werden könnte. Findet demgegenüber keine Memorisierung statt, erübrigt sich die Frage der urheberrechtlichen Qualifikation der maschinenlesbaren Wissensbilanz, zumal schlichtweg keine (relevante) Repräsentation etwaiger geschützter Werke im Modell enthalten ist. Auch auf der Ebene der Verwendung des Modells bzw. des Outputs ergeben sich Unterschiede. Den beiden Konstellationen ist allerdings gemeinsam, dass im Rahmen des Trainings selbst Vervielfältigungen und weitere Bearbeitungen der Trainingsinhalte erfolgen.

[17] In den nachfolgenden Ausführungen werden wir immer wieder auf diese Unterscheidung zurückkommen. Der Fokus liegt dabei klar auf dem Training eines grossen Sprachmodells jedenfalls ohne wortwörtliche Memorisierung. Dies deshalb, weil bereits heute effektive Massnahmen getroffen werden, um eine solche zu verhindern. Es ist jedoch davon auszugehen, dass eine Memorisierung nicht vollständig verhindert werden kann, weder im wortwörtlichen Bereich (wie bei bekannten bzw. häufig vorkommenden Slogans) noch – und erst recht nicht – im Bereich der *inhaltlichen* Memorisierung. So mag ein Sprachmodell womöglich nicht Wort für Wort etwa die Sätze aus Harry Potter memorisiert haben (weil entsprechende Gegenmassnahmen getroffen wurden), aber es kennt die Welt des Harry Potter und die darin vorkommenden Figuren, Orte und weiteren charakteristischen Merkmale des Harry-Potter-Universums, so dass es mühelos eine sich darin spielende Geschichte erstellen kann, wenn Zusatzmassnahmen (z.B. ein entsprechendes sog. *Alignment*[13] oder Prompt- oder Output-Filter) dies nicht verhindern. Wir werden deshalb auch ausführlich auf die Situation der Memorisierung aus urheberrechtlicher Sicht eingehen und zeigen, wie damit umzugehen ist.

C. Keine urheberrechtlich relevante Handlung

1. Ansatz 1: Es wird kein Werkgenuss ermöglicht

[18] Wie wir gesehen haben, erfordert das Training von Sprachmodellen Bearbeitungen und vor allem (kurzzeitige) Kopien der betreffenden Trainingsinhalte. Damit ist beim Training von Sprachmodellen ein Eingriff in das urheberrechtliche Ausschliesslichkeitsrecht und insbesondere in das Vervielfältigungsrecht gemäss Art. 10 Abs. 2 Bst. a URG fraglich.[14] Für die Beantwor-

13 Dem Modell wird z.B. beigebracht, entsprechende Prompts nicht zu befolgen.

14 Sandra Marmy-Brändli/Isabelle Oehri, Das Training künstlicher Intelligenz, in: sic! 2023, S. 655 ff., S. 657; Mathis Berger, Künstliche Intelligenz und Immaterialgüterrecht, in: Jusletter IT vom 4. Juli 2024, N 7 ff.; Philippe Gilliéron, Intelligence artificielle: la titularité des données, in: RSDA 2021, S. 435 ff., S. 449. In der Praxis wird teilweise auch argumentiert, das Training *als solches* stelle eine urheberrechtlich relevante Nutzungsart dar. Das ist

tung dieser Frage ist entscheidend, ob in diesem Zusammenhang überhaupt von einer urheberrechtlich relevanten Werkverwendung gesprochen werden kann.[15] Soweit ersichtlich, hatte die Schweizer Gerichtspraxis bislang noch keine Gelegenheit, sich hierzu zu äussern.[16] Es ist deshalb durch Auslegung und anhand der urheberrechtlichen Grundsätze zur ermitteln, ob insbesondere die Kopiervorgänge im Rahmen des Trainings von Sprachmodellen als Eingriffe in das Vervielfältigungsrecht zu qualifizieren sind. Nur, wenn dies zu bejahen ist, ist die Anwendbarkeit einer Schrankenbestimmung bzw. die Zustimmung des Rechteinhabers für die Rechtmässigkeit des Trainings mit urheberrechtlich geschützten Inhalten vorausgesetzt.

[19] Gemäss Art. 10 Abs. 2 Bst. a URG hat der Rechteinhaber das ausschliessliche Recht, Werkexemplare wie Druckerzeugnisse, Ton-, Tonbild- oder Datenträger herzustellen. Trotz dieser analog anmutenden Formulierung ist das Urheberrecht technologieneutral ausgestaltet[17] und der Begriff der Vervielfältigung weit gefasst.[18] Er umfasst unabhängig von der Anzahl der hergestellten Werkexemplare sowohl körperliche als auch unkörperliche, vorübergehende wie auch dauerhafte Vervielfältigungen, welche unmittelbar oder mittelbar einem Werkkonsum, d.h. der Wahrnehmung des Werks durch die menschlichen Sinne dienen.[19]

[20] Die Qualifikation der im Rahmen des Trainings von Sprachmodellen erforderlichen Kopien als urheberrechtliche Vervielfältigungen erscheint auf den ersten Blick (auch) aufgrund des Gesetzeswortlauts naheliegend.[20] Auch systematische Überlegungen sprechen teilweise für diese Sichtweise: Mit der Einführung von Art. 24a URG (vorübergehende Vervielfältigungen) hat der Gesetzgeber klargestellt, dass auch nur temporäre Vervielfältigungen urheberrechtlich relevante Handlungen darstellen können (ansonsten Art. 24a URG überflüssig wäre). Die für das Training von Sprachmodellen erforderlichen Kopien unterscheiden sich vorderhand nicht von den Vervielfältigungen nach Art. 24a URG: Auch sie sind grundsätzlich nur vorübergehender Natur und ermöglichen den menschlichen Werkkonsum in der Theorie (jedenfalls, wenn die spätere Benutzung des Sprachmodells miteinbezogen wird, wofür aber andere Stellen verantwortlich sein können[21]). Das würde grundsätzlich dafür sprechen, auch die Kopiervorgänge beim Training eines Sprachmodells als urheberrechtliche Vervielfältigung zu qualifizieren.

[21] Einem solchen Verständnis stehen aber mehrere Argumente entgegen. In einem ersten Schritt gilt es festzuhalten, dass die Begriffe des URG urheberrechtsautonom auszulegen sind. Wie nicht

---

deshalb nicht von vornherein abwegig, weil Art. 10 Abs. 2 URG nicht abschliessend formuliert ist und durchaus noch weitere Nutzungsarten denkbar sind, die dann unter die Generalklausel gemäss Art. 10 Abs. 1 URG fallen. Allerdings setzt sich das Training – wie beschrieben – aus Bearbeitungen und Vervielfältigungen zusammen, und diese Nutzungsarten sind vom Gesetz bereits explizit genannt. Es erscheint daher weder notwendig noch zielführend, das Training eines grossen Sprachmodells selbständig als urheberrechtlich relevante Nutzungsart zu qualifizieren.

15 Diese Frage stellt sich auch die jüngste Literatur: Florent Thouvenin/Peter G. Picht, AI & IP: Empfehlungen für Rechtsetzung, Rechtsanwendung und Forschung zu den Herausforderungen an den Schnittstellen von Artificial Intelligence (AI) und Intellectual Property (IP), in: sic! 2023, S. 507 ff., S. 517; Ivan Cherpillod, Intelligence artificielle et droit d'auteur, in: sic! 2023, S. 445 ff., S. 446 f.

16 Marmy-Brändli/Oehri (Fn. 14), S. 657.

17 Reto M. Hilty, Urheberrecht, 2. Aufl., Bern 2020, N 304.

18 Manfred Rehbinder/Lorenz Haas/Kai-Peter Uhlig, Orell Füssli Kommentar Urheberrechtsgesetz mit weiteren Erlassen und internationalen Abkommen, 4. Aufl., Zürich 2022, URG 24a N 1; Marmy-Brändli/Oehri (Fn. 14), S. 658.

19 Statt vieler: Rehbinder/Haas/Uhlig (Fn. 18), URG 10 N 9; Berger (Fn. 14), N 8.

20 Thouvenin/Picht (Fn. 15), S. 517.

21 Dazu Kapitel V.E.

jeder Umgang mit einem Werk eine «Verwendung» im Sinne von Art. 10 Abs. 1 URG darstellt,[22] ist auch nicht jeder Vorgang, der rein technisch betrachtet eine Vervielfältigung ist, ohne weiteres als Vervielfältigung im Sinne des Urheberrechts zu qualifizieren. In technischer Hinsicht finden beim Training von Sprachmodellen zweifelsohne Vervielfältigungen statt. Nach dem Gesagten steht damit allerdings noch nicht fest, dass es sich dabei auch um Vervielfältigungen im Sinne des Urheberrechts handelt.[23] Das ist vielmehr durch Auslegung zu ermitteln.

[22] Im Rahmen dieser Auslegung ist von zentraler Bedeutung, dass die dem Rechteinhaber vorbehaltenen Nutzungshandlungen in Art. 10 Abs. 2 Bst. a-f URG nicht beim eigentlichen Werkgenuss ansetzen – dieser ist urheberrechtlich frei –, sondern Handlungen beschreiben, die im Sinne einer Werkvermittlung notwendig sind, um letztlich einen Werkgenuss (durch Dritte) zu ermöglichen.[24] Nun ist es aber so, dass die Kopien der Trainingsinhalte (und die damit zusammenhängenden Bearbeitungen) beim Training eines grossen Sprachmodells gerade nicht geeignet sind, den Konsum dieser Inhalte durch Dritte (Menschen) zu ermöglichen, auch nicht indirekt durch die Verwendung des Modells (indem der Inhalt später im Output auftaucht). Das gilt ohne weiteres dann, wenn keine Memorisierung stattfindet. Jedoch kann dies auch in den Konstellationen der Memorisierung vertreten werden: Die für das Training benutzten Inhalte liegen für ein solches nämlich auch dann im Modell nicht in einer Form vor, die für den menschlichen Konsum geeignet ist bzw. vom Menschen wahrgenommen werden kann. Zudem ergibt erst die Kombination mit dem Prompt den Output.[25] Somit dienen auch bei einer Memorisierung die Vervielfältigungen der Trainingsinhalte weder unmittelbar noch mittelbar der menschlichen Wahrnehmung der einzelnen, im Training wahrgenommenen Werke und eignen sich grundsätzlich auch nicht, eine solche Wahrnehmung zu ermöglichen. Sie ist wie gezeigt auch nicht das Ziel. Vielmehr formen sich diese Vervielfältigungen im Rahmen des Trainings zu einem Trainingsdatensatz, der einzig und alleine dem Machine Learning und damit übergeordneten statistischen Erkenntnissen dient. Das gilt auch für die Crawler-Datensammlungen, die oft als Quelle benutzt werden und noch halbwegs lesbare Inhalte enthalten. Auch sie sind jedoch in einer Form aufbereitet, die für den menschlichen Konsum nicht geeignet ist, geschweige denn gedacht.[26]

[23] Die im Rahmen des Trainings von grossen Sprachmodellen erforderlichen Kopien der Trainingsinhalte stellen nach dem Gesagten zwar technisch betrachtet Vervielfältigungen dar. Weil jedoch die urheberrechtliche Beurteilung der Vorgänge einem urheberrechtlichen Begriffsverständnis folgt und durch die (technischen) Vervielfältigungen der Trainingsinhalte kein menschlicher Werkgenuss ermöglicht wird (auch nicht mittelbar), was für den urheberrechtlichen Verwendungsbegriff allerdings vorausgesetzt wäre, handelt es sich bei diesem Lösungsansatz nicht um Vervielfältigungen im Sinne des Urheberrechts. Somit sind die Kopien der Trainingsinhalte, die für das Training eines grossen Sprachmodells erstellt werden, urheberrechtlich frei bzw. ohne Relevanz.

---

22 Denis Barrelet/Willi Egloff, Das neue Urheberrecht – Kommentar zum Bundesgesetz über das Urheberrecht und verwandte Schutzrechte, 4. Aufl., Bern 2020, URG 10 N 8; Marmy-Brändli/Oehri (Fn. 14), S. 658; Rehbinder/Haas/Uhlig (Fn. 18), URG 10 N 3.

23 So auch: Thouvenin/Picht (Fn. 15), S. 517.

24 BGE 143 II 617, E. 5.3.2; Hilty (Fn. 17), N 292; Ansgar Kaiser, Der fehlende Werkgenuss beim Text und Data Mining, in: Florent Thouvenin et al. (Hrsg.), Kreation Innovation Märkte, Berlin 2024, S. 251 ff., S. 253; Rehbinder/Haas/Uhlig (Fn. 18), URG 10 N 3.

25 Dazu ausführlich Kapitel V.C.3.c).

26 Ein Beispiel von CommonCrawl, abrufbar unter: https://vischerlnk.com/3Z0gmwd.

2. Ansatz 2: Das Training ist ein (freier) «Werkgenuss» der KI

[24] Ein zweiter Ansatz zieht einen Analogieschluss zum menschlichen Werkkonsum. Dieser ist – wie bereits erwähnt – unbestrittenermassen urheberrechtlich frei. Wenn ein Mensch z.B. eine Website liest, ist das Lesen noch keine urheberrechtlich relevante Handlung, die der Zustimmung des Rechteinhabers bedürfte; menschliches Lesen und Lernen ist aus urheberrechtlicher Perspektive vielmehr ohne Zustimmung des Rechteinhabers möglich.[27] Das Training von Sprachmodellen mit unter Umständen urheberrechtlich geschützten Inhalten unterscheidet sich bei genauer Betrachtung nicht wirklich von diesem Sachverhalt: Vergleichbar zum menschlichen Gehirn, das ebenfalls zuerst unzählige Texte hören und lesen muss, bevor es selbst in der Lage ist, eigene Texte zu schaffen, benötigt das Sprachmodell das Training mit (urheberrechtlich geschützten) Inhalten, um die Bedeutungen der Tokens und ihre Beziehungen zueinander zu erlernen.[28] Allerdings bedarf maschinelles Lesen und Lernen im Unterschied zur Situation beim Menschen einer technischen Vervielfältigung des Trainingsinhalts. Dies stellt jedoch keinen «fundamentalen»[29] Unterschied zur Situation beim Menschen dar, der eine unterschiedliche Behandlung zu begründen vermöchte, weil die Vervielfältigungen beim Werkgenuss des Sprachmodells technisch bedingt und diesem damit inhärent sind. Ein Werkgenuss des Sprachmodells ist Stand heute ausschliesslich mit solchen technischen Vervielfältigungen denkbar.

[25] Auch aus diesen Gründen lässt sich vertreten, dass die (technischen) Vervielfältigungen (und die damit zusammenhängenden Bearbeitungen) von Trainingsinhalten im Rahmen des Trainings von Sprachmodellen nicht als Vervielfältigungen im Sinne des Urheberrechts zu qualifizieren sind und damit das Urheberrecht nicht verletzen. Welchen Output mit einem trainierten Modell erzeugt und wofür er benutzt wird, ist – *notabene* wie beim Menschen – eine andere, davon zu trennende Frage.

3. Ansatz 3: Fehlende urheberrechtliche Relevanz des «Wissens» der KI

a. Vorbemerkungen

[26] Für den Fall der wortwörtlichen Memorisierung von bestehenden Werken stellt sich die Frage, ob eine solche Memorisierung als urheberrechtlich relevante Vervielfältigung und die Weitergabe des Modells folglich als Verbreitung bzw. als Zugänglichmachung zu qualifizieren ist. Neben den wenigen Fällen der wortwörtlichen Memorisierung ist weiter fraglich, wie mit dem Umstand umzugehen ist, dass ein Inhalt aus einem bestehenden Werk mit noch so vielen Details in einer Art und Weise Eingang in das Sprach- und Sachwissen des Modells findet, als dass er (a) dort in einer Konkretisierung vorliegt, die als maschinenlesbare Repräsentation der Wissensbilanz von Trainingsinhalten gilt, und (b) sich aufgrund eines geeigneten Prompts im Rahmen einer konkreten Benutzung im Output in einer Form wiederfinden kann, bei welcher der indivi-

27 Insoweit noch gl.M.: David Bomhard, Text und Data Mining auf Grundlage von Webcrawling und Webscraping, in: InTeR 2023, S. 174 ff., S. 175 (Ausführungen beziehen sich auf das deutsche Recht).

28 Siehe zum Ganzen diesen Gedanken ebenfalls aufgreifend: Thouvenin/Picht (Fn. 15), S. 517.

29 Hier a.M.: Bomhard (Fn. 27), S. 175, der in der aus technischen Gründen erforderlichen Vervielfältigung von Trainingsdaten den fundamentalen Unterschied zum menschlichen Werkgenuss sieht und daher für das Training einer KI stets eine Zustimmung fordert.

duelle Charakter des Originalwerks noch erkennbar und nicht verblasst ist.[30] Das ist zwar nicht sehr wahrscheinlich, kann aber vorkommen.[31]

b. In den Trainingsdaten enthaltene Werke sind nicht mehr im Modell

[27] Die Beantwortung dieser Fragen ist zunächst davon abhängig, ob die in den Trainingsdaten enthaltenen Werke während des Trainings und danach, d.h. während des Vertriebs, Betriebs und Einsatzes des Modells, fortwährend in einer Form im Modell enthalten sind, die in den jeweiligen Schutzbereich der Originalwerke fällt. Das ist bereits deshalb fraglich, weil die Werke – wie bereits mehrfach erwähnt – selbst im Falle der wortwörtlichen Memorisierung in ihre Bruchstücke bzw. Token zerlegt sind.

[28] Es gilt, die Rechtslage in der analogen Welt in Erinnerung zu rufen: Dort ist es dem sachenrechtlichen Eigentümer ohne weiteres zulässig, etwa ein Gemälde oder eine Fotografie in seine bzw. ihre Einzelteile zu zerlegen (es sei denn, es handelt sich um ein Originalwerk im Sinne von Art. 15 Abs. 1 URG). Durch dieses Zerlegen, bspw. eines Gemäldes oder einer Fotografie, wird das Werk der bildenden Kunst bzw. das fotografische Werk zerstört. Es besteht als solches nicht mehr, selbst wenn seine Einzelteile noch vorhanden sind. Die einzelnen Teile können daher weder für sich alleine noch in ihrer ungeordneten Gesamtheit eine Vervielfältigung im urheberrechtlichen Sinne darstellen.

[29] Das Training eines grossen Sprachmodells ist gleich gelagert (auch für den Fall der Memorisierung): Indem die in den Trainingsdaten enthaltenen Werke erstens in ihre Token zerlegt werden und diese zweitens nur dazu dienen, die Justierung der Gewichte des Modells vorzunehmen und danach «weggeworfen» werden, findet im Ergebnis ein Zerlegen des Werks statt. Die Token bleiben bildlich gesprochen nicht beisammen, sondern aus ihrem Inhalt und ihrer Beziehung zueinander werden verstreut über das gesamte Modell die «Schräubchen» der Milliarden von Gewichten etwas in die eine oder die andere Richtung gedreht. Dieser Vorgang führt zu einer Zerlegung der einzelnen Werke und somit zu ihrer Zerstörung. Die Werke sind in der Folge als solche nicht mehr vorhanden. Vorhanden sind nur noch die Informationen, die sich aus der Vielzahl von Trainingsinhalten als statistisch relevant erwiesen haben, und zwar abgespeichert in einer Form, die nichts mehr mit einem Text zu tun hat. Die Beziehungen zwischen den einzelnen Informationselementen sind rein statistisch-mathematischer Natur, etwa dass das Wort «König» eine viel engere Beziehung zu «Königin» hat als zu «Fahrrad». Diese maschinenlesbare Repräsentation der Wissensbilanz von Trainingsinhalten kann daher bei dieser Betrachtungsweise keine Vervielfältigung der Werke im Sinne des Urheberrechts sein. Sie ist somit urheberrechtlich ohne Relevanz.[32]

---

30 Der Begriff des «Verblassens» ist im urheberrechtlichen Sinne zu verstehen, d.h., dass ein Werk als Teil eines grösseren Werks in den Hintergrund rückt.

31 Ob der Ersteller des Modells für diese Nutzung tatsächlich verantwortlich gemacht werden kann, können wir an dieser Stelle einmal offenlassen. Wir gehen darauf in Kapitel V.E ausführlich ein.

32 A.M.: Tim W. Dornis/Sebastian Stober, Urheberrecht und Training generativer KI-Modelle, Technologische und juristische Grundlagen, in: Roland Broemel et al. (Hrsg.), Recht und Digitalisierung, Band 19, Baden-Baden 2024, S. 71 ff.: In ihrem Gutachten im Auftrag der Initiative Urheberrecht (die Interessen von Rechteinhabern vertritt) kommen die Autoren zum Schluss, dass es keine Rolle spiele, was in einem Modell tatsächlich geschieht. Sie argumentieren, es komme nicht darauf an, ob eine Vervielfältigung beabsichtigt sei oder nicht (S. 75), übersehen aber, dass eine Vervielfältigung sachlogisch nur dann vorliegen kann, wenn am anderen Ende eine Kopie entsteht («Ver*viel*fältigung» – aus eins mach mehrere), was selbst nach ihrer Ansicht hier wohl nicht der Fall ist. Ihre Argumentation stellt stattdessen darauf ab, dass das Training und die Nutzung eines Modells als eine Einheit betrachtet

[30] Daran ändert auch der Umstand nichts, wonach in den Trainingsdaten enthaltene Werke im Falle der Memorisierung im Output wieder auftauchen können. Diesfalls kann zwar eine urheberrechtlich relevante Vervielfältigung vorliegen. Jedoch stellt das erneute Zusammensetzen der Einzelteile bzw. Token zum vorbestehenden Werk eine neuerliche Generierung bzw. «Schöpfung» eben jenes dar und nicht etwa ein Abrufen eines im Modell gespeicherten Werks. Auf die urheberrechtliche Zulässigkeit einer abermaligen «Schöpfung» gehen wir weiter unten ein.[33]

c. Theorie des «Verblassens» bzw. des «inneren Abstands»

[31] Für den Fall, dass dieser Ansicht nicht gefolgt bzw. die Auffassung vertreten wird, dass die in den Trainingsdaten enthaltenen Werke auch nach dem Zerlegen in ihre Einzelteile bzw. Token in einer Form im Sprachmodell memorisiert sein können, die potenziell in den Schutzbereich der vorbestehenden Werke fällt, muss Folgendes beachtet werden: Der Inhalt der vorbestehenden Werke ist in einer Form im Modell gespeichert, die vom Menschen nicht wahrgenommen werden bzw. in der jedenfalls ein Mensch das einzelne Werk mit seinen individuellen Zügen nicht mehr erkennen kann, und erst die Kombination mit dem Prompt erzeugt den Output. Jeder memorisierte Inhalt verschwindet quasi in einer «Suppe» von Milliarden von Parametern (Gewichte) und wird mit allen anderen Inhalten zu einer Gesamtheit (dem Modell) geformt, in welcher die individuellen Charakteristika des einzelnen Werks nicht mehr erkennbar sind bzw. verblassen. Dieses Verblassen führt aus dem urheberrechtlichen Schutzbereich hinaus. Dass sich mit mathematischen Methoden und den nötigen weiteren Hilfsmitteln (konkret: dem Prompt) einzelne charakteristische Bestandteile der «Suppe» eruieren lassen (konkret: sich ein Output erzeugen lässt, der in den Schutzbereich eines vorbestehenden Werks fällt), ändert nichts daran, dass die einzelnen Werke bzw. ihre charakteristischen Merkmale in der «Suppe» (hier: im Modell) in den Hintergrund treten.[34] Auch die Erstellung einer solchen «Suppe» aus unzähligen Einzelwerken ist urheberrechtlich kein relevanter Vorgang.[35] Erstellung und Weitergabe der «Suppe» sind also freigestellt, es sei denn, die dazu nötigen Vervielfältigungen im Vorfeld wären von dieser Freistellung nicht mitabgedeckt oder der gesamte Vorgang des Trainings würde als neue Form der Verwendung im Sinne von Art. 10 Abs. 1 URG betrachtet. Letzteres erscheint allerdings schon deswegen fraglich, weil es wie gezeigt nicht dem Genuss eben dieses Werks dient, sondern im Grunde dazu, das Werk zum Datenpunkt einer statistischen Erhebung zu machen.

---

wird (S. 73), was nicht zutrifft. Sie berufen sich ferner darauf, dass es nur einer körperlichen Festlegung des Werkes bedürfe, die geeignet sei, das Werk den menschlichen Sinnen auf irgendeine Weise mittelbar oder unmittelbar wahrnehmbar zu machen (S. 78). Diese Argumentation übersieht, dass es hierzu den Prompt braucht, der aber nicht im Modell enthalten ist. Der Prompt ist nicht nur Zugangsschlüssel zum ansonsten vollständigen «Werk», sondern eine inhaltliche Zutat, vergleichbar mit einem Zweikomponenten-Kleber, der nur und erst beim Mischen beider Komponenten zum Kleber wird. Technisch sind im Modell zwar die Bausteine für das Werk gewissermassen vorhanden, aber ohne den Prompt finden sie nicht zusammen.

33 Dazu Kapitel V.E.

34 Hier kann der Vergleich mit einem Fotomosaik helfen (https://de.wikipedia.org/wiki/Fotomosaik): Es besteht aus zahlreichen vollständig wiedergegebenen Einzelbildern, die aber zusammen ein neues Gesamtbild formen. Dessen untergeordnete Teilelemente, um deren individuellen Charakter es nicht mehr geht, verlieren in dieser Verwendung ihre urheberrechtliche Relevanz.

35 Hier bietet sich ein Vergleich an zur analogen Situation, in der Teile zahlreicher Bilder-Puzzles (von welchen alle ein geschütztes Werk zeigen) in einen Topf geworfen und vermischt werden. Der Vergleich hinkt insofern, als dass bei einem Sprachmodell von allen ähnlichen Teilen zuerst ein statistischer Durchschnitt verwendet wird. Dies führt dazu, dass die Originalbilder sich nicht mehr vollständig rekonstruieren lassen. Es ist also nicht wie beim einem geschredderten Papier, das sich theoretisch immer wieder zusammensetzen liesse.

Abbildung 2: Theorie des «Verblassens» bzw. Suppentheorie

[32] Wird selbst dieser «Suppentheorie» nicht gefolgt, so fällt eine Urheberrechtsverletzung durch die wortwörtliche bzw. inhaltliche Memorisierung aber auch deswegen weg, weil zwischen dem vorbestehenden Werk und dessen Memorisierung ein hinreichender «innerer Abstand» besteht. Die Theorie des inneren Abstands besagt, dass eine urheberrechtlich freie Benutzung selbst dann vorliegen kann, wenn der individuelle Charakter des Originalwerks nicht verblasst, sofern sich das Zweitwerk ausdrücklich mit dem verwendeten Originalwerk auseinandersetzt und damit den «inneren Abstand» schafft.[36] Das Zweitwerk muss seinem Wesen nach als selbständig erachtet werden. Eine Parodie oder eine Buchbesprechung können Beispiele für solche Zweitwerke sein, deren Verwendung nicht von der Zustimmung des Rechteinhabers des Originalwerks abhängt (wobei die Parodie zusätzlich vom Gesetzgeber mit einer Schrankenbestimmung bedacht worden ist).

[33] Ein solcher innerer Abstand liegt hier vor: Das Training eines Sprachmodells ist nicht auf die Vervielfältigung der Trainingsinhalte ausgelegt, sondern darauf, deren Bedeutungen, Aussagen und Zusammenhänge zu ermitteln und diese in einer vom Werk unabhängigen, maschinenlesbaren Form festzuhalten. Es ist daher durchaus zu vergleichen mit dem Sprach- und Literaturwissenschafter, der etwa die Welt des Harry Potter erforscht und z.B. analysiert, wie die Figuren gestaltet sind, wie sie miteinander interagieren, welche Konzepte in den Geschichten verwendet werden – und natürlich mit welchen sprachlichen Elementen dies geschieht. Dazu wird er diese Elemente strukturiert erfassen, verzeichnen und analysieren. Genau dies geschieht auch beim Training eines grossen Sprachmodells. Natürlich können die Aufzeichnungen des Sprachwissenschaftlers benutzt werden, um eine neue Geschichte zu Harry Potter zu entwerfen. Wer das ohne

36 Urteil des Obergerichts Zürich vom 7. Juli 2009, E. IV.2.1, in: sic! 2010, 889 ff., 892; Marco Handle, Der urheberrechtliche Schutz der Idee, in: Manfred Rehbinder/Reto M. Hilty/Cyrill P. Rigamonti (Hrsg.), SMI – Schriften zum Medien- und Immaterialgüterrecht, Band/Nr.100, Bern 2013, N 322 ff.

Zustimmung des Rechteinhabers oder gesetzliche Erlaubnis tut, wird möglicherweise eine Urheberrechtsverletzung begehen. An der urheberrechtlichen Zulässigkeit der Aufzeichnungen des Wissenschaftlers ändert dies jedoch nichts: Sie haben zum Werk einen hinreichenden inneren Abstand und sind damit ausserhalb des urheberrechtlichen Schutzbereichs zu verorten bzw. urheberrechtlich frei. Analog ist auch das Training eines Sprachmodells auf Basis von geschützten Werken urheberrechtlich freigestellt.[37]

[34] Die maschinenlesbare Repräsentation der Wissensbilanz der Trainingsinhalte ist also unproblematisch aus urheberrechtlicher Perspektive.

d. Rechtliche Folge der fehlenden Relevanz

[35] Wir haben bereits ausgeführt, dass Art. 10 URG diejenigen Handlungen beschreibt, die im Sinne einer Werkvermittlung notwendig sind, um letztlich einen Werkgenuss (durch Dritte) zu ermöglichen. Es fragt sich daher, den Genuss welchen «Werkes» die Vervielfältigungen und Bearbeitungen im Rahmen des Trainings genau ermöglichen. Bei genauer Betrachtung ermöglichen die Handlungen im Rahmen des Trainings eines grossen Sprachmodells maximal den Werkgenuss der maschinenlesbaren Repräsentation der Wissensbilanz der Trainingsinhalte. Der Trainingsinhalt als solcher kann davon nicht betroffen sein, weil er im Modell ja nur in Form der maschinenlesbaren Repräsentation memorisiert sein kann, und zwar als Erkenntnis nicht aus einem einzelnen Inhalt, sondern einer Vielzahl solcher – eine Synthese also. Zumal die maschinenlesbare Repräsentation von solchem Wissen nach Gesagtem urheberrechtlich frei ist, muss dies in logischer Konsequenz auch für die im Rahmen des Trainings stattfindenden Vervielfältigungen und Bearbeitungen gelten.

37 Mit diesem Ansatz kann ferner ein weiteres Problem «gelöst» werden, nämlich der Umstand, dass die Parameter eines Sprachmodells eine maschinenlesbare Repräsentation von Wissen zu einem Werk beinhalten können, ohne dieses (also etwa die Bücher) im Training je gesehen zu haben. Das kann bei bekannten Werken wie den Geschichten von Harry Potter der Fall sein: In der Sekundärliteratur, in Zeitungsberichten und auf Websites findet sich so viel Material hierzu, dass sich das Wissen um die Welt und die Geschichten von Harry Potter zwangsläufig einverleiben wird. Dies ist auch richtig, denn es stellt Allgemeinwissen dar. Die Art und Weise, wie Sprachmodelle funktionieren, führt dazu, dass das Modell dieses verteilte Wissen wie Mosaiksteine oder Puzzleteilchen an bestimmten Stellen in seiner «Suppe» zu einem Gesamtbild zusammenführt und über dieses Gesamtbild verfügt, so wie Menschen, die vieles über Harry Potter gelesen, gesehen und gehört und ein Gesamtbild der Welt sowie der Geschichte des Harry Potters im Kopf haben. So würde das Modell, wenn man es dazu auffordern würde und keine Gegenmassnahmen getroffen worden wären, auch daraus eine Geschichte zu Harry Potter erstellen können, welche die Urheberrechte des Rechteinhabers an Harry Potter verletzen könnte. In diesem Fall ist aber nicht die Werknutzung der Vorlagen das urheberrechtliche «Problem» (in der Annahme, dass diese selbst nicht urheberrechtlich geschützt sind und die Rechte des Rechteinhabers an Harry Potter nicht verletzen, weil sie über den nötigen inneren Abstand verfügen, sich auf das Zitatrecht gestützt werden kann oder sonst eine Schutzausnahme greift), sondern das Ergebnis des Trainings in Form der maschinenlesbaren Repräsentation des Allgemeinwissens zu Harry Potter. In Übereinstimmungen mit den Ausführungen weiter oben lässt sich allerdings sagen, dass bei dieser maschinenlesbaren Repräsentation die individuellen Züge der Originalwerke verblassen bzw. die digitale Repräsentation das Produkt der Auseinandersetzung des Modells mit den Trainingsinhalten darstellt und damit über den notwendigen inneren Abstand verfügt, um aus dem Schutzbereich des Urheberrechts zu fallen. Ferner lässt sich natürlich auch vertreten, dass die Trainingsinhalte urheberrechtlich nicht geschützt sind in Bezug auf ein bestimmtes Werk (hier: Sekundärliteratur in Bezug auf die Rechte an Harry Potter), dann würde ja selbst eine gegenständliche Kopie dieser Inhalte urheberrechtlich nicht von Relevanz sein. Erst ein neues Kreieren einer analogen Geschichte zu Harry Potter könnte in den Schutzbereich des Originalwerks fallen. Es wäre also eine reine Frage der Zulässigkeit des Outputs.

D. Training als urheberrechtlich relevante, jedoch erlaubte Handlung

1. Einschlägige Schrankenbestimmungen

[36] Ein weiterer Lösungsansatz für das Training von Sprachmodellen ist die Anwendung einer passenden Schrankenbestimmung des Urheberrechts. Wir beschränken uns hier auf die Analyse derjenigen Schranken, die tatsächlich in Frage kommen. Es sind dies zwei:

- **Betriebsinterner Gebrauch** (Art. 19 Abs. 1 Bst. c URG): Diese Schranke erlaubt das Vervielfältigen von Werkexemplaren in Betrieben für die interne Information oder Dokumentation. Nicht erlaubt ist die (weitgehend) vollständige Vervielfältigung von im Handel erhältlichen Werkexemplaren.[38] Vollständige bzw. weitgehend vollständige Vervielfältigungen sind jedoch zulässig, wenn sie beim Abrufen von erlaubterweise zugänglich gemachten Werken hergestellt werden wie z.B. beim Download solcher Inhalte, wobei nur die jeweils erste, bestimmungsgemässe Kopie, die mit dem Download entstanden ist, gedeckt ist.[39] Grundsätzlich muss die Vervielfältigung gegenüber der zuständigen Verwertungsgesellschaft vergütet werden. Durch diese Vergütung ist für die Rechteinhaber bereits ein gewisser Schutz sichergestellt, was eine grosszügige Auslegung des betrieblichen Eigengebrauchs nahelegt. Dies ist im Folgenden stets vor Augen zu halten. Wenn die Vervielfältigung (also die Speicherung) mit der Absicht erfolgt, daraus einen Output zu generieren, der (auch) an Dritte ausserhalb der Organisation gerichtet ist, dient die Vervielfältigung auf den ersten Blick nicht mehr nur der internen Information und Dokumentation.[40] Dem kann jedoch entgegengehalten werden, dass jede betriebsinterne Nutzung einen gewissen kommerziellen Zweck erfüllt. Selbst wenn Inhalte vorderhand zur Information und Weiterbildung der Angestellten verwendet werden, wird damit regelmässig (mindestens) ein positiver Einfluss auf den Geschäftsbetrieb bezweckt.[41] Wird z.B. den Mitarbeitenden einer Anwaltskanzlei relevantes Fachwissen in Gestalt von Auszügen aus wissenschaftlicher Literatur zur Verfügung gestellt, geschieht dies i.d.R. mit dem Ziel, die Mitarbeitenden auf dem neusten Stand zu halten und so eine kompetente und schnelle Beratung von Klienten zu ermöglichen. Die Mitarbeitenden analysieren und verarbeiten die erhaltene Information, um einen entsprechenden Output zu generieren.[42] Dieser Sichtweise scheint auch das Handelsgericht Zürich in einem kürzlich ergangenen – und kritisierten[43] – Urteil zu folgen, in dem es ausdrücklich festhält, dass sich der Zweck der Information nicht in der Wissensvermittlung (für Mitarbeitende) selbst erschöpfen muss, sondern auch bloss der Arbeitserleichterung dienen

38 Zu weiteren Ausnahmen siehe weiter unten.

39 Rehbinder/Haas/Uhlig (Fn. 18), URG 19 N 45.

40 Marmy-Brändli/Oehri (Fn. 14), S. 659 f.; vgl. auch: Philip Kübler, Wie generative KI-Systeme Rechte nutzen, medialex 05/23, 6. Juni 2023, N 12, abrufbar unter: https://medialex.ch/2023/06/06/wie-generative-ki-systeme-rechte-nutzen: Weil der Zweck des maschinellen Kopierens im Training eines generativen KI-Systems nicht die intern orientierte Information, sondern die extern orientierte Produktion sei, könne ein KI-System den Zweck der internen Information oder Dokumentation nicht für sich beanspruchen.

41 Dahingehend: Daniel Schönberger, Deep Copyright: Up- and Downstream Questions Related to Artificial Intelligence (AI) and Machine Learning (ML), S. 18, abrufbar unter: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3098315.

42 Siehe zum Ganzen auch: Elias Mühlemann/Nicole Ritter, Teil 10: Urheberrecht und KI: Verantwortlichkeit von Anbietern und Nutzern, abrufbar unter: https://www.vischer.com/know-how/blog/teil-10-urheberrecht-und-ki-verantwortlichkeit-von-anbietern-und-nutzern/.

43 Rehbinder/Haas/Uhlig (Fn. 18), URG 19 N 30.

darf.[44] Mit Blick auf das Training eines Sprachmodells lässt sich argumentieren, dass die urheberrechtlich relevante Handlung, nämlich die Vervielfältigung, tatsächlich nur zu internen Zwecken erfolgt, und zwar um dem Modell Sprach- und Sachwissen zu vermitteln. Das, was das Modell später an Output produziert, sind nur aus dem Inhalt gewonnene Informationen, die selbst nicht urheberrechtlich geschützt sind. Ähnlich, wie dies zum Beispiel eine Anwältin machen würde, nachdem sie sich über ein Thema in der internen Bibliothek informiert hat, um dem Klienten eine Antwort zu schreiben, erstellt das Modell bei seiner späteren Nutzung gestützt auf die so verarbeiteten Informationen einen Output. Auch hier stellt sich aber das Problem der Memorisierung: Ist das ursprünglich für das Training verwendete Werk im Output resp. der Zusammenfassung noch erkennbar und wird somit der Öffentlichkeit zugänglich gemacht, wird der Rahmen dieser Schranke überschritten.[45] Die Schranke des betriebsinternen Eigengebrauchs funktioniert weiter dort nicht richtig, wo für das Training eines Sprachmodells Inhalte benutzt werden, die noch im Handel gegen Bezahlung bezogen werden können und trotzdem vollständig oder weitgehend vollständig kopiert werden. Zwar wäre es – wie erwähnt – erlaubt, sich trotzdem eine Kopie zu machen, wo solche Inhalte online erlaubterweise zugänglich sind. Mit einer Kopie ist es aber in der Regel nicht getan, da die Inhalte nicht «in einem Zug» verarbeitet werden können, sondern es zahlreiche Zwischenschritte mit entsprechenden Kopien gibt. Eine Möglichkeit wäre es immerhin, dass von solchen Inhalten in einem ersten Durchgang des Trainings nur ein Teil verwendet wird (z.B. 50 Prozent) und sodann im zweiten Durchgang der Rest bzw. die andere Hälfte, da die Ausnahme von der Schranke nur weitgehend vollständige Kopien erfasst.[46] Bei einer funktionalen Betrachtungsweise liesse sich sogar die Verwendung von 90 Prozent eines Werks von einer weitgehend vollständigen Kopie abgrenzen, wenn sich die 90 Prozent auf jeden Satz oder Absatz beziehen, wie dies gewisse Anti-Memorisierungs-Techniken wie z.B. die *Goldfish-Loss*-Methode[47] tatsächlich vorsehen. Niemand würde auf den Erwerb eines Buches verzichten, wenn man stattdessen eine Kopie erhalten würde, bei welcher von jedem Satz oder Absatz 10 Prozent fehlen, zumal die Kopie in der Regel nicht mehr vernünftig zu gebrauchen wäre. Jedenfalls findet die Schranke des betriebsinternen Eigengebrauchs keine Anwendung auf Computerprogramme. Dies ist von einiger praktischer Relevanz, werden beim Training von grossen Sprachmodellen doch in grossen Mengen auch Computerprogramme eingelesen. Schliesslich erlaubt die Schranke des betriebsinternen Eigengebrauchs auch keine Vervielfältigungen von Werken der bildenden Kunst sowie von graphischen Aufzeichnungen von Musikwerken (Noten). Die Schrankenbestimmung des betrieblichen Eigengebrauchs unterliegt somit diversen relevanten Einschränkungen. Sie ist daher nur beschränkt nützlich für das Training von Sprachmodellen.

- **Wissenschaftsschranke** (Art. 24d URG): Diese Schranke erlaubt die (auch vollständige) Vervielfältigung von Inhalten, sofern sie der wissenschaftlichen Forschung dient und durch die dafür verwendeten technischen Verfahren bedingt ist, d.h. die Erreichung des wissen-

44 Urteil des Handelsgerichts Zürich HG190187-O vom 6. September 2021, E. 4.3.5.

45 In diesem Sinne auch: Mühlemann/Ritter (Fn. 42).

46 Eine weitgehend vollständige Vervielfältigung liegt gemäss BGE 133 III 473, E. 3.1 und Barrelet/Egloff (Fn. 22), URG 19 N 30 vor, wenn angesichts des Umfangs der Kopie für eine durchschnittliche Konsumentin der Kauf des vollständigen Exemplars uninteressant wird. Das ist spätestens dann der Fall, wenn der grösste Teil des urheberrechtlich geschützten Stoffes vervielfältigt wird.

47 Siehe: https://arxiv.org/abs/2406.10209.

schaftlichen Zwecks verlangt den Einsatz eines technischen Verfahrens, welches wiederum die Vervielfältigung erfordert. Gemeint sind gemäss der Lehre Dinge wie die Aufbereitung von Daten, Umformatierungen, Scans mit Texterkennung oder Kennzeichnungen.[48] Eine Vergütung ist nicht geschuldet.[49] Allerdings muss der Zugang zu den Inhalten ein rechtmässiger sein, d.h. wenn dafür bezahlt werden muss (z.B. über eine Registrierung), ist dafür zu bezahlen.[50] Das EU-Recht kennt unter dem Titel «Text and Data Mining» (**TDM**) eine ähnliche Regelung, die jedoch enger ausgestaltet ist und dem Rechteinhaber ein Opt-out-Recht im Falle kommerzieller Nutzungen gibt.[51] Das tut das Schweizer Recht (bisher) nicht; ein entsprechender Opt-out durch einen Rechteinhaber wäre in der Schweiz somit nicht wirksam.[52] Ob vertragliche Nutzungsverbote oder Auflagen wirksam sind, ist in der Lehre umstritten.[53] Nach der hier vertretenen Ansicht sind sie insoweit unbeachtlich, als sie den Kerngehalt von Art. 24d URG tangieren, d.h. die Schranke ihrer Wirkung berauben: Es ist gerade Sinn und Zweck der Schrankenbestimmung, die Abhängigkeit von der Zustimmung des Rechteinhabers für Vervielfältigungen zu Forschungszwecken zu verhindern,[54] und es kann nicht sein, dass dieser Zweck mit einem Satz in Registrierungs- oder Lizenzbedingungen ausser Kraft gesetzt werden kann. Zudem verlangt selbst der Wortlaut nur einen rechtmässigen *Zugang* zum Werk – ob die allenfalls gewährte Lizenz wieder dahinfällt, spielt keine Rolle. Die Kernfrage bei Art. 24d URG ist jedoch, inwieweit das Training eines Sprachmodells überhaupt ein erfasster Forschungszweck ist.[55] Klar ist zunächst, dass sowohl nicht kommerzielle als auch kommerzielle wissenschaftliche Forschung erfasst ist.[56] Als wissenschaftliche Forschung gilt die systematische, methodische Suche nach neuen Erkenntnissen, gleich welcher Disziplin.[57] Der wissenschaftliche Zweck muss klar und konkret sein und über etwaige andere Zwecke dominieren.[58] Damit legitimiert die Schranke auch Projektvorhaben, die Forschung und kommerzielle Anwendungen miteinander verknüpfen; soweit das systematische Anlernen des Systems der Produktentwicklung dient und die gewonnenen Erkenntnisse für weitere Produktentwicklungen verallgemeinert werden, ist es unerheblich, wenn ein Hauptziel des Projekts in der späteren Kommerzialisie-

48 Rehbinder/Haas/Uhlig (Fn. 18), URG 24d N 4; Barrelet/Egloff (Fn. 22), URG 24d N 6.

49 BBl 2018, S. 603 und S. 628.

50 Rehbinder/Haas/Uhlig (Fn. 18), URG 24d N 2.

51 Art. 3 f. der Richtlinie (EU) 2019/790.

52 Siehe für einen ausführlichen Vergleich der unterschiedlichen Regelungen: Damian Hartmann, Text and Data Mining and Copyright in Switzerland and the European Union, in: sic! 2023, S. 157 ff., S. 165 f.

53 Für den zwingenden Charakter von Art. 24d URG und damit die Wirksamkeit vertraglicher Nutzungsverbote oder Auflagen verneinend: Michael Isler, Text und Data Mining in der medizinischen Forschung, in: LSR 2022, S. 111 ff., S. 116; Barrelet/Egloff (Fn. 22), Vorbemerkungen zu URG 19–28 N 6; Marmy-Brändli/Oehri (Fn. 14), S. 664; Michael Isler, Wissenschaftsschranke (Art. 24d URG), in: Peter Mosimann (Hrsg.), Das revidierte Urheberrecht, Bern 2020, N 230; Melanie Graf/Kirsten Johanna Schmidt, Data Mining und wissenschaftliche Forschung – de lege lata und de lege ferenda, in: sui-generis 2017, S. 185 ff., S. 199 f.; a.A.: Berger (Fn. 14), N 13; Hilty (Fn. 17), N 442 f., jedoch nicht spezifisch mit Bezug auf Art. 24d URG.

54 Vgl. hierzu: BBl 2018, S. 602.

55 Dahingehend auch: Kübler (Fn. 40), N 16 ff.; Städeli/Mary (Fn. 7), S. 249; Berger (Fn. 14), N 13.

56 Statt vieler: Michèle Burnier, Révision du droit d'auteur suisse: questions choisies sous l'angle du Big data, in: PI – Propriété intellectuelle Band/Nr. 13, S. 275 ff., S. 286; Rehbinder/Haas/Uhlig (Fn. 18), URG 24d N 7; Isler, Wissenschaftsschranke (Fn. 53), N 232.

57 BBl 2018, S. 628; Hilty (Fn. 17), N 514.

58 BBl 2018, S. 603 und S. 628 f.; Marmy-Brändli/Oehri (Fn. 14), S. 66; Isler, Text und Data Mining (Fn. 53), S. 117.

rung der Anwendung liegt.[59] Dennoch wird in der Literatur der Standpunkt vertreten, «das Training für kommerzielle Anwendungen sog. ‹künstlicher Intelligenz›» falle nicht unter einen wissenschaftlichen Zweck,[60] wobei dies nicht wirklich begründet wird und in dieser Pauschalität auch unzutreffend ist. Denn ein KI-Modell ist per Definition eine «Erkenntnis», weil es in Form seiner Parameter (Gewichte) das Ergebnis seiner Analyse der Trainingsinhalte verkörpert und anhand von generalisierten (nur maschinell interpretierbaren) Regeln und Mustern Auskunft darüber gibt, wie in eben diesen Trainingsinhalten Sprache verwendet wird, so dass gestützt darauf beliebige neue Texte mit derselben Sprache generiert werden können. Diese Erkenntnis wird unzweifelhaft systematisch wie auch methodisch erreicht. Dass diese Erkenntnis (sprich: das Sprachmodell) später noch verkauft werden kann, ändert nichts daran, weil kommerzielle Forschung ebenfalls abgedeckt ist. Damit ist das Training von grossen Sprachmodellen von Art. 24d URG erfasst. Inzwischen liegt auch ein erstes (allerdings nicht rechtskräftiges) Urteil aus Deutschland vor, in dem sogar das Anbieten eines Datensatzes für das Training von KI-Modellen als Vervielfältigung für wissenschaftliche Forschung im Sinne der dort geltenden TDM-Regelung erachtet wurde, obwohl mit dem Datensatz noch kein Erkenntnisgewinn verbunden war, sondern er nur Input für Trainings war.[61] Einschränkend gilt auch hier, dass die Wissenschaftsschranke für Computerprogramme nicht gilt.

[37] Beide Schrankenbestimmungen erlauben ohne Zustimmung des Rechteinhabers lediglich Vervielfältigungen, nicht jedoch andere Bearbeitungen.[62] Nach Art. 11 Abs. 1 Bst. a URG hat der Rechteinhaber das alleinige Recht zu bestimmen, ob, wann und wie sein Werk verändert werden darf. Damit stellt sich die Frage, ob die Bearbeitungen im Rahmen der Vorbereitung eines Trainings Bearbeitungen des Werks im urheberrechtlichen Sinne sind und daher ohne Zustimmung des Rechteinhabers verboten wären. Die Rede ist namentlich von Umformatierungen (z.B. Umwandlung in Wortbruchstücke bzw. Tokens oder die Entfernung von Formatierungen) wie auch Schwärzungen (z.B. Entfernung von personenbezogenen Daten). Nach der hier vertretenen Ansicht stellen diese Bearbeitungen keine Bearbeitungen des Werks im Sinne von Art. 11 URG dar bzw. sind bereits durch den Vervielfältigungsprozess abgedeckt: *Erstens* kommen sie bei jeder Vervielfältigung im digitalen Bereich vor. Wird ein Text gescannt und mit einer Texterkennung bearbeitet, wird er in seine einzelnen Elemente zerlegt (bis auf Pixelstufe), wobei die jeweiligen Elemente einzeln weiterverarbeitet und störende Elemente für die Bilderkennung ausgeblendet werden. Das Werk ist im Innern des Systems zwar inhaltlich dasselbe, aber es wird völlig anders dargestellt. Das Vervielfältigungsrecht würde jede Bedeutung verlieren, wären solche untergeordneten, letztlich technisch bedingten Bearbeitungen nicht mitgemeint. Dies ist in der Lehre mindestens implizit anerkannt: So bezeichnet sie bei der Wissenschaftsschranke Vorgänge wie die Aufbereitung von Daten, Umformatierungen, Scans mit Texterkennung oder Kennzeichnungen von Daten als erfasst, ja sogar als erforderlich, damit die Schranke greift. *Zweitens* liegt keine Urheberrechtsverletzung vor, wenn eine Umgestaltung durch einen Dritten den individuellen

59 Isler, Text und Data Mining (Fn. 53), S. 117.

60 Rehbinder/Haas/Uhlig (Fn. 18), URG 24d N 6; ähnlich auch: Berger (Fn. 14), N 13.

61 Urteil des Landgerichts Hamburg 310 O 227/23 vom 27. September 2024 (nicht rechtskräftig).

62 Wobei der betriebsinterne Eigengebrauch das Verbreitungsrecht und das Recht zur Zugänglichmachung im Betrieb ebenfalls erfasst. Vgl. etwa: BGE 133 III 478.

Charakter des vorbestehenden Werks verblassen lässt.[63] Die Bearbeitungen zur Vorbereitung des Trainings bewirken, dass die urheberrechtlich geschützten Trainingsinhalte in ihre Teile (bzw. Token) zerlegt werden. Der urheberrechtliche Schutz kann sich zwar auch auf einzelne Teile eines Werks beziehen, soweit diesen Teilen individueller Charakter zukommt (Art. 2 Abs. 4 URG). Je kleiner diese Teile jedoch sind, desto weniger kommt ihnen individueller Charakter zu. Einem einzelnen Token kommt kein individueller Charakter zu, was ebenfalls die Schlussfolgerung zulässt, dass die angesprochenen Bearbeitungen zulässig sein müssen.[64] *Drittens* sind die Interessen des Rechteinhabers durch solche Bearbeitungen in keiner Weise tangiert, weshalb ihm diesbezüglich nach der *ratio legis* des Urheberrechts auch kein Ausschliesslichkeitsrecht zukommen soll. Vor diesem Hintergrund erfordert die Aufbereitung der Trainingsinhalte nur für die Zwecke des Trainings keine separate Schrankenbestimmung, wenn das Training bzw. die dafür erforderlichen Vervielfältigungen von einer solchen erfasst sind.

2. Zustimmung des Rechteinhabers

[38] Ein weiterer Ansatz, die Zustimmung des Rechteinhabers, dürfte nur für einen Teil der Trainingsinhalte zur Anwendung gelangen. Wenig Schwierigkeiten bereiten zunächst jene Fälle, in denen ausdrückliche Lizenzbedingungen vorliegen. Es ist in der Praxis allerdings zu unterscheiden zwischen der Lizenz des Anbieters eines Datensatzes und der Lizenz durch den Inhaber der Rechte an den darin enthaltenen Inhalten. Aus urheberrechtlicher Sicht braucht es in der Regel nur die Lizenz bzw. Erlaubnis des letzteren, nicht jedoch des ersteren, weil die Zusammenstellung eines Trainingsdatensatzes mangels Individualität nur selten zu einem eigenständigen Urheberrechtsschutz führt, der eine entsprechende Lizenzierung überhaupt erst erforderlich macht. Bei von öffentlichen Internet-Seiten gecrawlten Inhalten wird es eine ausdrückliche Zustimmung jedoch kaum je geben.[65] Es stellt sich somit die Frage, ob regelmässig immerhin eine implizite Zustimmung in die Verwendung der Inhalte vorliegt. Jedenfalls bei vom Rechteinhaber im Internet frei zugänglich gemachten Inhalten kann das mit guten Gründen vertreten werden: Jeder, der heute im Internet Inhalte bereithält, muss davon ausgehen, dass sie nicht nur von Menschen, sondern auch von Maschinen gelesen und für Suchmaschinen oder das Training von KI verwendet werden.[66] Wenn der Rechteinhaber darauf verzichtet, Massnahmen gegen eine solche vernünftigerweise zu erwartende Verwendungsweise zu treffen, dann spricht dies für eine implizite Zustimmung in diese Verwendung. Diese Argumentation funktioniert freilich nur, solange sich der Rechteinhaber selbst nicht (ausdrücklich) anders äussert, zum Beispiel über Nutzungsbedingungen, Abmahnschreiben oder eingebaute, maschinenlesbare Codierungen, die ein Crawlen bzw. Scraping nach entsprechenden Standards untersagen, und natürlich auch nur dort, wo Inhalte mit Erlaubnis des Rechteinhabers im Internet publiziert worden sind.[67]

[39] Eine weitere Grenze der impliziten Zustimmung dürfte die Memorisierung von Inhalten sein: Der Rechteinhaber wird seine Inhalte mitunter gerade zu dem Zweck publizieren, dass an-

63 Hilty (Fn. 17), N 376 ff.

64 So ähnlich auch: Berger (Fn. 14), N 15 f.

65 Ausserdem ist es wegen der äusserst grossen Menge an Werken, die für das Trainieren und Testen von KI-Systemen erforderlich ist, praktisch unmöglich, die entsprechenden Rechteinhaber zu identifizieren, zu kontaktieren und um ihre Zustimmung zur Nutzung der Werke zu bitten. Vgl. hierzu: Thouvenin/Picht (Fn. 15), S. 517.

66 Gleicher Auffassung: Marmy-Brändli/Oehri (Fn. 14), S. 663.

67 Marmy-Brändli/Oehri (Fn. 14), S. 663 Fn. 80.

dere davon lernen oder sich inspirieren lassen, weshalb ohne weiteres vertreten werden kann, dass Maschinen dies mindestens genauso tun dürfen wie Menschen, wenn sie mit dem Gelernten oder der Inspiration gleich umgehen: Maschinen dürfen das Gelernte anwenden bzw. die Inspiration wirken lassen, es bzw. sie aber nicht 1:1 wiedergeben, ohne dafür vom Rechteinhaber oder dem Gesetzgeber ermächtigt zu sein, wie beispielsweise im Falle des Zitatrechts.[68] Unter dem Aspekt der Zustimmung ist demnach nicht die Memorisierung der Inhalte an sich problematisch, sondern die Generierung von Output mit gleich wiedergegebenem und geschütztem Inhalt, ohne insbesondere die Regelungen des Zitatrechts zu befolgen (was Sprachmodelle in der Regel nicht einfach so tun). Dieser Vorgang ist freilich nicht demjenigen zuzurechnen, der das Modell trainiert.[69]

E. Verletzung des Urheberrechts durch den Output der KI

[40] An diesem Punkt ist die Trennung zwischen Wissen im Sprachmodell (enthalten in den Parametern) und dem Output des Modells wichtig: Der Output entsteht erst durch die Eingabe eines passenden Prompts. Kommt dieser Prompt in der Verwendung des Modells nie vor, wird auch das Modell nie einen solchen Output erzeugen. Der Prompt ist jedoch abhängig vom Benutzer und seiner Anwendung des Modells. Er kann demjenigen, der das Modell trainiert hat, jedenfalls in Bezug auf das Training des Modells nicht entgegengehalten werden, zumal es auf dieser «Ebene» noch an zwei für eine Rechtsverletzung zwingenden Komponenten fehlt, dem Prompt und dem Output. Die Rechtsverletzung würde sich zudem erst im Output widerspiegeln, beschlägt also nicht das Modell an sich. Es verhält sich wie mit einem Zweikomponenten-Kleber – jede Komponente für sich klebt nicht und ist kein Kleber; sie sind es erst in der Kombination. Die Frage, wie wahrscheinlich es ist, dass ein Benutzer des Modells einen bestimmten Prompt eingibt, beschlägt nicht das Modell, sondern dessen Nutzung und die Mitverantwortung desjenigen, der sie allenfalls ermöglicht (dazu nachfolgend mehr).

[41] Hinzu kommt, dass selbst wenn sich mit einem Modell ein urheberrechtsverletzender Output erzeugen liesse, nicht automatisch jeder Benutzer eines solchen Modells eine Urheberechtsverletzung begeht. Verwendet beispielsweise ein Unternehmen ein Allzweck-Sprachmodell, das auch mit Inhalten zu Harry Potter trainiert worden ist, für eine Industrieanwendung, und kommt es in der Industrieanwendung nie zu einem entsprechenden Prompt, so wird dieser Benutzer keine Urheberrechtsverletzung begehen. Auch wenn ein solches Modell als Ganzes weitergegeben und vervielfältigt würde, wäre dies noch keine urheberrechtlich relevante Handlung in Bezug auf den Output, den dieses Modell generieren könnte. Mit anderen Worten: Die Weitergabe eines Modells, das Harry Potter zitieren kann, ist noch keine Verletzung der Urheberrechte an Har-

68 Es könnte argumentiert werden, dass bestimmte Werke nur der Unterhaltung dienen und daher die Übertragung auf die Maschine scheitert, weil diese über keine Emotionen verfügt. Hierbei wird jedoch übersehen, dass selbst dort, wo ein Werk primär der Unterhaltung dient, dies nie ein ausschliesslicher Zweck ist, sondern ein Rechteinhaber immer auch Sekundärzwecke wie z.B. eine inhaltliche Analyse gestattet. Der Autor eines Romans schreibt diesen zwar primär zur Unterhaltung, aber er ist ebenso einverstanden damit, dass ihn der Buchkritiker des Feuilletons, der Lehrer mit seiner Klasse oder die an bestimmten Genres interessierte Literaturwissenschaftlerin hinsichtlich der verwendeten Sprache analysiert. Zu verschiedenen Harry Potter-Bänden sind beispielsweise mehrere solche Analysen erschienen. Diese stellen freilich auch aufgrund ihres inneren Abstands zum Originalwerk keine Verletzung desselbigen dar, selbst wenn sie das Werk im Rahmen ihrer Analyse auch ausserhalb des Zitatrechts wiedergeben.

69 Dazu Kapitel V.E.

ry Potter, weil sich eine solche erst im Output, d.h. in der konkreten Verwendung des Modells, manifestieren würde.

[42] Doch selbst wenn ein Unternehmen ein grosses Sprachmodell für die Erstellung und Verwendung eines Outputs benutzt, welcher fremde Urheberrechte verletzt, ist derjenige, der das Modell dem Unternehmen bereitstellt, nicht automatisch dafür mitverantwortlich. Es liegen hier zwei grundsätzlich separate Handlungen vor: Diejenige des Trainings des Modells und diejenige der Verwendung. Eine andere, dergestalt in der Lehre noch nicht gestellte und erörterte Frage zielt darauf, ob derjenige, der ein Modell einem Benutzer bereitstellt, wobei dieses sodann zur Begehung einer Urheberrechtsverletzung benutzt wird, als *Teilnehmer* oder gar als Mittäter an dieser Verletzung anzusehen ist. Die Bereitstellung bzw. Weitergabe des Modells könnte als Tatbeitrag oder als Hilfestellung zur Urheberrechtsverletzung angesehen werden.

[43] Die Beantwortung dieser Frage möchten wir vorliegend grundsätzlich ausklammern, zumal sie – wie soeben ausgeführt – nicht unmittelbar das Training eines grossen Sprachmodells beschlägt. Allerdings weisen wir darauf hin, dass diese Frage im Zusammenhang mit der Verantwortlichkeit von Internet-Hosting-Providern für Urheberrechtsverletzungen durch ihre Kunden bereits gerichtlich erörtert worden ist. Wir verweisen hierzu etwa auf die Zuger *Rapidshare*-Rechtsprechung, in welcher eine gewerbsmässige Gehilfenschaft eines Providers, dessen Server für die Verteilung von Raubkopien genutzt wurden, abgelehnt wurde, weil sein Geschäftsmodell nicht einzig auf die Beihilfe zur Urheberrechtsverletzung ausgerichtet war.[70] Des Weiteren hat das Bundesgericht im Zusammenhang mit einer Urheberrechtsverletzung entschieden, dass nicht jede beliebige Teilnahmehandlung, die lediglich «irgendwie» fördernden Einfluss hat, jedoch nicht in hinreichend engem Zusammenhang mit der Tat selbst steht, einen Tatbeitrag darstellt.[71] In diesem Entscheid ging es um einen Anbieter von Internet-Zugängen, der die technische Infrastruktur für den Zugang zum weltweiten Internet bereitgestellt hat, was im Rahmen zivilrechtlicher Ansprüche nicht für eine Teilnahme an einer Urheberrechtsverletzung unbekannter Dritter genügte.[72] Der Anbieter eines Sprachmodells, der das Modell zwar selbst trainiert hat, es aber nicht selbst betreibt und nur unter der Auflage einer urheberrechtskonformen Nutzung abgibt, wird vertreten können, dass auch er nur eine technische Infrastruktur bereitstellt, die für sich nicht urheberrechtsverletzend ist. Dies wird sie erst, wenn sie in entsprechender Weise (d.h. mit passendem Prompt und unzulässiger Verwendung des Outputs) verwendet wird. Der Anbieter eines Sprachmodells wird in der Kausalkette zwar näher am Täter sein als der Internet-Access-Provider, jedoch weniger nah als der Hosting-Provider, jedenfalls solange er das Modell nicht auf seinem Rechner betreibt und als Service anbietet.

[44] Die Provider-Rechtsprechung im Bereich des Persönlichkeitsrechts lässt sich zwar nicht ohne weiteres auf den Bereich der Urheberrechtsverletzungen übertragen. Das hat aber vor allem damit zu tun, dass in ersterem Falle nach herrschender Praxis bereits jede (auch unbewusste) Form des «Mitwirkens» an einer Persönlichkeitsverletzung nach Art. 28 Abs. 1 ZGB selbständig zivilrechtlich verfolgt werden kann, vorausgesetzt allerdings, dass es einen Kausalzusammenhang

70 Hierzu weiterführend: https://steigerlegal.ch/2022/03/07/rapidshare-urteil-volltext/, archiviert unter: https://perma.cc/8UC2-TBT8.

71 BGE 145 III 72, E.2.3.1.

72 Ebd., E. 2.3.2.

gibt.[73] Das Obergericht Solothurn erwog in einem Fall betreffend die Registrierungsstelle von IP-Adressen gar, dass ein Mitwirken genüge und es nicht einen Kausalzusammenhang brauche, der adäquat sei, um die tiefen Anforderungen an ein «Mitwirken» nicht zu unterlaufen,[74] was freilich an der Sache vorbeizielt, weil die Adäquanz eine Figur des Haftpflicht- und Strafrechts ist, um die es im betreffenden Fall nicht ging.

[45] Das Bezirksgericht Zürich kam im Zusammenhang mit Internet-Suchmaschinen zum Ergebnis, dass die Eingabe von Suchbegriffen in einer Suchmaschine nicht der Betreiberin zuzurechnen ist, jedenfalls soweit die Suchbegriffe nicht vorgeschlagen werden. Die Anzeige der Treffer durch die Betreiberin der Suchmaschine wertete es nicht als Persönlichkeitsverletzung.[75] Zuvor noch hatte dasselbe Gericht in einem anderen Entscheid der Betreiberin einen Beitrag zugerechnet, weil sie Informationen im Netz einer breiten Masse von Nutzern zugänglich gemacht habe, die sonst nicht hätten gefunden werden können.[76]

[46] Das Handelsgericht Zürich ging in einem neuen Entscheid betreffend FIFA und Google weiter: Der Betrieb einer Suchmaschine und die damit verbundene Verarbeitung von Informationen genügte dem Gericht nicht zur Passivlegitimation im Falle angeblich persönlichkeitsverletzender Artikel. Die streitgegenständlichen Artikel wurden in der Suchmaschine auf der ersten Seite der Suchtreffer nur bei der Eingabe von Suchbegriffen mit Wörtern aus dem Titel der Artikel angezeigt. Das Gericht befand, dass es damit einzig der Benutzer der Suchmaschine war, der den Zusammenhang zwischen den Artikeln und der FIFA herstellte. Google würde zwar die Suchmaschine zur Verfügung stellen und damit das Auffinden von Artikeln begünstigen. Dies sei jedoch noch keine rechtserhebliche Mitwirkungshandlung, weil diese nicht in hinreichend engem Zusammenhang mit der Tat selbst stehe; ein blosses «Wirken» genüge nicht, es brauche ein «Mitwirken». Ein Verhalten von Google, das – neben dem Betrieb der Suchmaschine – das Auffinden der Artikel konkret ermöglicht oder begünstigt habe, sei nicht behauptet worden, so das Gericht.[77]

[47] Diese Rechtsprechung lässt sich mit guten Gründen auf die Anbieter von grossen Sprachmodellen und darauf aufbauenden Allzweck-KI-Chatbots wie «ChatGPT» oder «Copilot» übertragen: Suchmaschinenbetreiber verzeichnen allerlei Inhalte aus dem Internet in ihren Datenbanken, indexieren diese und machen sie damit suchbar. Anbieter von KI-Chatbots verzeichnen dieselben Inhalte, nur dass sie die Inhalte nicht 1:1 in ihren Modellen abbilden, sondern lediglich davon abgeleitetes, aggregiertes Sprach- und Sachwissen. Dieses wiederum stellen sie zur freien Abfrage im Rahmen einer Dienstleistung zur Verfügung. In Bezug auf Inhalte, die allfällige Drittrechte verletzen, aber nicht wiedergegeben werden (weil Massnahmen jedenfalls gegen eine wortwörtliche Memorisierung getroffen wurden), gehen sie also noch einen Schritt weniger weit, da sie solche Inhalte nicht 1:1 weiterverbreiten bzw. zugänglich machen. Selbst bei memorisierten Inhalten stellen sie nur Bausteine zur Verfügung, die sie für den Benutzer gemäss seinem Prompt zusammenstellen, wie die Suchmaschine, die in ihrem Index einen passenden Inhalt findet und Zugang dazu verschafft. Und wer lediglich Kopien der von ihm trainierten Modelle anbietet (z.B. in Form von Dateien zur Installation auf einem eigenen Rechner), geht noch einen Schritt we-

73 Urteil des Bundesgerichts 5A_792/2011 vom 14. Januar 2013, E. 6.2 f., BGE 141 III 513, E. 5.3.1; zur Abgrenzung siehe Urteil des Bundesgerichts 5A_658/2014 vom 6. Mai 2015, E. 4.2.

74 Urteil des Obergerichts Solothurn ZKBER.2022.17 vom 3. November 2022, E. 5 ff.

75 Urteil des Bezirksgerichts Zürich CG190002 vom 26. Oktober 2020, E. 2.1.3.

76 Urteil des Bezirksgerichts Zürich CG160047 vom 1. Juni 2018, E. 6.2.9.

77 Urteil des Handelsgerichts Zürich HG220030-O vom 21. August 2024, E. 3.2.4.2.6.

niger weit – vergleichbar mit dem, der dem Betreiber einer Suchmaschine den rohen Index zur Verfügung stellt.

[48] Kommt es nun bei einem Benutzer eines solchen KI-Chatbots zur Eingabe von Prompts, die einen rechtswidrigen Output generieren, so ist dies nicht anders zu beurteilen als die Eingabe von spezifischen Suchbegriffen in einer Suchmaschine, die zu rechtswidrigen Artikeln führt. Analog kann die blosse Bereitstellung des KI-Chatbots (oder sogar nur des Modells) noch nicht als relevante Mitwirkung an der Verletzung gelten, jedenfalls wenn für den betreffenden Output ein spezieller Prompt nötig ist, den Benutzer nicht einfach so eingeben und der vom Anbieter auch nicht vorgeschlagen wird. Wenn dem bereits im Persönlichkeitsrecht so sein soll, wo jedes Mitwirken genügt, muss dies im Bereich des Urheberrechts erst recht gelten, da hier höhere Anforderungen an relevante Tatbeiträge bestehen. Die Verantwortlichkeit der Anbieter von Allzweck-KI-Chatbots sowie (und erst recht) den ihnen zugrunde liegenden KI-Modellen für die von ihren Nutzern mit speziellen Prompts generierten Outputs ist damit jedenfalls im Schweizer Recht stark eingeschränkt. Das erscheint richtig: Ein KI-Modell kann beliebig viele unterschiedliche Outputs generieren. In Kombination mit dem geeigneten Prompt können diese immer auch rechtsverletzend sein. Soweit es jedoch der Benutzer ist, der es darauf anlegt, und der Anbieter des Chatbots oder gar des Modells dies nicht spezifisch begünstigt,[78] fehlt es letztlich an der nötigen adäquaten Kausalität[79] bzw. am Mitwirken. Dafür lassen sich grosse Sprachmodelle zu vielseitig einsetzen. Der Vollständigkeit halber davon abzugrenzen ist notabene der Fall, in welchem ein Anbieter eines KI-Dienstes das von ihm benutzte Sprachmodell nebst dem Prompt des Benutzers mit weiteren Informationen füttert (z.B. aktuellen Informationen, Inhalten aus Datenbanken oder aus dem Internet), um den Output aufzuwerten. Ein solches sog. «Retrieval Augmented Generation» (RAG) und die urheberrechtliche Zulässigkeit einer Verwendung von fremden Werken im Input bzw. Prompt eines Sprachmodells ist gesondert zu beurteilen; darum geht es hier nicht.

F. Exkurs: Forum Shopping

[49] Für urheberrechtlich geschützte Werke besteht als Folge des im internationalen Immaterialgüterrecht geltenden Territorialitätsprinzips in jedem Staat ein räumlich begrenztes Schutzrecht nach Massgabe des jeweiligen nationalen Urheberrechts.[80] Das hat zur Folge, dass dieses «inländische» Urheberrecht nur durch eine zumindest teilweise im Inland begangene Handlung verletzt werden kann.[81] Mit anderen Worten ist der Handlungsort der mutmasslich urheberrechtsverletzenden Nutzung für die sachrechtliche Lokalisierung entscheidend.[82]

[50] Im Bereich des Trainings von grossen Sprachmodellen ist die relevante Nutzungshandlung das Speichern der Trainingsinhalte im Trainingskorpus des KI-Anbieters sowie das eigentliche Training des KI-Systems, nicht aber dessen späterer Einsatz durch Endkunden. Dementsprechend

78 Z.B. mit dem Betrieb eines Modells oder KI-Systems speziell für rechtswidrige Zwecke oder ein Training zur Förderung rechtsverletzender Outputs.

79 Einen natürlichen Kausalzusammenhang gibt es immer, denn ohne Modell gibt es keinen Output, genauso wie es ohne Suchmaschinen-Index keine Suchtreffer gibt.

80 Gerhard Schricker/Ulrich Loewenheim, Urheberrecht, 6. Aufl., München 2020, Vorbemerkung UrhG 120 N 109.

81 Schricker/Loewenheim (Fn. 80), Vorbemerkung UrhG 120 N 126, N 131 ff. und N 142 ff.

82 Niklas Maamar, Urheberrechtliche Fragen beim Einsatz von generativen KI-Systemen, in: ZUM 2023, S. 481 ff., S. 486.

richtet sich das anwendbare Recht danach, wo der KI-Anbieter das Training des grossen Sprachmodells durchführt.[83] Empirisch lässt sich das dadurch belegen, dass das KI-Training üblicherweise in Ländern mit «liberalen» Urheberrechtsregelungen und weitreichenden TDM-Schranken durchgeführt wird.[84]

[51] KI-Anbieter haben also die Möglichkeit, eine für sie günstige, weil «liberale» Rechtsordnung für das Training von grossen Sprachmodellen zur Anwendung zu bringen – selbst wenn das so trainierte KI-System anschliessend beispielsweise in der Schweiz auf den Markt gebracht wird.[85] Handelt es sich dabei um eine Rechtsordnung, die das Training von grossen Sprachmodellen mit urheberrechtlich geschützten Inhalten beispielsweise freistellt, steht einem späteren Inverkehrbringen dieses Sprachmodells in der Schweiz zumindest aus urheberrechtlicher Sicht nichts mehr im Wege. Dies ist überdies, wie oben vermerkt, unter dem AI Act relevant.[86]

G. Fazit

[52] Beim Training von grossen Sprachmodellen werden neben gemeinfreien Inhalten auch solche verwendet, die urheberrechtlichen Schutz geniessen. Doch auch das Training mit urheberrechtlich geschützten Inhalten ist rechtlich bereits *de lege lata* grundsätzlich zulässig – das gilt sowohl für das Training mit als auch für das Training ohne Memorisierung. Dies lässt sich zunächst damit vertreten, dass die Vervielfältigungen der Werke für das Training eines grossen Sprachmodells keine urheberrechtlich relevante Handlung darstellen, da diese Prozesse nicht der menschlichen Wahrnehmung der einzelnen Trainingsinhalte dienen, das Urheberrecht aber nur Handlungen schützt, die letztlich geeignet sind, den Werkgenuss des Menschen zu ermöglichen. Wird das abgelehnt, kann vertreten werden, dass im Training ein «Werkgenuss» der KI zu sehen ist, der analog zum menschlichen Werkgenuss urheberrechtlich frei sein muss, weil er ihm in relevanten Punkten entspricht. Die technisch erforderlichen Vervielfältigungen sind diesem «Werkgenuss» inhärent und begründen daher keinen wesentlichen Unterschied zum menschlichen Werkgenuss.

[53] Das Training mit Memorisierung ist insofern «problematischer», als dass im Vergleich zum Training ohne Memorisierung eine maschinenlesbare Repräsentation der Wissensbilanz von Trainingsinhalten hinzutritt, die ihrerseits auf ihre urheberrechtliche Zulässigkeit überprüft werden muss. Zur Erstellung dieser maschinenlesbaren Repräsentation werden die einzelnen Werke jedoch in ihre Einzelteile bzw. Token zerlegt. Dieser Vorgang hat zur Folge, dass die in den Trainingsdaten enthaltenen Werke zerstört werden und als solche fortan nicht mehr existieren. Die maschinenlesbare Reproduktion der Wissensbilanz von Trainingsinhalten kann bereits deshalb keine urheberrechtlich relevante Vervielfältigung sein.

[54] Sollte diesem Ansatz nicht gefolgt und der maschinenlesbaren Wissensbilanz dennoch so etwas wie Werkqualität und damit eine potenzielle Fähigkeit zur Rechtsverletzung zugemessen werden, ist des Weiteren auf die Theorie des «Verblassens» bzw. des «inneren Abstands» hinzuweisen: Bei der maschinenlesbaren Wissensbilanz verblassen die individuellen Charakteristika

83 So etwa auch: Maamar (Fn. 82), S. 486.

84 Marcus von Weiser, Generative KI und Urheberrechtsschranken, in: GRUR-Prax 2023, S. 516 ff., S. 520 N 39; Dornis/Stober (Fn. 32), S. 120.

85 Siehe zum Ganzen und mit Bezug auf die EU: Maamar (Fn. 82), S. 486.

86 Siehe Rz. 9 oben.

der urheberrechtlich geschützten Trainingsinhalte bzw. das Modell schafft zu diesen Inhalten einen hinreichenden inneren Abstand, womit die maschinenlesbare Wissensbilanz vom urheberrechtlichen Schutzbereich der Trainingsinhalte nicht erfasst ist. Die Betrachtung rein des Modells ist damit schon gar nicht dazu geeignet, einen Werkgenuss der Trainingsinhalte zu ermöglichen, weshalb für die im Rahmen des Trainings erforderlichen Vervielfältigungen und Bearbeitungen sowie auch für die Weitergabe des Modells an einen Dritten das beim Training ohne Memorisierung Gesagte gleichermassen Geltung hat. Weil aufgrund der Memorisierung jedoch der Output des Modells eine Urheberrechtsverletzung darstellen kann, stellt sich beim Training mit Memorisierung weiter die Frage nach der Verantwortlichkeit desjenigen, der das Modell trainiert und für die Nutzung bereitstellt. Grundsätzlich wollen wir diese Frage an dieser Stelle nicht abschliessend erörtern. Es lässt sich allerdings sagen, dass – analog zur Situation bei Hosting-Providern und Anbietern von Suchmaschinen, wo es schon entsprechende Rechtsprechung gibt – gute Gründe dafür sprechen, eine Haftung jedenfalls desjenigen, der ein Allzweck-Modell trainiert und als solches der Allgemeinheit anbietet, grundsätzlich abzulehnen.

[55] Wird dennoch von einer urheberrechtlich relevanten Handlung ausgegangen, kann mindestens bei bestimmten Werken vertreten werden, dass die Erlaubnis des Rechteinhabers vorliegt. Wo dies nicht explizit aufgrund von Lizenzen erfolgt ist, so doch implizit durch die öffentliche Bereitstellung von Inhalten im Internet ohne Einschränkungen. Zudem können zwei gesetzliche Schrankenbestimmungen angeführt werden: Erstens der betriebsinterne Gebrauch und zweitens die Wissenschaftsschranke. Beide weisen zwar praktisch relevante Einschränkungen auf (z.B. gelten sie nicht für Computerprogrammcode, der beim Training von Sprachmodellen oft relevant ist). Trotz allem erscheinen sie dort als praxistauglich, wo ein rechtmässiger Zugang zu den Trainingsinhalten besteht (Wissenschaftsschranke) bzw. sichergestellt wird, dass die relevanten Werke nicht im Output auftauchen (betriebsinterner Gebrauch).

[56] Schliesslich können KI-Anbieter für das Training von grossen Sprachmodellen mittels *forum shopping* eine «liberale» Rechtsordnung zur Anwendung bringen, die das Training von grossen Sprachmodellen mit urheberrechtlich geschützten Inhalten z.B. freistellt. Einem späteren Inverkehrbringen des Sprachmodells in der Schweiz steht dann aus urheberrechtlicher Sicht nichts mehr entgegen, weil die Handlung des Trainings beendet ist. Dasselbe gilt, wenn es in einer anderen Rechtsordnung verbreitet werden soll, welche ebenfalls dem Schutzlandprinzip folgt.

[57] Unsere Ausführungen zeigen, dass eine Anpassung des Schweizer Urheberrechts nicht zwingend ist. Es ist allerdings auch davon auszugehen, dass die Verwendung von Inhalten für das Training von KI-Modellen aufgrund der auf dem Spiel stehenden tatsächlichen oder gefühlten wirtschaftlichen Interessen ein rechtspolitischer Zankapfel bleiben wird: Die eine Seite wird verlangen, dass der Standort Schweiz für die Entwicklung von KI-Modellen attraktiv(er) gemacht wird, während die andere Seite Schutz ihrer Inhalte vor der Ausbeutung insbesondere durch US-Technologie-Konzerne fordern wird. Ob sich dies durch eine Anpassung des Urheberrechts, wie sie von beiden Seiten wohl verlangt werden wird, wirklich erreichen lässt, bleibt unklar, da das Recht der Realität in solchen Fällen regelmässig hinterherhinkt, soweit es gegen die Macht des Faktischen überhaupt wirksam ankommen kann. Schliesslich wird es auch jene geben, die eine Angleichung des Schweizer Urheberrechts an dasjenige der EU verlangen werden, namentlich in Bezug auf die TDM-Regelung – etwa durch Einführung eines Opt-out-Rechts im Rahmen der Forschungsschranke. Ein solches wirft allerdings zusätzliche Fragen auf. Die Schweiz ist bisher zudem nicht schlecht gefahren, indem sie Technologie nur zurückhaltend reguliert (der Bericht

des Bundesrates zum Regulierungsbedarf in Sachen künstlicher Intelligenz lag zum Zeitpunkt dieses Beitrags noch nicht vor).

VI. Datenschutz

A. Vorbemerkungen

[58] Für die datenschutzrechtliche Zulässigkeit des Trainings eines grossen Sprachmodells müssen nach Datenschutzgesetz (**DSG**) verschiedene Voraussetzungen erfüllt sein, die wiederum abhängig davon sind, ob es sich bei der verantwortlichen Stelle um eine private Person handelt oder um ein Bundesorgan (z.B. eine Forschungs- oder Bildungseinrichtung der öffentlichen Hand).

[59] Für private Personen gelten folgende Voraussetzungen:

- Es müssen die Bearbeitungsgrundsätze von Art. 6 DSG eingehalten sein;
- sollte dies nicht gegeben sein oder ein Widerspruch vorliegen, ist eine Rechtfertigung erforderlich (Art. 30 Abs. 2 Bst. a oder b DSG i.V.m. Art. 31 Abs. 1 DSG);
- es muss die Informationspflicht nach Art. 19 ff. DSG eingehalten sein;
- etwaige Auslandsübermittlungen haben gemäss Art. 16 ff. DSG zu folgen; und
- werden besonders schützenswerte Personendaten an Dritte gegeben, ist ebenfalls eine Rechtfertigung erforderlich (Art. 30 Abs. 2 Bst. c DSG i.V.m. Art. 31 Abs. 1 DSG).

[60] Soweit es sich wie um ein Bundesorgan handelt (zu denken ist hier insbesondere an die Forschungsanstalten und die Eidgenössischen Technischen Hochschulen wie etwa die ETH Zürich), gelten folgende Voraussetzungen:

- Es müssen die Bearbeitungsgrundsätze von Art. 6 DSG eingehalten sein, soweit das Gesetz keine Abweichung davon vorsieht;
- es muss die Informationspflicht nach Art. 19 DSG eingehalten sein;
- es ist eine gesetzliche Grundlage erforderlich (Art. 34 Abs. 1 DSG);
- diese Grundlage muss in einem Gesetz im formellen Sinn enthalten sein, soweit es um besonders schützenswerte Personendaten oder ein Profiling geht oder ein besonders schwerwiegender Eingriff droht (Art. 34 Abs. 2 DSG);
- etwaige Auslandsübermittlungen haben Art. 16 ff. DSG zu folgen.

[61] Ob diese Anforderungen erfüllt sind, muss je nach Quelle beurteilt werden. Es macht Sinn, folgende Differenzierungen und Annahmen zu treffen:

- Es kann zwischen **öffentlichen Quellen** und nicht öffentlichen Quellen unterschieden werden. Enthält eine öffentliche Quelle Personendaten, bedeutet dies nicht, dass die Daten mit Wissen und Willen der betroffenen Person öffentlich gemacht worden sind. In den meisten Fällen wird dies nicht der Fall sein.[87] Die Unterscheidung ist jedoch wichtig, um den

[87] Luca Dal Molin/Kirsten Wesiak-Schmid, Datenschutz im Unternehmen, Zürich/St. Gallen 2023, § 1 N 202 nennen als Beispiel veröffentlichte Fotos im Internet, auf denen Drittpersonen entgegen deren Wissen abgebildet sind, oder die Bekanntmachung von Personendaten entgegen dem Willen der betroffenen Personen in einer Zeitung.

Eingriff in die Persönlichkeit zu beurteilen, der mit einer Verwendung für das Training einhergeht. Dies ist wiederum für die Beurteilung der Verhältnismässigkeit, einer etwaigen Rechtfertigung und Notwendigkeit einer gesetzlichen Grundlage in einem formellen Gesetz wichtig. Dieser letzte Punkt ist insofern zu relativieren, als dass dort, wo die Publikation einen besonders gewichtigen Eingriff darstellt und die publizierten Daten nicht leicht zugänglich sind (z.B. Veröffentlichung von gestohlenen privaten Daten im Darknet), der Eingriff durch eine Verwendung dieser Daten entsprechend gewichtig sein wird. Umgekehrt werden die Hürden zur Verwendung von Daten dort, wo sie mit Wissen und Willen der betroffenen Person publiziert worden sind, datenschutzrechtlich weniger Probleme aufwerfen.[88] Im vorliegenden Fall gehen wir davon aus, dass ausschliesslich öffentliche Quellen für das Training verwendet werden. Das sind Quellen dann, wenn sie einer unbestimmten Anzahl von Personen zugänglich sind, selbst wenn hierfür ein Entgelt entrichtet oder ein Vertrag abgeschlossen werden muss.

- Eine **Memorisierung von Personendaten** dürfte nur sehr selten vorkommen und damit auch nur sehr wenige Personen betreffen. Handelt es sich bei den Quellen um öffentliche Quellen, sind erfahrungsgemäss nur Personen betroffen, über die in der Öffentlichkeit viel berichtet wird (oder worden ist) oder die unter ihrem Namen selbst viel publiziert haben (auch wenn sie ansonsten keine Personen der «Zeitgeschichte» sind). Wir bezeichnen sie als «öffentliche Personen». Ob memorisierte Personendaten beim späteren Gebrauch ausgegeben werden, hängt davon ab, ob dem Modell entsprechende Prompts gegeben werden.[89]
- Wir gehen davon aus, dass ein gewisses Interesse an einer Memorisierung von **Daten über öffentliche Personen** besteht, d.h. daran, dass ein Modell sich zu bekannten Personen der Zeitgeschichte äussern kann. Von einem besonderen Interesse an Äusserungen über Personen, die zwar keine Personen der Zeitgeschichte sind, aber sehr viel publiziert haben, gehen wir nicht aus, doch werden sie auch nicht als störend erachtet.
- Das Training eines grossen Sprachmodells dient an sich **keinen personenbezogenen Zwecken**. Zwar kann und wird es Personendaten öffentlicher Personen enthalten, die in den Trainingsdaten häufig vorkommen, aber das ist nicht das primäre Ziel. Dieses ist vielmehr die Erstellung eines Systems mit Sprach- und Sachwissen (Allgemeinwissen), das basierend auf einem Input einen passenden Output generieren kann.[90] Wo jedoch das Allgemeinwissen auch Wissen über bestimmte Personen der Zeitgeschichte mitumfasst, ist dessen Memorisierung ein sekundäres Ziel.

  In der Fachwelt wird neuerdings sogar noch einen Schritt weitergegangen und vertreten, dass die Bearbeitung von Daten für die Zwecke des Trainings eines Sprachmodells gar keine Bearbeitung von Personendaten sei, weil die Stelle, die es vornimmt, die betroffenen Personen aufgrund der Umstände möglicherweise gar nicht mit einem Aufwand, den sie zu

88 Corrado Rampini/Rehana C. Harasgama, Basler Kommentar Datenschutzgesetz/Öffentlichkeitsgesetz, 4. Aufl., Basel 2024, DSG 30 N 26.

89 Siehe dazu: Rosenthal, Teil 19 (Fn. 1).

90 Siehe dazu: David Rosenthal, Teil 17: Was in einem KI-Modell steckt und wie es funktioniert, abrufbar unter: https://www.vischer.com/know-how/blog/teil-17-was-in-einem-ki-modell-steckt-und-wie-es-funktioniert/.

treiben bereit ist, identifizieren kann.[91] Wir erörtern diese Theorie vorliegend einstweilen nicht; sie könnte jedoch auch im Schweizer Recht vertreten werden.

- Das Training ist ein für sich **geschlossener Bearbeitungsvorgang**. Es umfasst das Aufbereiten der Trainingsdaten, die Durchführung des Trainings (nicht überwacht oder überwacht) und die Validierung bzw. das Testen des Modells. Für unsere Zwecke gehen wir davon aus, dass die Trainingsdaten aufbewahrt werden, was jedoch für das Sprachmodell nicht relevant ist. Die Nutzung des Sprachmodells oder dessen weitere Bearbeitung gehört nicht zu diesem Bearbeitungsvorgang.
- Es werden für das Training grundsätzlich **keine rechtswidrigen oder unsittlichen Quellen** wie etwa Daten aus dem Darknet oder von Sites mit Hate-Speech verwendet. Es werden auch keine Daten von Quellen verwendet, die sich zum Zeitpunkt der Erhebung mittels anerkannter Standards gegen eine Verwendung ihrer Daten für Trainingszwecke (z.B. mittels Anti-Crawling-Kennzeichen) ausgesprochen haben.[92]
- Wir gehen davon aus, dass das resultierende grosse Sprachmodell der **Öffentlichkeit zugänglich** gemacht wird (also nicht bloss eigenen Zwecken dient und vertraulich bleibt). Das Quellenmaterial wird jedoch nicht öffentlich zugänglich gemacht.

[62] Nebst den materiellen Anforderungen des Datenschutzrechts ist in formeller Hinsicht für ein Training unseres Erachtens eine **Datenschutz-Folgenabschätzung** durchzuführen. Das Training eines Sprachmodells stellt zudem eine Bearbeitungsaktivität dar, die in einem **Verzeichnis der Bearbeitungstätigkeiten** aufzunehmen ist. Ob auch eine **Protokollierung** (namentlich der Speicherung und Löschung der Personendaten) und ein **Bearbeitungsreglement** erforderlich sind, hängt einerseits davon ab, ob ein Bundesorgan zugange ist, und andererseits, ob besonders schützenswerte Personendaten in grossem Umfange bearbeitet werden (was unwahrscheinlich ist).

B. Einhaltung der Bearbeitungsgrundsätze, Pflicht zur Information

[63] Die Bearbeitungsgrundsätze und die Pflicht zur Information werden wie folgt eingehalten:

- **Zweckbindung:** Die Erhebung von Trainingsdaten für das Training stellt in der Regel eine indirekte Beschaffung von Personendaten für einen neuen Zweck dar: Die Daten werden nicht bei der betroffenen Person selbst erhoben (Ausnahme: Die eigene Website einer privaten Person wird eingelesen) und für sie wird in der Regel nicht erkennbar gewesen sein, dass die Daten vom Verantwortlichen für ein Training benutzt werden sollen (Ausnahme: Die Daten stammen von einer Datenquelle, die dies den Personen kommuniziert hat, weil sie deren Daten für solche Zwecke weitergeben können wollte, z.B. eine Plattform für soziale Medien, welche Daten für KI-Trainings verkauft). Es stellt sich somit die Frage, ob das Training mit dem Zweck, zu welchem die Daten von der betroffenen Person ursprünglich beschafft worden sind, mindestens vereinbar ist. Die Verwendung von für einen Primärzweck beschafften Daten für einen sekundären, nicht personenbezogenen Zweck in pseudonymisierter oder anonymisierter Form gilt als ein typisches Beispiel für

91 Peter Craddock: Op-ed: AI training data = (non-)personal data? And is consent really relevant?, in: LinkedIn, 14. Oktober 2024, abrufbar unter: https://vischerlnk.com/3A27ZYm.

92 Siehe jedoch Kapitel VIII.

einen mit dem ursprünglichen Zweck «vereinbaren» Zweck.[93] Ein solcher Fall liegt hier in der Regel vor: Wird ein Sprachmodell mit Inhalten trainiert, die Personendaten enthalten, werden diese normalerweise nicht im Modell abgebildet, d.h. das Modell merkt sich nur aggregierte Informationen darüber, wie Sprache gebildet wird. Es erfolgt somit eine Anonymisierung. Ausnahme bildet der Fall, dass ein bestimmtes Personendatum (z.B. der Geburtstag einer bekannten Persönlichkeit) genügend oft in den Trainingsdaten vorkommt und dadurch Eingang in das Sprachwissen des Modells findet. In diesem Fall erfolgt eine Pseudonymisierung, weil die betroffenen Personen anhand des Sprachwissens nicht direkt identifiziert sind, sondern es eine passende Abfrage braucht, damit sie erkennbar werden (anhand der Parameter im Modell lässt sich die Information schon gar nicht erkennen, weshalb gewisse Stimmen davon ausgehen, dass ein Sprachmodell per se keine Personendaten enthalten kann[94]; diese – aus unserer Sicht falsche – Ansicht setzt sich aber nicht wirklich durch[95]).[96] In diesen Fällen kann es geschehen, dass die Vereinbarkeit mit dem Zweck in Frage gestellt wird. Dem lässt sich allerdings folgendes entgegenhalten: Ist das häufige Vorkommen dem Umstand geschuldet, dass die Trainingsdaten öffentliche Inhalte wiedergeben, dann wird es sich in der Regel um Personendaten öffentlicher Personen handeln, weil naturgemäss nur bei öffentlichen Personen Personendaten in grösserer Zahl und in unterschiedlichen öffentlichen Inhalten zu finden sind (mit oder gegen ihren Willen). Wenn dem aber so ist, dann wird eine solche Person inzwischen damit rechnen müssen, dass diese Informationen über sie auch für das Training von grossen Sprachmodellen verwendet wird, weil inzwischen allgemein bekannt ist, dass hierfür möglichst viele öffentliche Inhalte verwendet werden. Es kann jedenfalls mit guten Gründen vertreten werden, dass es bei solchen Personen nicht «unerwartet, unangebracht oder beanstandbar» ist (um die Worte aus der Botschaft des Bundesrats zu zitieren),[97] dass eben diese Personendaten nicht nur für zahlreiche Publikationen, sondern auch das Training eines grossen Sprachmodells verwendet werden. Somit ist der Grundsatz der Zweckbindung (Art. 6 Abs. 3 DSG) grundsätzlich sowohl bei memorisierten als auch bei nicht memorisierten Personendaten in Trainingsdaten erfüllt.

- **Verhältnismässigkeit:** Die Bearbeitung von Personendaten muss gemäss Art. 6 Abs. 2 DSG verhältnismässig sein, d.h. zum Zweck des Trainings geeignet, erforderlich und zumutbar (d.h. verhältnismässig im engeren Sinne).[98] Bei diesem Grundsatz wird normalerweise angeführt, dass das Training eines grossen Sprachmodells davon lebt, dass es möglichst viele Trainingsdaten beinhaltet, d.h. die Qualität mit zunehmendem Datenvolumen auch besser

93 David Rosenthal, Datenschutz beim Einsatz generativer künstlicher Intelligenz, in: Jusletter vom 6. November 2023, N 28.

94 Hamburgischer Beauftragter für Datenschutz und Informationsfreiheit, Diskussionspapier: Large Language Models und personenbezogene Daten, 15. Juli 2024, abrufbar unter: https://datenschutz-hamburg.de/news/hamburger-thesen-zum-personenbezug-in-large-language-models.

95 Dazu inzwischen: European Data Protection Board, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, 17. Dezember 2024, abrufbar unter: https://www.edpb.europa.eu/system/files/2024-12/edpb_opinion_202428_ai-models_en.pdf.

96 David Rosenthal, Wie und warum ein grosses Sprachmodell den «Geburtstag» von (öffentlichen) Personen kennen kann, Illustration vom 4. August 2024, abrufbar unter: https://www.rosenthal.ch/downloads/VISCHER_LLM_Geburtstag_Beispiel.pdf.

97 BBl 2017, S. 7025.

98 Lukas Bühlmann/Michael Reinle, Basler Kommentar Datenschutzgesetz/Öffentlichkeitsgesetz, 4. Aufl., Basel 2024, DSG 6 N 51 ff.; m.w.Verw.: Dal Molin/Wesiak-Schmid (Fn. 87), § 1 N 133.

wird.[99] Dies ist insofern etwas an der Sache vorbeigedacht, als es nicht auf ein möglichst grosses Volumen an Personendaten ankommt, sondern an Trainingsdaten. Das Training mit Personendaten ist gerade nicht das Ziel. Sie lassen sich jedoch in den Trainingsdaten oft nicht vermeiden, etwa wenn öffentliche Artikel, Aufsätze, Websites oder Einträge aus Wikipedia verwendet werden. Geeignet sind sie für das Training freilich trotzdem: Wenn dem Modell vermittelt werden soll, wie Sätze in einer bestimmten Sprache formuliert werden müssen, damit sie stimmen, müssen darin auch alle Elemente vorkommen, welche die Sprache mit sich bringt – und dies sind nun einmal auch Informationen, die aus den Sätzen Personendaten machen, wie beispielsweise Namen oder andere identifizierende Elemente. Insofern sind Personendaten auch nötig für das Training. Nötig sind allerdings nicht wahre Informationen, sondern nur beispielhafte Formulierungen. Es genügen Trainingssätze wie «Peter Muster wurde am 1. Januar 1980 geboren» um das Konzept des Satzes zu vermitteln. Der Inhalt muss grundsätzlich nicht zutreffen. Eine Ausnahme besteht dort, wo mit einer höheren Zahl an Wiederholungen zu rechnen ist, also bei öffentlichen Personen. Hier besteht durchaus das Interesse, dass das Sprachmodell nicht nur Sätze betreffend Geburtstage richtig formulieren kann, sondern wenn es um den Geburtstag einer prominenten Person geht, die Angaben möglichst auch richtig liefert. Daraus ergibt sich, dass das Modell Personendaten für das Training nur dort brauchen kann, wo es um öffentliche Personen geht. In den anderen Fällen sind sie nicht nötig und könnten mit Pseudonymen ersetzt oder geschwärzt werden, ohne dem Training zu schaden (dies ist insofern einzuschränken, als dass ein Modell natürlich auch lernen muss, wie z.B. eine Telefonnummer im Sprachkontext eingesetzt wird und es daher auch Sätze braucht, wo dies geschieht; die Telefonnummer muss jedoch nicht bis auf die letzte Nummer stimmen).[100] Der dritte Aspekt der Verhältnismässigkeit, die Zumutbarkeit, bereitet hingegen weniger Schwierigkeiten: Für betroffene Personen, deren Personendaten nicht memorisiert werden, ist der Eingriff in ihre Persönlichkeit nur sehr beschränkt. Es geht nicht um sie, und ihre Personendaten sind auch nicht im Modell. Demgegenüber steht das Interesse desjenigen, der das Sprachmodell trainiert. Sein Interesse ist es, möglichst viele Trainingsdaten zu verwenden. Dabei verfügt er nur über begrenzte Möglichkeiten, diese vorgängig mit vertretbarem Aufwand von unnötigen Personendaten zu säubern, soweit solche überhaupt erkannt werden können. Wird jedoch genau dies getan, ist unseres Erachtens nicht nur die Zumutbarkeit gegeben, sondern auch die Notwendigkeit der Bearbeitung jener Personendaten, die in den gesäuberten Trainingsdaten verbleiben. Denn der Zweck des Trainings kann nur dann erfüllt werden, wenn sich das Training mit vernünftigem Aufwand durchführen lässt. Müssten die Trainingsdaten einzeln auf etwaige Personendaten hin überprüft werden, wäre ein Training von vorneherein nicht mehr möglich und die Zweckerreichung damit vereitelt. In diesem Sinne ist somit auch der Grundsatz der Verhältnismässigkeit eingehalten.

- **Treu und Glauben, einschliesslich Transparenz:** Die Bearbeitung muss weiter nach Treu und Glauben erfolgen (Art. 6 Abs. 2 DSG). Dies bedeutet über die bereits erwähnten Kriterien der Zumutbarkeit und der Vereinbarkeit hinaus, dass die Bearbeitung in einer nach-

99 Steffen Albrecht, ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen, TAB-Hintergrundpapier Nr. 26, 2023, abrufbar unter: https://www.bundestag.de/resource/blob/944148/30b0896f6e49908155fcd01d77f57922/20-18-109-Hintergrundpapier-data.pdf.

100 Zum Ganzen: Rosenthal, Teil 19 (Fn. 1).

vollziehbaren, transparenten Weise erfolgen muss.[101] Es bedeutet, dass sie allgemein nicht als stossend, unfair oder störend empfunden werden darf.[102] Von letzterem kann unserer Ansicht nach im Falle des Trainings eines grossen Sprachmodells nicht ausgegangen werden (in der Annahme, dass das Sprachmodell berechtigten privaten oder öffentlichen Interessen dient und nicht beispielsweise illegalen Zwecken). Wenn bereits das Urheberrecht – wie gezeigt – die Analyse von fremden Werken gestatten will, um daraus Erkenntnisse zu ziehen (hier: Sprachwissen), belegt dies, dass unsere Rechtsordnung solche Vorgänge grundsätzlich zulassen, ja sogar möglich machen will, selbst wenn die Vorgänge in fremde Rechtssphären eingreifen – jedenfalls solange diese nicht erheblich beeinträchtigt werden. So verhält es sich auch hier, jedenfalls vorausgesetzt, es wird kein personenbezogener Zweck verfolgt und es kommt grundsätzlich nicht zu einer Memorisierung. Die genannten Ausnahmen von Personen, deren Daten bereits häufig in der Öffentlichkeit vorkommen, stehen dem unseres Erachtens wie erwähnt nicht entgegen, sondern werden normalerweise durch ein berechtigtes öffentliches oder auch privates Interesse gedeckt sein. Bei Daten aus Data Breaches oder bei sonst rechtswidrigen Datenquellen wird dies zwar nicht unbedingt gelten, aber solche kommen nach den genannten Prämissen nicht zum Einsatz. Gleiches gilt für Daten von Websites, die sich gegen Crawler für die Zwecke von KI-Trainings ausgesprochen haben. Unter diesen Umständen erscheint das Training eines grossen Sprachmodells auch über den Grundsatz der Verhältnismässigkeit hinaus weder als unfair noch als stossend oder störend. Etwas schwieriger gestaltet es sich mit der Erkennbarkeit der Datenbearbeitung, d.h. deren Transparenz. Die Transparenz soll es den betroffenen Personen ermöglichen, sich gegen die Verwendung ihrer Daten auszusprechen.[103] Da die Daten nicht von den betroffenen Personen direkt erhoben werden, ist es jedenfalls in den meisten Fällen nicht möglich, diese direkt zu informieren. Es bleibt dem Verantwortlichen somit lediglich die Möglichkeit, über seine Beschaffung und weitere Bearbeitung von Personendaten allgemein zu informieren, so beispielsweise über entsprechende Ankündigungen oder seine Datenschutzerklärung. Dieses Vorgehen hat insofern das Manko, als dass die betroffenen Personen gar keinen Anlass haben, sich auf den betreffenden Seiten nach entsprechenden Informationen zu erkundigen. Allerdings verlangt der Grundsatz von Treu und Glauben nach unserer Ansicht auch keine absolute Transparenz, sondern eine Transparenz, wie sie fairerweise, d.h. angesichts der Umstände, angezeigt erscheint. Diese Umstände beinhalten auch die Schwere des Eingriffs, welche die Bearbeitung der Daten zur Folge hat. Sie ist, wie gezeigt, nicht gewichtig, weil die Bearbeitung grundsätzlich weder personenbezogen erfolgt noch zur Speicherung von Personendaten führt. Personendaten sind, mit anderen Worten, grundsätzlich blosses Beiwerk. In vergleichbaren Fällen sieht das Schweizer Recht normalerweise keinen besonderen Schutzbedarf, wie etwa beim Recht am eigenen Bild, das ebenfalls nicht greift, wenn Personen auf einer Aufnahme lediglich Beiwerk sind.[104] So verhält es sich letztlich auch mit der Informationspflicht nach Art. 19 DSG. Sie besteht zwar grundsätzlich auch im vorliegenden Fall, entfällt jedoch, wenn die Information nicht mög-

101 Bühlmann/Reinle (Fn. 98), DSG 6 N 50.

102 M.w.H.: Bühlmann/Reinle (Fn. 98), DSG 6 N 49.

103 Siehe: David Rosenthal, Teil 16: Wie Unternehmen beim Einsatz von KI Transparenz gewährleisten können, abrufbar unter: https://www.vischer.com/know-how/blog/teil-16-wie-unternehmen-beim-einsatz-von-ki-transparenz-gewaehrleisten-koennen/.

104 Andreas Meili, Basler Kommentar Zivilgesetzbuch I, 7. Aufl., Basel 2022, ZGB 28 N 20.

lich oder mit einem unverhältnismässigen Aufwand verbunden wäre (Art. 20 Abs. 2 DSG). Das ist hier gegeben. Trotzdem empfehlen wir denjenigen, die ein grosses Sprachmodell trainieren wollen, dies in ihrer Datenschutzerklärung entsprechend auszuführen. In diesem Fall sind nach unserer Ansicht sowohl der Grundsatz der Bearbeitung nach Treu und Glauben wie auch die Informationspflicht erfüllt.

- **Speicherbegrenzung:** Personendaten dürfen nur so lange aufbewahrt werden, als dies für die Erfüllung des Zwecks erforderlich ist (Art. 6 Abs. 4 DSG). Hier ist nach Personendaten im Modell und solche in den Trainingsdaten zu unterscheiden. Soweit eine Memorisierung erfolgt ist, teilen die Personendaten naturgemäss das Schicksal des grossen Sprachmodells: Solange es gebraucht wird, braucht es auch die darin allenfalls enthaltenen Personendaten. Dies wird in der Praxis kaum Schwierigkeiten bereiten. Schon eher stellt sich die Frage nach der Aufbewahrung der Trainingsdaten, und zwar über das Training hinaus. Die Frage, ob Trainingsdaten aufbewahrt werden sollen, muss allerdings in einem grösseren Kontext betrachtet werden. Sie können nämlich nicht nur dem eigentlichen Training dienen, sondern auch weiteren Zwecken, wie etwa dem Nachweis der Quellen und Trainingsdaten, die tatsächlich verwendet wurden (beispielsweise für den Fall von Rechtsstreitigkeiten, aber auch aus Gründen der Transparenz). Sie können relevant sein für die wissenschaftliche Forschung, etwa um Aspekte der Memorisierung zu untersuchen. Sie können auch für Re-Trainings erforderlich werden. Jeder dieser Gründe kann hinreichend sein, um Trainingsdaten nicht sofort nach dem Training zu löschen. Der Grundsatz der Speicherbegrenzung (ein Teilaspekt der Verhältnismässigkeit) wird somit in der Regel nicht verletzt sein, solange es noch einen guten Grund gibt, die Trainingsdaten aufzubewahren.[105] Der Zugriff auf diese muss allerdings entsprechend eingeschränkt sein.
- **Richtigkeit:** Wer Personendaten bearbeitet, muss sich über deren Richtigkeit vergewissern, wobei die Richtigkeit sich nach dem Zweck bestimmt, für welchen die Personendaten bearbeitet werden (Art. 6 Abs. 5 DSG).[106] Das Erfordernis erscheint auf den ersten Blick unlösbar zu sein, da ein Training Trainingsdaten in einer Menge erfordert, die inhaltlich nicht mehr im Einzelnen geprüft werden kann. Schon nur bei Internet-Crawler-Daten muss mit zahlreichen Informationen von Websites gerechnet werden, welche sachlich falsch sind. Dies ist aus unserer Sicht hier jedoch nicht relevant, zumal der Zweck des Trainings eines grossen Sprachmodells nicht dem Aufbau von Sachwissen aus einzelnen Inhalten dient. Vielmehr geht es um den Aufbau einerseits von Sprachwissen, d.h. das Modell soll erfassen, wie Sprache in den Trainingsdaten verwendet wird, und andererseits von Allgemein- und Fachwissen, d.h. das Modell soll erfassen, was an Informationen genügend häufig vorkommt, um es zu solchem werden zu lassen. Das Training ist der Bearbeitungsvorgang, mit welchem diese Zwecke erreicht werden – und hinsichtlich dieser Zwecke sind die Personendaten in den Trainingsdaten rechtlich gesehen dann richtig, wenn sie die Trainingsinhalte diesbezüglich unverfälscht wiedergegeben. Zweck ist nämlich nicht die Schaffung von irgendwelchem übergeordneten Wissen, sondern dem *in den jeweiligen Trainingsinhalten* enthaltenen Wissen. Betrifft das Sachwissen auch öffentliche Personen, kann es bei im Training hinreichend häufig gesehenen Informationen auch deren Personendaten enthalten. In Be-

105 In diesem Sinne auch: Stephanie Volz, KI Sandboxen für die Schweiz?, in: SZW 2022, S. 51 ff., S. 56.

106 Bühlmann/Reinle (Fn. 98), DSG 6 N 247; Nadja Braun Binder, Künstliche Intelligenz und automatisierte Entscheidungen in der öffentlichen Verwaltung, in: SJZ 115/2019, S. 467 ff., S. 474.

zug auf diesen Zweck verlangt der Grundsatz der Richtigkeit in der Tat, dass angemessene Massnahmen zur Gewährleistung der Richtigkeit vorgenommen werden.[107] Auch dies bereitet aber normalerweise keine Schwierigkeiten: Modelle nehmen, wie bereits mehrfach erwähnt, nicht jedes Personendatum als solches auf, sondern nur jene Informationen, die in grösserer Zahl gleich vorkommen, mit anderen Worten also durch unterschiedlichste Quellen «bestätigt» sind. Das Geburtsdatum einer prominenten Person, das in zahlreichen Publikationen genannt wird, ist wie bereits erwähnt ein solches Beispiel. Da grundsätzlich nur in solchen Fällen eine Memorisierung stattfindet, erfüllt sie automatisch auch das Erfordernis eines Faktenchecks. Erforderlich ist immerhin, dass für das Training eines grossen Sprachmodells eine möglichst grosse Vielfalt an Trainingsdaten verwendet wird und qualitativ gute Quellen bevorzugt werden. In diesem Sinne dient also ein Mehr an Personendaten auch dem Datenschutz. Die Richtigkeit der memorisierten Personendaten ist im Übrigen von der Richtigkeit der Personendaten im Output eines Sprachmodells zu unterscheiden; dieses kann auch Personendaten enthalten, die nicht als solche memorisiert worden sind, sondern eine Halluzination[108] darstellen (oder aus dem Input des Benutzers stammen). Da es sich bei der Verwendung eines LLM um eine separate Bearbeitung handelt, tangiert dies das Training nicht. Der Grundsatz der Richtigkeit ist somit unter diesen Voraussetzungen erfüllt.

- **Rechtmässigkeit:** Schliesslich dürfen Personendaten nicht unrechtmässig beschafft oder sonst bearbeitet werden (Art. 6 Abs. 1 DSG). Diese Vorgabe bezieht sich auf Rechtsverletzungen ausserhalb des DSG, wie sie beispielsweise vorkommen können, wenn der Verrat eines Geschäftsgeheimnisses ausgenutzt wird (Art. 162 StGB).[109] Es könnte vertreten werden, dass auch die Verwendung von Personendaten, die im Darknet als Beute aus einem Hacker-Angriff publiziert worden sind, eine solche unrechtmässige Bearbeitung darstellt. Aus diesem Grund wird für die Zwecke des Trainings eines grossen Sprachmodells darauf zu achten sein, dass keine Daten aus illegalen oder sittenwidrigen Quellen verwendet werden, was jedoch ohnehin den Gepflogenheiten entspricht. Der Umstand, dass vereinzelt auch auf «normalen» Websites verratene Geschäftsgeheimnisse oder illegale Inhalte enthalten sein können, sollte vorliegend kaum Schwierigkeiten bereiten: Erstens meint «unrechtmässig» vorliegend nur Bestimmungen, die dem Schutz der Persönlichkeit der betroffenen Personen dienen (und nicht anderen «Opfern», wie dem Unternehmen, dessen Geschäftsgeheimnisse tangiert sind), und zweitens sind jedenfalls einzelne Personendaten wie gezeigt lediglich «Beiwerk», d.h. es geht nicht um sie und sie fallen dem Verantwortlichen nur zufällig zu. Sie sollten unseres Erachtens unter den obigen Prämissen nicht dazu führen, dass eine Datenbearbeitung als unrechtmässig gilt. Auch der Grundsatz der Rechtmässigkeit wird somit eingehalten sein.

---

107 Bruno Baeriswyl, Stämpflis Handkommentar Datenschutzgesetz, 2. Aufl., Bern 2023, DSG 6 N 62 ff.; Reto Fanger, Orell Füssli Kommentar Datenschutzgesetz, 2. Aufl., Zürich 2023, DSG 6 N 11.

108 Sie sind kein Problem der Richtigkeit des Trainings oder der Personendaten im Modell, denn gemeint ist damit in der Regel nicht der Umstand, dass ein Modell falsche Informationen enthält (das ist bei genügend zahlreichen «falschen» Trainingsdaten natürlich möglich), sondern dass es zu einem bestimmten Thema gar keine oder keine ausreichend klaren Informationen enthält und es daher solche zur Lückenfüllung erfindet, weil es dem Benutzer einen lückenlosen Text liefern will. Dies ist daher nicht ein datenschutzrechtliches Problem des Modells, sondern seiner Benutzung.

109 David Rosenthal/Yvonne Jöhri, Handkommentar zum Datenschutzgesetz, Zürich 2008, aDSG 6 N 3.

- **Datensicherheit:** Es muss durch geeignete technische und organisatorische Massnahmen eine dem Risiko angemessene Datensicherheit gewährleistet sein (Art. 8 DSG). Sicherheit versteht sich als Wahrung der Vertraulichkeit, der Integrität und der Verfügbarkeit der Daten und der Nachvollziehbarkeit der Datenbearbeitung.[110] Im vorliegenden Fall bedeutet dies, dass für die Speicherung der Trainingsdaten und die Trainingsvorgänge selbst sichergestellt ist, dass die genannten Schutzziele nicht verletzt werden. Im Falle von Sprachmodellen ist besonders auf technikspezifische Angriffsmethoden zu achten, wie beispielsweise das Risiko des «Vergiftens» eines Sprachmodells durch Trainingsdaten, die mit nicht authentischen Informationen versetzt sind, um sie auf diese Weise ins Sprachmodell einzuführen. So wäre es denkbar, bestimmtes gefälschtes Faktenwissen über eine Person durch genügend häufige Wiederholungen in einem Trainingsdatensatz einer Memorisierung zuzuführen. Das gefälschte Faktenwissen würde daraufhin im Modell seinen Niederschlag finden und bei passenden Prompts abgerufen. Dem kann einerseits teilweise entgegengewirkt werden, indem Trainingsdatensätze selbst gebildet werden, und andererseits durch eine Verwendung von als zuverlässig bekannten, bereits bestehenden Datensätzen. Die Massnahmen müssen keine perfekte Datensicherheit gewährleisten, sondern eine dem Risiko angemessene. Wir gehen aufgrund bisheriger Erfahrungen davon aus, dass sich damit auch der Grundsatz der Datensicherheit einhalten lässt.
- **Verhinderung von Verzerrungen bzw. Diskriminierung:** In den obigen Grundsätzen nicht erwähnt ist die Verhinderung von Verzerrungen (Bias) und Diskriminierung, die im Zusammenhang mit datenschutzrechtlichen Anforderungen auch an grosse Sprachmodelle häufig ins Feld geführt wird.[111] Diese Anforderungen betreffen die Verwendung der Sprachmodelle bzw. deren Tauglichkeit für bestimmte Anwendungen, nicht aber das Training als Vorgang der Bearbeitung von Personendaten. Soweit keine Memorisierung von Personendaten erfolgt, sind die betroffenen Personen von etwaigen Verzerrungen im Modell jedenfalls in Bezug auf ihre Personendaten, die für das Training verwendet werden, nicht betroffen. Betroffen sind allfällige andere Personen, mit deren Personendaten das Modell zum Einsatz kommt. Betroffen sind allenfalls auch jene Personen, deren Personendaten memorisiert werden. Das Problem ist insofern abhängig von den für das Sprachmodell verwendeten Datenquellen. Jede einzelne Quelle für sich stellt diesbezüglich datenschutzrechtlich zwar kein Problem dar. Wird jedoch im Hinblick auf den Zweck des Sprachmodells keine ausgewogene Auswahl an Datenquellen getroffen, kann dies in Bezug auf diese Personen zu einer Memorisierung von einseitigen Personendaten und somit besagten Verzerrungen im Modellwissen führen, was wiederum sowohl den Grundsatz der Richtigkeit als auch jenen von Treu und Glauben und der Verhältnismässigkeit tangieren kann, soweit ein Anwendungsfall den Abruf entsprechender Daten vorsieht. Auf diesen Umstand ist daher zu achten, auch wenn er die datenschutzrechtliche Zulässigkeit des Trainings nur insofern betreffen kann, als der genannte Verwendungszweck bereits zu diesem Zeitpunkt feststeht.

[64] Die vorstehenden Ausführungen zeigen, dass das Training eines grossen Sprachmodells unter den genannten Prämissen grundsätzlich nicht zu einer Verletzung der Bearbeitungsgrundsät-

110 Christa Stamm-Pfister, Basler Kommentar Datenschutzgesetz/Öffentlichkeitsgesetz, 4. Aufl., Basel 2024, DSG 8 N 2 f.; vgl. BBl 2017, S. 7031; Clara-Ann Gordon/Luisa Egli, Orell Füssli Kommentar Datenschutzgesetz, 2. Aufl., Zürich 2023, DSG 8 N 1.

111 Siehe zur Gefahr der Diskriminierung: Braun Binder (Fn. 106), S. 473 ff.

ze von Art. 6und 8 DSG führt. Daher ist für **private Verantwortliche** eine Rechtfertigung nach Art. 31 DSG grundsätzlich nicht erforderlich (siehe aber sogleich).

C. Rechtfertigung und Rechtfertigungsgründe

1. Private Verantwortliche

[65] Eine Rechtfertigung wäre für private Verantwortliche allerdings dann notwendig, wenn das Training zu einer Memorisierung von besonders schützenswerten Personendaten führen würde und vernünftigerweise damit zu rechnen wäre, dass letztere im Falle einer Verwendung oder Weitergabe des Sprachmodells abgerufen würden (Art. 30 Abs. 2 Bst. c DSG).[112] Ein Beispiel wäre, dass im Modell die politische Gesinnung einer bekannten Persönlichkeit memorisiert würde (so wissen die grossen Sprachmodelle beispielsweise, welche Ansichten Donald Trump vertritt). Eine Rechtfertigung kann auch in jenen Fällen erforderlich werden, in denen der Bearbeitung für die Zwecke des Trainings widersprochen wird (Art. 30 Abs. 2 Bst. b DSG).[113] Dies ist nicht auszuschliessen und insbesondere dann der Fall, wenn auf Websites oder in anderen Quellen, die von der betroffenen Person selbst stammen und Personendaten enthalten, ein entsprechender Widerspruch vermerkt ist.[114] Ähnlich gelagert ist auch der Fall, in welchem Personendaten aus einer Quelle gesammelt werden, die hinsichtlich ihrer Zweckbestimmung eine Verwendung der Personendaten für die Zwecke des Trainings ausgeschlossen hat.

[66] Als Rechtfertigungsgründe kommen in Frage:

- **Bearbeitung für nicht personenbezogene Zwecke** (Art. 31 Abs. 2 Bst. e DSG): Dieser Rechtfertigungsgrund legt ein überwiegendes privates Interesse am Training eines grossen Sprachmodells soweit nahe, als das Training nicht der Memorisierung von Personendaten dient. Diese Voraussetzung wird – wie schon mehrfach erwähnt – üblicherweise erfüllt sein, weil ein Training dem Aufbau von Sprach- und Allgemeinwissen dient, nicht der Aufnahme einzelner Personendaten. Ausnahmen bestehen in Bezug auf Personen des öffentlichen Lebens, deren Personendaten memorisiert werden, aber in Bezug auf diese greift dieser Rechtfertigungsgrund ohnehin nicht. Die Bearbeitung ihrer Daten muss daher anders gerechtfertigt werden. Auf den Rechtfertigungsgrund kann sich nur berufen, wer die drei aufgeführten Vorgaben erfüllt: (1) Die Daten müssen anonymisiert werden, sobald der Bearbeitungszweck dies zulässt.[115] Dies ist in Bezug auf das Training kein Problem, da die Aggregierung der Informationen aus den Trainingsinhalten, zu welcher der Trainingsprozess naturgemäss führt, auch die nötige Anonymisierung zur Folge hat. Die Aufbewahrung der Trainingsdaten hingegen dient weiteren Zwecken, die separat gerechtfertigt werden müssen. (2) Soweit möglich, dürfen besonders schützenswerte Personendaten Dritten nur in anonymisierter Form zugänglich gemacht werden.[116] Diese Vorgabe wird in der Praxis in der Regel ebenfalls keine relevante Hürde darstellen. (3) Die Ergebnisse werden so veröffentlicht, dass die

112 Rampini/Harasgama (Fn. 88), DSG 30 N 20; Thomas Steiner/Christian Laux, in: Adrian Bieri/Julian Powell (Hrsg.), Kommentar zum Schweizerischen Datenschutzgesetz mit weiteren Erlassen, Zürich 2023, DSG 30 N 23 ff.

113 David Rosenthal, Das neue Datenschutzgesetz, in: Jusletter vom 16. November 2020, N 38 f.

114 M.w.H.: Steiner/Laux (Fn. 112), DSG 30 N 14 f.

115 Rampini/Harasgama (Fn. 88), DSG 31 N 63; BBl 2017, S. 7076: Eine faktische Anonymisierung genügt.

116 Rosenthal (Fn. 113), N 42; Steiner/Laux (Fn. 112), DSG 31 N 33.

betroffenen Personen nicht bestimmbar sind.[117] Diese Vorgabe ist naturgemäss normalerweise ebenfalls erfüllt, unter Vorbehalt der Fälle, in denen eine Memorisierung stattfindet. Jedoch genügt eine Memorisierung alleine für die Bestimmbarkeit noch nicht: Bestimmbar werden die betroffenen Personen nur (aber immerhin) dann, wenn vernünftigerweise damit gerechnet werden muss, dass ein entsprechender Prompt eingegeben wird, der zum Abruf der betreffenden Information führt.[118] Der Rechtfertigungsgrund der Bearbeitung zu nicht personenbezogenen Zwecken greift somit grundsätzlich. Ausgenommen sind jene Fälle, in denen Personendaten memorisiert werden und mit deren Abruf gerechnet werden muss.

- **Personen des öffentlichen Lebens** (Art. 31 Abs. 2 Bst. f DSG): Weiter besteht ein überwiegendes privates Interesse in der Regel daran, Personendaten über eine Person des öffentlichen Lebens zu sammeln, soweit sich die Personendaten auf das Wirken dieser Person in der Öffentlichkeit beziehen. Dieser Rechtfertigungsgrund dürfte viele der in den Trainingsdaten enthaltenen Personendaten abdecken, die von jenen Personen stammen, die ein höheres Potenzial der Memorisierung aufweisen. Der Rechtfertigungsgrund erfasst zwar nur das Sammeln solcher Daten, doch dieses umfasst gemäss Lehre auch damit verbundene Tätigkeiten wie das Erfassen, Auftrennen und Katalogisieren.[119] Es wäre somit denkbar, auch das Training als eine Art «Katalogisierung» zu betrachten, zumal es nicht a priori einen Unterschied macht, ob die gesammelten Daten in einer konventionellen Datenbank oder einem Sprachmodell «abgelegt» werden. Jedoch würde die Weitergabe des Sprachmodells nicht mehr als «Sammeln» gelten, sofern die fraglichen Personendaten darin memorisiert wären und mit ihrem Abruf zu rechnen wäre. Diese Fälle, zu denen es kommen kann, deckt dieser Rechtfertigungsgrund somit nicht mehr ab.
- **Überwiegendes privates und öffentliches Interesse** (Art. 31 Abs. 1 DSG): Auch wenn keiner der in Art. 31 Abs. 2 DSG aufgezählten Rechtfertigungsgründe greift, kann trotzdem ein überwiegendes privates oder öffentliches Interesse an einer Datenbearbeitung bestehen.[120]

  Im vorliegenden Zusammenhang besteht dieses Interesse darin, ein grosses Sprachmodell bauen zu können. Dazu ist eine möglichst grosse Menge an Trainingsdaten erforderlich, die wiedergeben, wie Sprache verwendet wird. Weiter besteht zur Erhöhung der Modellqualität ein Interesse daran, dass diese Trainingsdaten möglichst divers sind, was die Erschliessung möglichst vieler unterschiedlicher Quellen erfordert. Das Training eines Sprachmodells ist wiederum nur möglich, wenn es mit vernünftigem Aufwand betrieben werden kann. Der Aufwand für eine zuverlässige Anonymisierung aller Trainingsdaten würde eine unüberwindbare Hürde darstellen. Bei den Textmengen, die verarbeitet werden, können nur hochautomatisierte Prozesse verwendet werden, die auch rechnerisch keinen grossen Aufwand darstellen, da sonst das Training bereits an der Datenbereinigung scheitert. Wird zu umfassend anonymisiert, verlieren die Trainingsdaten auch Informationen, die für korrekte und ausgeglichene Trainingsergebnisse wichtig sind. Wird das Modell der Öffentlichkeit zur Verfügung gestellt, hat auch diese ein Interesse daran, dass das Modell möglichst gut trainiert worden ist. Schliesslich hat derjenige, der ein Modell trainiert, ein Interesse, den

117 Monika Pfaffinger, in: Bruno Baeriswyl/Kurt Pärli/Dominika Blonski(Hrsg.), Datenschutzgesetz (DSG), 2. Aufl., Bern 2023, DSG 31 N 87.

118 Zum Ganzen: Rosenthal, Teil 19 (Fn. 1).

119 Rosenthal/Jöhri (Fn. 109), aDSG 13 N 70.

120 Rampini/Harasgama (Fn. 88), DSG 31 N 27 und N 69.

Aufwand so gering wie möglich zu halten. Soweit es um Personendaten von Personen des öffentlichen Lebens geht, kann je nach Verwendungszweck eines Sprachmodells auch ein Interesse des Anbieters und seiner Nutzer und letztlich auch ein öffentliches und privates Interesse daran bestehen, dass es diese Personen kennt und beispielsweise über solche Personen Aussagen treffen kann. Diese Aussagen sollen zudem möglichst richtig sein. Dies wiederum setzt voraus, dass möglichst viele Daten zu einer Person bearbeitet werden, weil dies erstens die Wahrscheinlichkeit von Halluzinationen reduziert (d.h. dass das Modell eine erfragte bzw. benötigte Information nicht hat, sie aber als notwendig erachtet und daher erfindet) und weil durch die Aggregierung der gelesenen Informationen, die im Rahmen eines Trainings eines Sprachmodells naturgemäss erfolgt, das Gewicht und damit der Einfluss einzelner Informationen und damit auch Fehlinformationen in der Gesamtheit einer bestimmten Information abnimmt. Enthalten einige wenige Artikel über eine Person eine bestimmte Falschinformation, so wird dieser im Sprachmodell dann weniger Gewicht beigemessen oder sie gar «ignoriert», wenn es entsprechend viel mehr andere Artikel gibt, die diese Falschinformation mit korrekter Information widersprechen.

Seitens der betroffenen Personen ist zu differenzieren zwischen jenen, deren Daten zwar für das Training verwendet, deren Daten aber nicht memorisiert werden, und solchen, deren Daten so häufig vorkommen, dass sie als Personendaten Eingang ins Modell finden:

Im ersten Fall, dem Normalfall, besteht zwar ein gewisser Kontrollverlust darüber, was mit den eigenen Daten geschieht und für welchen (fremden) Zweck sie verwendet werden. Die Nutzung erfolgt aber weder personenbezogen noch zeitigt sie eine nachhaltige Wirkung: Die Personendaten werden zwar vom Modell gelesen, finden aber als solche keinen Eingang darin; sie tragen lediglich zur Bildung des Sprachwissens des Modells bei, überleben diesen Vorgang nicht, d.h. werden weder wortwörtlich noch sinngemäss memorisiert. Handelt es sich um öffentliche Daten, ist der Eingriff in die Persönlichkeit der betroffenen Person sogar geringer. Denn ist eine Information öffentlich, muss damit gerechnet werden, dass dieselben Daten von einem Menschen konsumiert werden und bei diesem auch als einzelne Information im Gedächtnis hängen bleiben, was bei einem Sprachmodell nicht der Fall sein wird. Eine Nutzung solcher Inhalte für das Training eines Sprachmodells erfolgt wesentlich kontrollierter und beschränkter, ist also mithin weniger gravierend. Ferner wirken etwaige Massnahmen dem Kontrollverlust entgegen, wie z.B. die automatisierte Schwärzung von bestimmten identifizierenden Angaben wie bspw. Telefonnummer, Kreditkartennummern, Sozialversicherungsnummern, Adressen und Namen. Der vernünftigerweise mögliche Aufwand verhindert allerdings, dass diese Massnahmen perfekt sind. Vielmehr geht es darum, insgesamt die Verwendung von Personendaten einzudämmen. Am wirksamsten bleibt die – systembedingte – Aggregierung der Personendaten durch den Trainingsvorgang und damit verbundene Anonymisierung. In diesem Normalfall des Trainings ohne Memorisierung überwiegt aus unserer Sicht daher das private und öffentliche Interesse am Training.

Im zweiten Fall, dem Ausnahmefall, wo eine inhaltliche Memorisierung stattfindet, ist zu unterscheiden, ob die Memorisierung darauf zurückzuführen ist, dass es sich um eine Person des öffentlichen Lebens bzw. der Zeitgeschichte handelt. Ist dem so, wird üblicherweise ein öffentliches Interesse an der Memorisierung bestehen. Besondere Massnahmen, um dies sicherzustellen, werden über eine sorgfältige Auswahl der Trainingsdatensätze normalerweise nicht nötig sein. Es wird nämlich nur dann zu einer inhaltlichen Memorisierung kommen, wenn der betreffende Inhalt in sehr vielen Quellen in gleicher Form vorkommt,

was wiederum grundsätzlich nur dann der Fall sein wird, wenn tatsächlich ein öffentliches Interesse daran besteht.

Zwei Ausnahmen sind immerhin denkbar:

- Eine bestimmte, personenbezogene Information kommt deshalb so oft vor, weil die betroffene Person sie selbst häufig publik gemacht hat (z.B. weil für das Training Inhalte aus sozialen Medien verwendet worden sind, in welchen die Person sehr aktiv war). In diesem Fall kann gestützt auf Art. 30 Abs. 3 DSG vermutet werden, dass keine Verletzung der Persönlichkeit vorliegt, weil die Information mit Wissen und Willen der betroffenen Person allgemein zugänglich gemacht wurde[121] und die Nutzung durch Suchroboter (auch solche, die Inhalte für KI-Trainings bereitstellen) im Rahmen des Erwartbaren liegt. Eine Rechtfertigung bzw. Interessenabwägung erübrigt sich somit.
- Der zweite Fall betrifft «relative» Personen der Zeitgeschichte, an denen heute kein öffentliches Interesse mehr besteht, deren Daten aber in den Trainingsinhalten immer noch häufig vorkommen. Hier kann es zwar zu einer inhaltlichen Memorisierung kommen. Sie bedeutet jedoch nicht automatisch, dass diese Personendaten auch abgerufen werden. Sie werden nur dann in den Output finden, wenn danach spezifisch mittels eines entsprechenden Prompts gesucht wird, was bei relativen Personen der Zeitgeschichte unwahrscheinlich ist. Ist dem so, kann vertreten werden, dass das Sprachmodell trotz Memorisierung diese Personendaten nicht enthält, weil nicht mit ihrer Extraktion zu rechnen ist (sog. relativer Ansatz).[122] Kommt ein entsprechender Prompt dennoch vor, wird sich die betroffene Person möglicherweise beim betreffenden Verwender des Sprachmodells gegen die Ausgabe ihrer Daten wehren müssen, analog zur heutigen Situation bei den Suchmaschinen, bei welchen die Suchmaschinen auch dann ein Ergebnis in der Trefferliste sperren müssen, wenn es in der Originalquelle im Internet enthalten und (sogar rechtmässig) frei abrufbar ist (und es der Suchmaschine auch erlaubt ist, den Inhalt zu indizieren; gesperrt werden muss lediglich der Output im Falle bestimmter Suchanfragen). Dies kann der betroffenen Person auch zugemutet werden. Umgekehrt verhält es sich bei derjenigen Person, die ein Sprachmodell trainiert und aus zahlreichen allgemein zugänglichen Informationsquellen Angaben erfährt; andernfalls würde ein Training von vornherein verunmöglicht.

Im Ergebnis ist das Training eines grossen Sprachmodells grundsätzlich durch ein überwiegendes Interesse gerechtfertigt. In Einzelfällen mag dies nicht ohne weiteres zutreffen, doch kann den betroffenen Personen insgesamt zugemutet werden, dass in solchen Fällen nicht das Training untersagt, sondern die Filterung des Outputs der betreffenden Sprachmodelle verlangt wird, um die Persönlichkeit dieser Personen zu schützen, vergleichbar mit der für Suchmaschinen etablierten Praxis.

- **Einwilligung:** Im Kontext von Art. 30 f. DSG kann sich auch die Frage stellen, ob allenfalls eine Einwilligung einer betroffenen Person vorliegt oder die Daten mit ihrem Wissen und Willen allgemein zugänglich gemacht worden sind, womit keine Persönlichkeitsverletzung vorliegt, sofern die betroffene Person im letzteren Fall eine Bearbeitung nicht ausdrücklich

121 Pfaffinger (Fn. 117), DSG 30 N 67.

122 Zum Ganzen: Rosenthal, Teil 19 (Fn. 1).

untersagt hat (Art. 30 Abs. 3 DSG). Während diese beiden Aspekte im konkreten Einzelfall von Relevanz sein können (z.B. bei Ansprüchen einer spezifischen Person), sind sie als Grundlage zur Rechtfertigung eines Trainings nicht praktikabel, da sie eine Einzelfallprüfung erfordern. Eine solche ist aufgrund der Datenmenge aber nicht möglich. Selbst bei Daten von Plattformen, auf welchen vorwiegend von betroffenen Personen selbst publizierte Daten enthalten sind (z.B. Social-Media-Netzwerke), ist mit nicht wenigen Ausnahmen zu rechnen, so dass diese Grundlage nicht als hinreichend zuverlässig gelten dürfte.

2. Bundesorgane als Verantwortliche

[67] Es ist zu beachten, dass sich nur private Verantwortliche auf die vorgenannten Rechtfertigungsgründe berufen können. Bei **Bundesorganen** (oder sonstigen staatlichen Stellen, die nach kantonalen Gesetzen zu agieren haben) existiert neben der Einwilligung im Einzelfall und dem Schutz von Leib und Leben nur die eigene – und ohnehin erforderliche – **Rechtsgrundlage** als Rechtfertigungsgrund für Fälle, in denen betroffene Personen sich gegen eine Bearbeitung aussprechen oder einer der Bearbeitungsgrundsätze nicht eingehalten werden kann.[123] Dies bedeutet, dass in diesen Fällen geprüft werden muss, ob die gesetzliche Aufgabe, mit welcher das Training des Sprachmodells begründet wird, oder die gesetzliche Ermächtigung zur Datenbearbeitung, welche das Training des Sprachmodells erlaubt, auch diese Fälle abdeckt. Dies gilt auch für die Bekanntgabe von besonders schützenswerten Personendaten, die bei privaten Verantwortlichen eine Rechtfertigung erfordert. Bundesorgane benötigen keine, jedoch muss ihre Rechtsgrundlage im Falle von besonders schützenswerten Personendaten eine solche in einem formellen Gesetz sein (oder eine der Ausnahmen ist gegeben, wie beispielsweise die Datenbearbeitung von Personendaten zu nicht personenbezogenen Zwecken nach Art. 39 DSG[124]).

[68] Nehmen wir hier die Rechtsgrundlage der Eidg. Technischen Hochschulen als Beispiel. Art. 36c ETH-Gesetz (**ETHG**) sieht vor, dass die ETH und die Forschungsanstalten im Rahmen von Forschungsprojekten Personendaten, einschliesslich besonders schützenswerter Personendaten, bearbeiten, soweit dies für das entsprechende Projekt erforderlich ist. Sie haben sicherzustellen, dass dabei die Bestimmungen des Datenschutzgesetzes vom 25. September 2020 eingehalten werden. Mit dieser Regelung liegt eine umfassende Rechtsgrundlage in einem formellen Gesetz vor. Die Normdichte ist zwar nicht sehr hoch, doch ergibt sich dies aus der Natur der Sache: Erlaubt sind nur, aber immerhin jene Bearbeitungen, die für ein Forschungsprojekt *erforderlich* sind, auch wenn es sich dabei um besonders schützenswerte Personendaten handelt. Auf den vorliegenden Beispielfall umgemünzt stellt sich somit die Frage, ob das Training eines grossen Sprachmodells als «Forschungsprojekt» bezeichnet werden kann. Hierbei können wir auf unsere obigen Ausführungen zum Urheberrecht verweisen, wo für die entsprechende Schrankenbestimmung ebenfalls auf die «Forschung» abgestellt wird. Wie wir dort bereits ausgeführt haben, kann das Training eines grossen Sprachmodells nach unserer Ansicht durchaus als Forschung bezeichnet werden: Es verfolgt das Ziel, aus den Trainingsdaten durch ein maschinelles Verfahren Erkenntnisse über die Verwendung von Sprache zu gewinnen und diese Erkenntnisse in maschinell verwertbarer Form zu konservieren. Wenn an der ETH grosse Sprachmodelle trainiert werden, erfolgt dies zudem zur Beantwortung zahlreicher weiterer Fragen, die sich im noch zu einem

123 Rampini/Harasgama (Fn. 88), vor DSG 30-32 N 1.

124 Siehe hierzu die obigen Ausführungen zur Parallelnorm für private Datenbearbeiter.

beträchtlichen Teil unerforschten Gebiet des maschinellen Lernens von grossen Sprachmodellen stellen. Dabei sollen im Übrigen auch Fragen zu datenschutzrechtlich relevanten Themen wie die Auslöser von Memorisierung und deren Verhinderung geklärt werden, was wiederum den Einsatz von Personendaten voraussetzt, denn ohne entsprechende Trainings und Sprachmodelle können in diesem Bereich keine Erkenntnisse gewonnen werden. Der Umstand, dass bestimmte Modelle schlussendlich öffentlich zugänglich gemacht werden, steht unseres Erachtens der Qualifikation als Forschung nicht entgegen; es verhält sich hier nicht anders als bei der Publikation einer wissenschaftlichen Arbeit samt entsprechender Messergebnisse. Vor diesem Hintergrund kommen wir zum Ergebnis, dass die Rechtsgrundlage das Training eines grossen Sprachmodells abdeckt. Das Ausbleiben einer Memorisierung ist hierfür keine Bedingung. Sie kann im Gegenteil gerade Gegenstand der Forschung sein. Als Korrektiv wirkt in diesen Fällen nicht die Rechtsgrundlage, sondern der Grundsatz der Verhältnismässigkeit und namentlich die Verhältnismässigkeit im engeren Sinn, bei welcher analog zur Interessenabwägung im Bereich der privaten Datenbearbeitung eine Abwägung der Interessen an der Datenbearbeitung und dagegen vorgenommen werden muss. Hier können die obigen Überlegungen zur Rechtfertigung aus überwiegendem privaten und öffentlichen Interesse analog herangezogen werden.

D. Bekanntgabe von Personendaten ins Ausland

[69] Schliesslich weisen wir darauf hin, dass etwaige Bekanntgaben von Personendaten ins Ausland den Anforderungen von Art. 16 ff. DSG zu genügen haben.[125] Dies wird bei einem Training von grossen Sprachmodellen in der Schweiz keine Herausforderung sein, da es zu keiner solchen Bekanntgabe kommt, auch wenn die Trainingsdaten aus dem Ausland stammen. Dass ein so erstelltes Sprachmodell später auch im Ausland genutzt wird, ist hier ebenfalls nicht relevant, da das Training eines grossen Sprachmodells und dessen spätere Verwendung zwei unterschiedliche Bearbeitungsaktivitäten darstellen. Relevant sein kann die Frage immerhin dann, wenn ein Unternehmen sein Training in einer Cloud-Umgebung vornimmt, die sich im Ausland befindet. Hier stellen sich allerdings keine anderen Schwierigkeiten als beim Cloud-Einsatz sonst auch. Hier hat sich mit dem Angemessenheitsbeschluss des Bundesrates in Bezug auf das CH-US Data Privacy Framework (DPF) aber seit September 2024 eine erhebliche Entspannung ergeben.[126]

E. Fazit

[70] Datenschutzrechtlich ist das Training eines grossen Sprachmodells in der Schweiz unter Verwendung öffentlicher Inhalte auch ohne Einwilligung der betroffenen Personen grundsätzlich zulässig. Wesentlich ist, dass die Natur des Trainings eines Sprachmodells dazu führt, dass die Personendaten lediglich aggregiert Eingang in das Sprachmodell finden, womit der Eingriff in die

125 Demnach dürfen Personendaten nur dann in andere Länder gegeben werden, wenn diese über eine angemessene Datenschutzgesetzgebung verfügen oder, falls nicht, eine der Garantien von Art. 16 Abs. 2 DSG oder eine der Ausnahmen von Art. 17 Abs. 1 DSG greift. M.w.H.: Christian Kunz, in: Adrian Bieri/Julian Powell (Hrsg.), Kommentar zum Schweizerischen Datenschutzgesetz mit weiteren Erlassen, Zürich 2023, DSG 16 N 16.

126 David Rosenthal, Swiss-US DPF: Daten mit und ohne in die USA übermitteln, VISCHER Blog vom 16. August 2024, abrufbar unter: https://www.vischer.com/know-how/blog/swiss-us-dpf-daten-mit-und-ohne-in-die-usa-uebermitteln/.

Persönlichkeit der betroffenen Personen gering ausfällt. Normalerweise besteht daher kein Erfordernis einer Rechtfertigung. Ausnahmsweise kann dennoch bei privatwirtschaftlichen Datenbearbeitern ein Rechtfertigungsgrund gefordert sein. Üblicherweise liegt ein solcher denn auch aufgrund überwiegender privater oder öffentlicher Interessen vor. Dies gilt sogar bei memorisierten Personendaten, da die Memorisierung üblicherweise darauf zurückzuführen ist, dass die Daten in zahlreichen öffentlichen Quellen vorkommen. Unter diesen Umständen kann davon ausgegangen werden, dass daran ein öffentliches und privates Interesse besteht oder die Daten mit Wissen und Willen der betroffenen Personen zugänglich gemacht wurden. Bei Bundesorganen muss zusätzlich die jeweilige Rechtsgrundlage für eigene Datenbearbeitungen geprüft werden; im Falle der ETH liegt mit Art. 36c ETH-Gesetz eine solche für die Erstellung von grossen Sprachmodellen und der damit verbundenen weiteren Forschung beispielsweise vor.

VII. Lauterkeitsrecht

A. Geltungsbereich des UWG

[71] Das UWG ist auf Wettbewerbshandlungen anwendbar.[127] Darunter sind Handlungen zu verstehen, «welche objektiv auf eine Beeinflussung der Wettbewerbsverhältnisse angelegt sind und nicht in einem völlig anderen Zusammenhang erfolgen», wobei das Verhalten des Verletzers marktrelevant, marktgeneigt oder wettbewerbsgerichtet zu sein hat.[128] Vorausgesetzt wird weder ein subjektiver Wille zur Beeinflussung des Wettbewerbs noch eine tatsächliche Beeinflussung eben jenes.[129] Auch ist kein Wettbewerbsverhältnis zu den betroffenen Anbietern oder Abnehmern erforderlich.[130] Handlungen im rein privaten Rahmen oder sich nur unternehmensintern auswirkende Tätigkeiten sind demgegenüber vom Geltungsbereich des UWG ausgenommen.[131]

[72] Das Training eines grossen Sprachmodells fällt also dann nicht in den sachlichen Geltungsbereich des UWG, wenn es rein privaten, nicht kommerziellen Zwecken dient bzw. nur unternehmensinterne Auswirkungen zeitigt. Dies ist beispielsweise dann der Fall, wenn das zu trainierende Sprachmodell nur unternehmensintern und nur so eingesetzt wird, dass es die Wettbewerbsstellung im Aussenverhältnis nicht beeinflusst (z.B. durch eine Verbesserung der Konkurrenzfähigkeit aufgrund der Fähigkeit, Aufträge rascher, besser oder günstiger zu erledigen als die Konkurrenz); zu denken ist etwa an interne Experimente mit KI. Ausserhalb des rein privaten bzw. sich nur unternehmensintern auswirkenden Bereichs ist der Geltungsbereich des UWG demgegenüber eröffnet: Der unentgeltliche Zugriff auf und die Verwendung der frei verfügbaren Inhalte für das Training stellt eine Wettbewerbshandlung dar, weil sich derjenige, der das Modell trainiert, so den Aufwand erspart, der zur originären Erzeugung dieser Trainingsinhalte notwendig gewesen wäre.[132] Eine solche Handlung ist objektiv dazu geeignet, die Wettbewerbsverhältnisse zu beeinflussen.

---

127 Reto M. Hilty, Basler Kommentar UWG, Basel 2013, UWG 1 N 33.

128 BGE 120 II 76, E. 3a.

129 Eugen Marbach/Patrik Ducrey/Gregor Wild, Immaterialgüter- und Wettbewerbsrecht, 4. Aufl., Bern 2017, N 1126.

130 Reto Heizmann, Orell Füssli Kommentar Wettbewerbsrecht II, 2. Aufl., Zürich 2021, UWG 1 N 26.

131 Marbach/Ducrey/Wild (Fn. 129), N 1127; Heizmann (Fn. 130), UWG 1 N 26.

132 Siehe diesen Gedanken in anderem Zusammenhang enthaltend: Botschaft UWG 1983, S. 1047 ff., S. 1070; Urteil des Bundesgerichts 4C.342/2005 vom 11. Januar 2006, E. 3.2.

B. Insbesondere Art. 5 UWG

[73] Die lauterkeitsrechtliche Zulässigkeit des Trainings eines grossen Sprachmodells hängt vor allem von den Bestimmungen von Art. 5 UWG ab.[133] Demnach handelt unlauter, wer (i) ein ihm anvertrautes Arbeitsergebnis wie Offerten, Berechnungen oder Pläne unbefugt verwertet; (ii) ein Arbeitsergebnis eines Dritten wie Offerten, Berechnungen oder Pläne verwertet, obwohl er wissen muss, dass es ihm unbefugterweise überlassen oder zugänglich gemacht worden ist; und (iii) das marktreife Arbeitsergebnis eines andern ohne angemessenen eigenen Aufwand durch technische Reproduktionsverfahren als solches übernimmt und verwertet.

1. Verwertung eines Arbeitsergebnisses?

[74] In allen drei Fallkonstellationen von Art. 5 UWG geht es um die Verwertung eines (fremden) Arbeitsergebnisses. Arbeitsergebnisse sind «Produkte geistiger Anstrengung und materieller Aufwendungen, die ausserhalb des Bereichs der Spezialgesetzgebung zum Schutz von Immaterialgütern nicht geschützt sind».[134] Blosse Ideen, Gedankenblitze und nicht konkret ausgearbeitete Methoden stellen keine Arbeitsergebnisse dar, selbst wenn sie schriftlich fixiert bzw. materialisiert werden.[135] Unter der Verwertung wird jede wirtschaftliche Nutzbarmachung eines fremden Arbeitsergebnisses verstanden.[136]

[75] Es ist fraglich, ob das Training eines grossen Sprachmodells mit den entsprechenden Inhalten tatsächlich eine Verwertung von Arbeitsergebnissen darstellt:

- In Bezug auf das Erlernen von Sprachwissen ist dies bereits deswegen abzulehnen, weil bei den Arbeitsergebnissen das Produkt der geistigen Anstrengung regelmässig nicht die in den Inhalten verwendete Sprache ist. Sie ist vielmehr das Mittel, mit dem das eigentliche Arbeitsergebnis, nämlich das in den Inhalten verkörperte Sachwissen, zum Ausdruck gebracht bzw. materialisiert wird. Dagegen könnte argumentiert werden, dass auch die Sprache und damit das Sprachwissen ein Produkt geistiger Anstrengung sein kann – dann nämlich, wenn die Sprache nicht als Mittel zum Zweck, sondern als eigentlicher Zweck anzusehen ist («Sprachkunst»). Dies dürfte jedoch statistisch selten sein und damit vom Modell nicht häufig genug «gesehen» werden, sodass dieses Sprachwissen nicht Eingang in das Modell findet.
- In Bezug auf das Erlernen von Sachwissen ist zu beachten, dass es beim Training nicht um die Übernahme des konkreten Produkts der geistigen Anstrengung geht, sondern um das Herausgreifen einzelner, darin enthaltener Informationen lediglich im Sinne von Datenpunkten zur Bildung einer Gesamtaussage. Der Input in den Prozess ist also etwas anderes als der Output, und nur der Output wird verwertet. Es ist vergleichbar mit dem Menschen, der mehrere Aufsätze unterschiedlicher Autoren zu einem Thema liest und danach in einem eigenen Arbeitsprodukt in eigenen Worten jene Informationen festhält, die ihm nach der Lektüre insgesamt als gesichert erscheinen. Auch er hat keines der vorbestehenden Ar-

133 In diesem Sinne auch: Städeli/Mary (Fn. 7), S. 249 ff.; Philippe Gilliéron, Réflexions autour de la contractualisation des projets d'intelligence artificielle, in: sic! 2024, S. 423 ff., S. 425.

134 BGE 117 II 199, E. 2a/ee; BGE 122 III 484, E. 8b.

135 BGE 122 III 484, E. 8b.

136 Botschaft UWG 1983, S. 1069; Markus R. Frick, Basler Kommentar UWG, Basel 2013, UWG 5 N 53.

> beitsergebnisse verwertet, sondern lediglich seine Schlüsse daraus gezogen. Wie schon beim Sprachwissen, kommt es auch beim Sachwissen nicht auf den einzelnen Inhalt an; dieser wirkt sich nur aus, falls und wenn es andere Inhalte gibt, die gewissermassen dasselbe sagen, wenn es also «more of the same» gibt. Übernommen wird im Ergebnis nicht der einzelne Inhalt, sondern ein Aggregat aus allem. Es geht also nicht um die Meinung, Empfehlung, Zusammenstellung etc. des einzelnen zu einem Thema, sondern die «herrschende» Lehre dazu. Sie zu ermitteln und zu verwenden muss frei bleiben und soll auch durch das Lauterkeitsreicht nicht von denjenigen monopolisiert werden können, die sie niederschreiben oder dazu beitragen.

[76] Es lässt sich unseres Erachtens also mit guten Gründen argumentieren, dass das Training eines grossen Sprachmodells keine wirtschaftliche Nutzbarmachung von Arbeitsergebnissen und damit keine Verwertung eben jener darstellt, sondern lediglich Erkenntnisse aus diesen Arbeitsergebnissen ableitet.

2. Art. 5 Bst. a und b UWG

[77] Beide Fallkonstellationen erfordern auf die eine oder andere Art und Weise ein «Anvertrautsein». Bei der sog. direkten Vorlagenausbeutung (Fallkonstellation (i)) ergibt sich das bereits aus dem Wortlaut der Bestimmung. Diese Tatbestandsvariante zielt auf jene Situationen, in denen jemand in Übereinstimmung mit dem Erzeuger des Arbeitsergebnisses in dessen Besitz gelangt ist.[137] Demgegenüber dehnt die Fallkonstellation (ii) den Anwendungsbereich von Art. 5 UWG auf Personen aus, denen das Arbeitsergebnis nicht direkt vom Erzeuger anvertraut wurde, sondern die auf andere Weise unbefugt in dessen Besitz gelangt sind (sog. indirekte Vorlagenausbeutung).[138] Obwohl der Wortlaut des Gesetzes dies nicht ausdrücklich erwähnt, muss das Arbeitsergebnis nach einhelliger Meinung auch hier der Person, die es dem Verletzer aushändigt, vom Erzeuger anvertraut worden sein.[139] Anvertrautsein setzt eine vertragliche, vorvertragliche oder vertragsähnliche Beziehung voraus,[140] die in einem Verwertungsverbot zulasten des das Arbeitsergebnis Entgegennehmenden resultiert.[141]

[78] Sowohl die Fallkonstellation (i) als auch die Fallkonstellation (ii) erfordert also zumindest eine vertragsähnliche Beziehung (entweder zwischen dem Erzeuger und dem Verletzer (i) oder zwischen dem Erzeuger und dem Vermittler (ii)). Eine direkte vertragliche, vorvertragliche oder vertragsähnliche Beziehung zwischen dem Erzeuger und dem potenziellen Verletzer, also demjenigen, der das Sprachmodell trainiert, dürfte kaum je gegeben sein. Wir gehen jedoch auf diese Fallkonstellation, in der ein erhaltener Inhalt vertragswidrig für KI-Trainingszwecke verwendet wird, in Kapitel IX ein. Auch eine vertragliche bzw. vertragsähnliche Beziehung zwischen dem Erzeuger und dem Vermittler dürfte nicht den Regelfall darstellen: Inhalte werden im Internet sehr oft einfach von irgendwo übernommen und weiterverbreitet – man denke etwa an Social Media und sog. Reposts. Regelmässig handelt es sich um eine «Kette» des Informationsflusses,

137 Botschaft UWG 1983, S. 1069.

138 Lukas Fahrländer, DIKE UWG Kommentar, Zürich 2018, UWG 5 Bst. a und b N 24.

139 Botschaft UWG 1983, S. 1070; Frick (Fn. 136), UWG 5 N 58; Fahrländer (Fn. 138), UWG 5 Bst. a und b N 25.

140 BGE 133 III 431, E. 4.5.

141 Frick (Fn. 136), UWG 5 N 44.

wobei die Inhalte in der Regel nur dem ersten Vermittler vom Erzeuger anvertraut werden, nicht aber demjenigen Vermittler, dessen Inhalt dann effektiv für das Training des grossen Sprachmodells verwendet wird. Das Kriterium des Anvertrautseins schliesst daher schon viele Inhalte vom Anwendungsbereich der Fallkonstellationen (i) und (ii) aus.

[79] Bei der Fallkonstellation (ii) kommt hinzu, dass der Zweiterwerber, d.h. der Dritte, der das Arbeitsergebnis zu seinen Gunsten verwertet, wissen muss, dass das Arbeitsergebnis ihm unbefugterweise überlassen oder zugänglich gemacht worden ist. Die erforderliche subjektive Komponente beim Zweiterwerber bezieht sich mit anderen Worten auf den Verstoss gegen das vertragliche, vorvertragliche oder vertragsähnliche Verwertungsverbot.[142] Für eine zivilrechtliche Verantwortlichkeit genügt bereits fahrlässige Unkenntnis.[143] Das bedeutet jedoch nicht, dass den Zweiterwerber eine generelle Nachfragepflicht trifft; eine solche ist nur unter Umständen zu bejahen, die nahelegen, dass die vom Vermittler übergebenen Unterlagen nicht von ihm selbst vorbereitet sind.[144]

[80] Damit der Verwender der Inhalte für das Training eines Sprachmodells in Bezug auf das unbefugte Überlassen bzw. unbefugte Zugänglichmachen fahrlässig handelt, muss er die verkehrsübliche Sorgfalt verletzen. In diesem Zusammenhang können die Überlegungen zur Provider-Haftung[145] analog herangezogen werden: In der Lehre wird überwiegend die Meinung vertreten, dass Hosting-Provider nicht zur präventiven Prüfung aller Inhalte verpflichtet sind; nur, wenn sie explizit etwa auf eine Persönlichkeitsverletzung hingewiesen werden, müssen sie diese untersuchen und den fraglichen Inhalt gegebenenfalls entfernen.[146] Ein Grund dafür ist, dass der Aufwand für die Plattformbetreiber ansonsten enorm hoch und eigentlich nicht zu stemmen wäre. In Bezug auf Internet-Suchmaschinen befand das Genfer Tribunal de première instance, dass es den Betreibern der Suchmaschinen nicht zuzumuten sei, jede einzelne von ihr verzeichnete Webseite auf rechtswidrige Inhalte zu kontrollieren.[147] Gleiches muss auch im vorliegenden Zusammenhang gelten: Die verkehrsübliche Sorgfalt kann nicht dahingehend verstanden werden, dass im Rahmen des Trainings von grossen Sprachmodellen hinsichtlich jedes einzelnen Inhalts zu prüfen ist, ob Anhaltspunkte dafür bestehen, dass der Inhalt dem Vermittler anvertraut wurde und somit in Verletzung eines Verwertungsverbots dem Verwender überlassen bzw. zugänglich gemacht wird. Andernfalls würde das Training von grossen Sprachmodellen von vornherein verhindert, was nicht Telos von Art. 5 UWG bzw. des UWG als solchen sein kann. Es ist vielmehr so lange von der Gutgläubigkeit des Verwenders auszugehen, als dass dieser einerseits angemessene Massnahmen trifft, damit die Datensammelaktivitäten für das Training keine Inhalte aus bekanntermassen rechtswidrigen Quellen – wie z.B. Darknet-Foren zur Verbreitung gestohlener

142 Fahrländer (Fn. 138), UWG 5 Bst. a und b N 1.

143 Urteil des Bezirksgerichts Zürich GG040064/U vom 23. August 2005, in: sic! 2006, S. 112 ff., E. XV.3.1; Heizmann (Fn. 130), UWG 5 N 15; Frick (Fn. 136), UWG 5 N 59; Fahrländer (Fn. 138), UWG 5 Bst. a und b N 27.

144 Frick (Fn. 136), UWG 5 N 59.

145 Siehe für eine umfassende Abhandlung etwa: Der Bundesrat, Die zivilrechtliche Verantwortlichkeit von Providern, Bericht vom 11. Dezember 2015, abrufbar unter: https://www.bj.admin.ch/bj/de/home/publiservice/publikationen/berichte-gutachten/2015-12-11.html; auch: David Rosenthal, Internet-Provider-Haftung – ein Sonderfall?, in: Peter Jung (Hrsg.), Tagungsband Recht aktuell, Bern Edition Weblaw 2006.

146 Vgl. m.w.Verw.: Bericht Bundesrat (Fn. 145), S. 63; Alexander Kernen, Volle Verantwortlichkeit des Host Providers für persönlichkeitsverletzende Handlungen seines Kunden, in: Jusletter vom 4. März 2013, N 16; Betty-Annett Meier, II. Theoretischer Teil / F. – H., in: Roland Müller/Thomas Geiser/Kurt Pärli (Hrsg.), Bewertung des Arbeitgebers im Internet, 2018, S. 59.

147 Urteil des Tribunal de première instance C/9894/2007-17 vom 4. November 2008.

Daten – einschliessen, und andererseits Quellen spezifisch ausschliesst, bei denen er konkret und in relevanter Weise darauf hingewiesen wurde, dass sie das Lauterkeitsrecht verletzende Inhalte enthalten. Das dürfte erfüllt sein, wenn im Rahmen eines Crawlings für KI-Trainingszwecke gängige, verfügbare Sperrlisten benutzt werden, wie sie auch Unternehmen und Provider für das Filtern ihrer Internet-Zugänge nutzen (damit Mitarbeitende bzw. Kunden keine verbotenen oder gefährliche Websites nutzen) und bestimmte Techniken, wie sie für den Zugriff auf die meisten Darknet-Websites nötig sind (z.B. TOR-Browser), nicht einsetzen; eine vorgängige Einzelfallbeurteilung oder Recherche zur Erstellung eigener Sperrlisten ist unseres Erachtens nicht erforderlich. Alternativ kann auf bereits gecrawlte, im Markt angebotene Inhalte mit entsprechendem Ruf zurückgegriffen werden.

3. Art. 5 Bst. c UWG

[81] Hingegen kann Fallkonstellation (iii) auf den ersten Blick durchaus vorliegen: Erfasst sind hier zunächst alle öffentlichen Inhalte, die als marktreifes Arbeitsergebnis betrachtet werden können (d.h. als solche auch vermarktet werden können, wie z.B. Inhalte von Medien, sozialen Medien oder von Plattformen für User Generated Content).[148] Diese müssten einerseits «als solche» übernommen und verwertet werden, und dies hat «ohne angemessenen eigenen Aufwand» zu erfolgen. Werden Arbeitsergebnisse somit hinreichend verändert, bevor sie (erneut) zum Einsatz kommen, oder wird hinreichender Aufwand betrieben, sind die Voraussetzungen des Tatbestands nicht erfüllt.[149] Kriterien wie die «Marktreife» und das «Arbeitsergebnis» werden zwar bei vielen Inhalten im Internet nicht ohne weiteres erfüllt sein, aber angesichts der Mengen an Inhalten, die für das Training eines grossen Sprachmodells verwendet werden, werden zweifellos etliche Inhalte diese Kriterien trotzdem erfüllen. Somit kommt es darauf an, ob ein Arbeitsergebnis als solches und ohne angemessenen eigenen Aufwand *übernommen* und – nach herrschender Rechtsprechung[150] – auch als solches und ohne angemessenen eigenen Aufwand *verwertet* wird. Verwertung verlangt hierbei, dass das übernommene Arbeitsergebnis dem Markt zugeführt wird, denn nur dann beschlägt es den Schutzbereich des UWG.[151]

[82] Von all dem ist beim Training eines grossen Sprachmodells *prima vista* nicht auszugehen: Nicht nur ist der Aufwand für das Training in der Regel beträchtlich (dazu sogleich), auch werden Trainingsdaten nicht als solche übernommen, sondern die darin enthaltenen Sprach- und Sachinformationen werden aggregiert. Damit scheidet auch eine Verwertung als solche naturgemäss aus, jedenfalls solange sich das Arbeitsergebnis nicht «als solches» im Output des Sprachmodells wiederfinden wird. Wie schon beim Urheberrecht verhilft auch hier das Ausbleiben einer Memorisierung zur Rechtmässigkeit (mit dem Unterschied, dass das Lauterkeitsrecht auch bei Inhalten greifen kann, die nicht urheberrechtlich geschützt sind[152]): Kann die Memorisierung verhindert werden, fehlt es schon an der Übernahme des Arbeitsergebnisses «als solches», und erst recht an der Verwertung. Kommt sie demgegenüber vor, könnte argumentiert werden, dass das betref-

148 Rolf H. Weber/Lennart Chrobak, DIKE UWG Kommentar, Zürich 2018, UWG 5 N 15 und 18; vgl. auch Urteil des Appelationshofs Bern vom 21. Mai 2001, in: sic! 2001, S. 613 ff., E. 9; BGE 131 III 384, E. 4.2.

149 Heizmann (Fn. 130), UWG 5 N 20 und N 24.

150 BGE 131 III 384, E. 4.3.

151 Weber/Chrobak (Fn. 148), UWG 5 Bst. c N 25.

152 Marbach/Ducrey/Wild (Fn. 129), N 8; Hilty (Fn. 17), N 19 *e contrario*.

fende Arbeitsergebnis (z.B. ein Artikel, der diverse Male für das Training verwendet worden ist) im Modell als solches Eingang gefunden hat. Es kann bei dieser Ausgangslage aber auch ebenso vertreten werden, das Arbeitsergebnis sei zwar mit seinem Inhalt in das Modell eingeflossen, aber nicht «als solches» übernommen worden, sondern in anderer Form, nämlich als aggregiertes Sprach- und Sachwissen, das nur in Kombination mit einem passenden Prompt zur Erzeugung eines Textes genutzt werden kann, der dem ursprünglichen Arbeitsergebnis mehr oder weniger entspricht. Denn im Modell selbst werden Trainingsinhalte nie «als solche» gespeichert, sondern in einer davon abstrahierten Form, wie bereits vorne im Zusammenhang mit dem Urheberrecht gezeigt wurde.[153] Es müsste dann noch gezeigt werden, dass das Arbeitsergebnis mit dem passenden Prompt in seiner ursprünglichen Form abgerufen werden kann und mit einem solchen Prompt durch einen Dritten im Markt auch zu rechnen ist, um von einer Verwertung des Arbeitsergebnisses als solches auszugehen (und es nachzuweisen). Diese kombinierten Umstände dürften, falls überhaupt, selten vorliegen (vgl. jedoch den US-Fall, in welchem die New York Times gegen OpenAI und Microsoft geklagt hat[154]).

[83] Kann die Memorisierung nicht verhindert werden, jedenfalls in Einzelfällen nicht, stellt sich die Frage, ob wenigstens ein angemessener eigener Aufwand zur Übernahme oder Verwertung oder beidem erfolgt. Es ist ein doppelter Aufwandvergleich vorzunehmen: Einerseits ist die Leistung des Erstbewerbers mit der des Zweitbewerbers zu vergleichen und andererseits die Leistung des Zweitbewerbers mit seinem hypothetischen Aufwand bei eigenem Nachvollzug der einzelnen Produktionsschritte.[155] Unlauteres Verhalten ist nur dann einschlägig, wenn dem Zweitbewerber bei keinem der beiden Vergleiche ein angemessener Eigenaufwand attestiert werden kann.[156]

[84] Insbesondere ist dem Aufwand des Übernehmers bzw. Verwenders also der Aufwand des Erstbewerbers gegenüberzustellen.[157] Dabei gilt es zu berücksichtigen, dass das Training zwar automatisiert verläuft und für die Verarbeitung eines einzelnen Textes auch nicht sehr viel Rechenaufwand erforderlich ist. Das Training eines grossen Sprachmodells setzt jedoch insgesamt die Bereitstellung einer hohen Rechenleistungen voraus, weil es darauf angewiesen ist, mit sehr viel Text gefüttert zu werden, der auch entsprechend kuratiert werden muss. Entsprechend hoch ist die zur Durchführung eines Trainings erforderliche Investition.[158] Während vertreten werden kann, dass für den Aufwandsvergleich nur derjenige Bruchteil des Aufwands des Übernehmers berücksichtigt werden darf, der auf das fragliche Arbeitsergebnis entfällt, kann unseres Erachtens ebenso und mit überzeugenderen Gründen vertreten werden, dass auf die hohe Initialinvestition *in globo* abzustellen ist – dies, weil auch keine Verwertung des einzelnen Arbeitsergebnisses in Frage kommt, sondern nur eine Verwertung aller Arbeitsergebnisse zusammen in Form des grossen Sprachmodells. Des Weiteren ist richtigerweise zunächst jener Aufwand der Erstbewerber zu berücksichtigen, den diese für das allgemeine Sprachwissen aufgewendet haben, der in den Inhalten steckt, denn auf dieses kommt es beim Training eines Sprachmodells zunächst an. Unter diesen Umständen wird bereits aufgrund des «ersten» Aufwandvergleichs von einem angemesse-

153 Kapitel V.C.

154 Siehe: https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html.

155 Botschaft UWG 1983, S. 1071; BGE 131 III 384, E. 4.4.1.

156 BGE 131 III 384, E. 4.4.1; vgl. auch: Reto Arpagaus, Basler Kommentar UWG, Basel 2013, UWG 5 N 92.

157 Für das Training eines KI-Systems: Städeli/Mary (Fn. 7), S. 250; allgemein Weber/Chrobak (Fn. 148), UWG 5 Bst. c N 50.

158 Siehe hierzu etwa: Inside IT, So viel kosten grosse KI-Modelle, 22. April 2024, abrufbar unter: https://www.inside-it.ch/so-viel-kosten-grosse-ki-modelle-20240422.

nen eigenen Aufwand zur Übernahme und Verwertung auszugehen sein, weil der Aufwand, den das Unternehmen für die Sprachfertigkeiten seiner Mitarbeitenden in Bezug auf einen einzelnen Inhalt aufwendet, in der Regel vernachlässigbar ist und einzelne Ausnahmen aufgrund der Art und Weise, wie Sprachmodelle lernen, nicht relevant sind. Dasselbe wird auch für tatsächlich übernommenes Sachwissen gelten, weil dieses dafür so geläufig und verbreitet sein muss, dass es auch in zahlreichen anderen Quellen vorkommt.

[85] Schliesslich ist noch darauf hinzuweisen, dass der Schutz durch Art. 5 Bst. c UWG zeitlich befristet ist. Diese Bestimmung greift nämlich nur, solange die Kosten für die Erstellung der Arbeitsleistung nicht bereits amortisiert sind.[159] Damit dürfte bereits die zeitliche Schranke die Anwendung von Art. 5 Bst. c UWG auf das Training eines grossen Sprachmodells weitgehend ausschliessen.[160]

C. Übrige Bestimmungen

[86] Im Übrigen ist zu beachten, dass das Lauterkeitsrecht zwecks Schutz eines funktionierenden Wettbewerbs[161] im Ergebnis auch «inhaltliche» Vorgaben macht, indem es etwa für unlauter erklärt, wenn jemand andere und ihre Produkte oder Geschäftsverhältnisse durch irreführende oder unrichtige Äusserungen herabsetzt (Art. 3 Abs. 1 Bst. a UWG), oder wenn jemand Verwechslungen mit Waren, Werken, Leistungen oder dem Geschäftsbetrieb anderer herbeiführt (Art. 3 Abs. 1 Bst. d UWG). Diese Bestimmungen greifen unseres Erachtens aber erst, wenn ein grosses Sprachmodell in einer konkreten Anwendung eingesetzt wird, und nicht schon im Training.

[87] Das gilt im Übrigen auch für das Markenrecht, das hier nicht separat abgehandelt wird, aber mit dem Lauterkeitsrecht verwandt ist, wo es dem gewerblichen Rechtsschutz dient. In Trainingsdaten können durchaus Kennzeichen vorkommen, die markenrechtlich geschützt sind. Der Markenschutz gibt dem Inhaber aber nur das Recht, anderen den kennzeichenmässigen Gebrauch seiner Marke zu verbieten, also die Marke auf Waren oder für Dienstleistungen zu gebrauchen oder auf Geschäftspapier (vgl. Art. 13 des Markenschutzgesetzes, **MSchG**). Wird ein Trainingsinhalt, der eine fremde Marke enthält, für das Training eines Sprachmodells benutzt, stellt dies kein kennzeichenmässiger Gebrauch dar und kann daher unter diesem Titel auch nicht verboten werden.[162] Allerdings kann eine Marke auch urheberrechtlichen Schutz geniessen. Diesfalls gilt, was für urheberrechtlich geschützte Werke ausgeführt worden ist.[163]

D. Fazit

[88] Lauterkeitsrechtlich ist es von Vorteil, wenn eine Memorisierung von fremden Inhalten, die als «Arbeitsergebnis» eines Dritten betrachtet werden könnten, vermieden wird. Doch selbst wenn es zu einer solchen kommt, kann vertreten werden, dass die lauterkeitsrechtlichen Vorgaben von Art. 5 UWG nicht tangiert sind, weil lediglich aus diesen Arbeitsergebnissen abgeleitete

159 BGE 134 III 166, E. 4.2 f.

160 Städeli/Mary (Fn. 7), S. 250.

161 Botschaft 1983, S. 1039.

162 In diesem Sinne auch: Berger (Fn. 14), N 15.

163 Siehe dazu die Ausführungen in Kapitel V.

Erkenntnisse übernommen werden, und auch dies nur, wenn sie statistisch relevant sind, d.h. zum allgemeinen Sprach- und Sachwissen gehören, das nicht monopolisierbar ist. Auch sollten für das Training keine vertraulichen Inhalte verwendet werden, wo dies nach den entsprechenden Abmachungen untersagt worden ist. Auf Inhalte aus rechtswidrigen Quellen sollte im Rahmen des Trainings ebenfalls verzichtet werden.

VIII. Exkurs: Crawler-Verbote

[89] In der Praxis wird immer wieder auch die Frage diskutiert, ob die auf Websites immer häufiger anzutreffenden maschinenlesbaren Vermerke, wonach Crawling bzw. Scraping unerwünscht ist, beachtet werden müssen (z.B. «robots.txt»).[164] Solche Vermerke können urheberrechtliche, datenschutzrechtliche und lauterkeitsrechtliche Relevanz haben:

- Aus urheberrechtlicher Sicht führen sie dazu, dass im Falle eines Sperrvermerks für Crawler nicht von einer impliziten Zustimmung des Website-Betreibers ausgegangen werden kann. Allerdings ist die implizite Zustimmung ohnehin eine unsichere Rechtslage für ein Crawling; der Website-Betreiber müsste zugleich Rechteinhaber sein, sei es, um von einer impliziten Zustimmung auszugehen, sei es, um aus dem Widerspruch auf das Fehlen einer Zustimmung zu schliessen. Zudem greift wie bereits aufgezeigt die Schrankenbestimmung der wissenschaftlichen Forschung. Diese dient ja gerade dazu, Inhalte auch ohne Zustimmung des Rechteinhabers bearbeiten zu können. Sie verlangt nur, dass der Zugang zu den Werken rechtmässig erfolgt, was auch dann der Fall wäre, wenn ein Sperrvermerk vorhanden ist, der Zugriff aber nicht blockiert wird. Solche Sperrvermerke stehen rein urheberrechtlich jedenfalls in der Schweiz einem Scraping daher nicht entgegen.
- Aus datenschutzrechtlicher Perspektive ist ein Sperrvermerk zunächst nur beachtlich, wenn er von der betroffenen Person selbst kommt oder ihr zugerechnet werden kann. Das wird in der Mehrheit der Fälle nicht der Fall sein. Ist es der Fall, liegt ein Widerspruch gegen die Bearbeitung ihrer Personendaten vor. Eine weitere Bearbeitung stellt demnach eine Persönlichkeitsverletzung dar. Nach Schweizer Recht kann eine solche jedoch jedenfalls im privatwirtschaftlichen Bereich durch ein überwiegendes Interesse gerechtfertigt werden. Hierzu fällt zuallererst der Rechtfertigungsgrund der nicht personenbezogenen Bearbeitung von Personendaten in Betracht. Dieser versagt nur, wenn es zu einer Memorisierung der Personendaten kommt. Diesfalls wird zu prüfen sein, ob es sich um eine öffentliche Person handelt und daher ein überwiegendes Interesse an der Datenbearbeitung besteht. Wird die Memorisierung hinreichend gut verhindert, so dass tatsächlich nur mit der Memorisierung von Personendaten von Personen zu rechnen ist, die erheblich in der Öffentlichkeit stehen, dann ist ein Sperrvermerk datenschutzrechtlich irrelevant. Im Falle einer Bearbeitung durch ein Bundesorgan ist zu prüfen, ob die Rechtsgrundlage die «Übergehung» des Widerspruchs rechtfertigt, was letztlich im Rahmen der Verhältnismässigkeit zu prüfen sein wird und daher wiederum – bei der Verhältnismässigkeit im engeren Sinn – in einer Interessenabwägung mündet. Es gilt das für private Bearbeiter gesagte.

164 Marmy-Brändli/Oehri (Fn. 14), S. 663.

- Bleibt noch die lauterkeitsrechtliche Beurteilung. Hier verweisen wir auf obenstehende Ausführungen, warum jedenfalls die Voraussetzungen von Art. 5 Bst. c UWG im Regelfall nicht erfüllt sein werden. Sind diese Voraussetzungen nicht erfüllt und handelt es sich um öffentlich zugängliche Informationen, so ist unserer Ansicht nach die Übernahme von fremden Inhalten lauterkeitsrechtlich selbst dann nicht verpönt, wenn sie gegen die ausdrückliche Äusserung desjenigen erfolgt, der die Inhalte frei zugänglich bereitstellt. Dies fördert den Wettbewerb und schränkt ihn nicht unnötig ein (was durch die Nicht-Anwendbarkeit von Art. 5 Bst. c UWG deutlich wird). Vor diesem Hintergrund kann nach unserem Dafürhalten auch die Generalklausel des UWG nicht einschlägig sein, womit kein lauterkeitsrechtliches Verbot besteht.

[90] Das vorstehend Gesagte gilt im Übrigen auch für Crawler-Verbote in auf Websites enthaltenen Nutzungsbedingungen, die nicht maschinenlesbar sind bzw. sich nicht an Roboter richten. Dies liegt darin begründet, dass das blosse Aufschalten von Nutzungsbedingungen auf einer Website noch keinen Vertrag mit den Besuchern der Website begründet und sie daher rechtlich nicht bindet. So können sich Besucher von Websites auf etwaige Einwilligungen, die über solche Nutzungsbedingungen kundgetan werden, berufen (d.h. wenn diese eine bestimmte Nutzung von Inhalten erlauben), aber Verbote sind für sie unbeachtlich; sie müssen sich in diesen Fällen nur – aber immerhin – an die gesetzlichen Vorgaben des Urheber-, Datenschutz-, Lauterkeits- und des sonstigen Rechts halten.

[91] Anders verhält es sich, wenn sich Benutzer vor der Nutzung von Inhalten einer Website registrieren müssen. In diesen Fällen kommt regelmässig ein Vertrag zustande, in dessen Rahmen sich Benutzer verpflichten können, die Inhalte nicht zu «scrapen» bzw. zu «crawlen» oder für KI-Zwecke zu nutzen (dazu sogleich).

IX. Vertragsrecht

A. Vertragsverletzung?

[92] Stammen Trainingsinhalte nicht aus dem öffentlichen Internet oder anderen öffentlichen Quellen, ist unabhängig von urheberrechtlichen, datenschutzrechtlichen und lauterkeitsrechtlichen Vorgaben zu prüfen, ob sich derjenige, der mit den betreffenden Trainingsdaten ein grosses Sprachmodell trainieren will, allenfalls vertraglich verpflichtet hat, dies nicht zu tun oder nur unter bestimmten Bedingungen, wie z.B. der Bezahlung einer Lizenzgebühr. Hier sind keine allgemeinen Aussagen möglich, mit der Ausnahme, dass erfahrungsgemäss die meisten Geheimhaltungsklauseln in Verträgen auch ein Verbot der zweckfremden Nutzung von Geschäftsgeheimnissen der Gegenpartei enthalten (Zweckbindungsklausel). Selbst wenn beim Training eine Memorisierung ausgeschlossen werden kann oder das Modell Dritten nicht zugänglich gemacht wird, steht ein solches Umnutzungsverbot einer Verwendung der Daten für das Training entgegen.

[93] Eine Geheimhaltungsklausel ohne Zweckbindung (d.h. ohne Verbot einer zweckfremden Nutzung) steht dem Training eines Sprachmodells jedenfalls dann nicht entgegen, wenn eine wortwörtliche und auch nur sinngemässe Memorisierung von Inhalten (d.h. der abstrahierten Information) ausgeschlossen werden kann, jedenfalls dergestalt, dass kein Grund zur Annahme besteht, dass der Output einen Rückschluss auf den Geheimnisherrn (d.h. i.d.R. den Geschäfts-

partner) erlaubt. Eine *Membership Inference Attack*[165], die immer wieder als Bedrohungsszenario für KI-Modelle angeführt wird, stellt insofern in der Regel kein Problem dar, als demjenigen, dem ein Geheimnis bereits bekannt ist, ein solches nicht mehr «verraten» werden kann.[166]

[94] Ob das Training in Verletzung eines Vertrags erfolgt, muss durch Auslegung des Vertrags ermittelt werden. Sofern das Training eines Sprachmodells nicht klar geregelt oder untersagt ist, muss ermittelt werden, wie der Vertrag in diesem Punkt in guten Treuen zu verstehen ist. Führt dies zum Ergebnis, dass der Vertrag eine Lücke enthält, die Parteien also zum relevanten Thema («Dürfen bestimmte Daten für ein KI-Training benutzt werden und unter welchen Bedingungen?») nichts abgemacht haben, so muss diese Vertragslücke gefüllt werden. Hierbei ist einerseits das dispositive Gesetzesrecht, andererseits der hypothetische Parteiwille zu beachten (wenn angenommen werden muss, dass die Parteien den Vertrag ergänzt hätten, falls ihnen das Thema bekannt gewesen wäre).[167] Was Vorrang hat, ist umstritten.[168]

[95] In Bezug auf den hypothetischen Parteiwillen gehen wir – vorbehaltlich einer Zweckbindungsklausel – davon aus, dass Geschäftspartner in der Regel akzeptieren, dass der jeweilige Partner aus der Zusammenarbeit oder den fremden Leistungen eigenes Know-how gewinnt. Dabei handelt es sich um einen üblichen und akzeptierten Vorgang, selbst wenn dieses Know-how später monetarisiert werden kann. Es wird nur aber immerhin erwartet, dass die Vertraulichkeit gewahrt wird, d.h. gegenüber Dritten keine Geschäftsgeheimnisse offengelegt werden bzw. für Dritte keine Rückschlüsse auf den Geschäftspartner möglich sind.

[96] Diese Erkenntnis kann auch auf das Training von grossen Sprachmodellen angewandt werden, da ein solches Training vergleichbar ist mit der Gewinnung von Know-how. Dieses fällt beim Training von grossen Sprachmodellen nicht direkt bei den Mitarbeitenden des Unternehmens (und ggf. in deren Dokumentenablage) an, sondern beim vom Unternehmen kontrollierten Modell. Funktionell ist beides allerdings eine Ressource, welche das Unternehmen für seine anderweitigen geschäftlichen Aktivitäten nutzen kann. Die technische «Konservierung» von Know-how aufgrund von Geschäftsaktivitäten ist überdies nichts Neues. Jedes Unternehmen, das Produkte anbietet und dafür einen Kundensupport betreibt, wird sich ab einem gewissen Zeitpunkt eine «Knowledge-Base» anlegen, um die gesammelten Erfahrungen für die künftige Kundenbetreuung und Verbesserung seiner Produkte zu nutzen. Das Training eines KI-Modells ist eine andere Form der Konservierung von Know-how und in diesem Fall überdies mit dem sehr engen Fokus auf Sprachwissen, auch wenn sie den Vorteil hat, dass sie selbständig kommerzialisiert werden kann. Vor diesem Hintergrund ist davon auszugehen, dass vernünftige Parteien es einander unter der genannten Bedingung der Verhinderung der Memorisierung normalerweise nicht untersagen würden, die im Rahmen von Geschäftsaktivitäten anfallenden Daten auch für das Training von KI-Modellen zu verwenden.

---

165 Bei einer solchen versucht ein Angreifer herauszufinden, ob bestimmte Inhalte (z.B. die Daten einer bestimmten Person) für das Training verwendet worden sind, d.h. diese Daten «Mitglied» der Trainingsdaten waren. Dies wird anhand der vom Modell gelieferten Antworten ermittelt.

166 M.w.H. zur Membership Inference Attack: Lisa Käde/Stephanie von Maltzan, Algorithmen, die nicht vergessen – Sicherheitslücken in Machine-Learning-Modellen und deren Bedeutung für den Schutz der Daten und der Urheberrechte, in: InTeR 2020, S. 201 ff., S. 203.

167 Statt vieler: BGE 115 II 484, E. 4b.

168 Alfred Koller, OR AT Band I, 5. Aufl., Bern 2023, N 10.21: Das Bundesgericht hält immer wieder fest, dass der Richter bei Vorliegen einer Vertragslücke zuerst das dispositive Recht und erst dann, wenn dieses nicht weiterhilft, den hypothetischen Parteiwillen heranzuziehen hat (siehe z.B. BGE 115 II 484, E. 4b). Tatsächlich geht das Bundesgericht bei der Urteilsfindung in der umgekehrten Reihenfolge vor.

[97] Eine Einschränkung des Gesagten kann dort bestehen, wo besondere kommerzielle Interessen an dem für das dem Know-how bzw. Training benötigten Datenmaterial bestehen, d.h. davon ausgegangen werden muss, dass ein Geschäftspartner die kommerzielle Verwertung des von ihm beigesteuerten Datenmaterials durch den anderen Geschäftspartner nicht ohne Beteiligung am Geschäftserfolg zulassen würde. Das wird insbesondere dort der Fall sein, wo das Datenmaterial selbst einen kommerziellen Wert darstellt, den das Unternehmen, dem das Datenmaterial gehört, ausschöpfen will.

B. Rechtsfolgen der Verletzung vertraglicher Pflichten

[98] Es stellt sich in diesem Zusammenhang mitunter die Frage, welches die Rechtsfolgen einer Verletzung vertraglicher Pflichten sind. Grundsätzlich sind folgende Konsequenzen denkbar:

- **Kündigung:** Wird ein Vertrag verletzt, jedenfalls wesentlich, kann er in der Regel gekündigt werden. Das ist in den Fällen von besonderer Relevanz, in denen es sich um einen Lizenzvertrag handelt, womit die Kündigung des Vertrags auch zum Wegfall der Nutzungsbefugnis für das lizenzierte Werk führt. Beinhaltet ein Vertrag den Bezug einer Dienstleistung, kann anstelle einer ausserordentlichen Kündigung auch «nur» das Aussetzen der Dienstleistung vorgesehen sein. Diesfalls gilt der Vertrag weiterhin (inklusive Zahlungspflichten), aber die den Vertrag verletzende Partei kann nicht mehr davon profitieren.
- **Vertragsstrafe:** Jedenfalls nach Schweizer Recht steht es den Parteien frei, für den Fall der Verletzung einer bestimmten vertraglichen Regelung eine Konventionalstrafe zu vereinbaren (Art. 160 Abs. 1 des Obligationenrechts, **OR**). Konventionalstrafen sind heute nicht sehr verbreitet. Sie kommen aber mitunter im Zusammenhang mit Geheimhaltungsklauseln vor. Das ist hier von besonderer Relevanz, da solche Klauseln oft auch Zweckbeschränkungen in Bezug auf die von der Vertragsgegenseite erhaltenen Daten vorsehen. Die Verletzung einer solchen Zweckbindung kann sodann – je nach Formulierung – die Vertragsstrafe auslösen.
- **Schadenersatz:** Die Verletzung eines Vertrags kann einen Anspruch auf Ersatz des daraus entstehenden Schadens auslösen (vgl. etwa Art. 97 OR). Ein Schaden ist eine unfreiwillige Vermögensverminderung.[169] Woraus diese im Falle einer unzulässigen Verwendung von Daten eines Geschäftspartners zwecks Trainings eines Sprachmodells besteht, ist uns allerdings nicht klar. Ausserdem muss die unfreiwillige Verminderung des Vermögens vom Geschäftspartner nachgewiesen werden. Denkbar ist ein Schaden, wenn es zu einer Memorisierung und damit Bekanntgabe von Daten des Geschäftspartners kommt oder wenn dieser aufgrund der Verwendung des Trainingsmaterials zu Ansprüchen Dritter kommt (weil das Datenmaterial ursprünglich von Dritten stammt). Auch Rechtsverfolgungskosten des Geschäftspartners können einen Schaden begründen. Insgesamt dürfte der Anspruch auf Schadenersatz vorliegend jedoch schwierig durchzusetzen sein. Denkbar ist in bestimmten Konstellationen immerhin noch, dass der in seinen Vertragsansprüchen verletzte Geschäftspartner die Herausgabe des Gewinns fordert, den sein Vertragspartner durch die unerlaubte Verwendung von Datenmaterial erzielt hat. Wird ein Sprachmodell aber nicht kommerziell messbar genutzt, dürfte auch dies ein schwieriges Unterfangen sein und es können in

169 Statt vieler: BGE 144 III 155, E. 2.2.

der Praxis höchstens eingesparte Aufwendungen geltend gemacht werden, die allerdings ebenfalls nachzuweisen wären.

- **Andere Nachteile**: In einem Vertrag können die Parteien die Konsequenzen von Vertragsverletzungen im Rahmen des geltenden Rechts letztlich beliebig regeln. Es sind auch andere Konsequenzen denkbar, wie z.B. die Pflicht, die unerlaubte Verwendung des Datenmaterials rückgängig zu machen (z.B. durch Löschung des KI-Modells). Solches müsste unseres Erachtens jedoch konkret vereinbart worden sein.
- **Lauterkeitsrecht:** Kommt es durch die Verwendung von Datenmaterial eines Geschäftspartners für das Training von Sprachmodellen zu einer Vertragsverletzung, kann dies unter Umständen auch ein unlauteres und damit rechtswidriges Verhalten darstellen. Dies ist dann der Fall, wenn ein anvertrautes Arbeitsergebnis unbefugt verwertet wird (Art. 5 Bst. a und b UWG, siehe dazu oben im Kapitel VII.B). Das Lauterkeitsrecht verbietet ferner die Verwertung von Geschäftsgeheimnissen, die ein Dritter unter Verletzung seiner eigenen vertraglichen Pflichten weitergegeben hat (Art. 6 UWG). Beides untersagt auch das Strafgesetzbuch, das sowohl den Verrat von Geschäftsgeheimnissen als auch das Ausnutzen des Verrats eines Dritten unter Strafe stellt (Art. 162 StGB).

[99] Welche dieser Konsequenzen wahrscheinlich oder möglich ist, muss im Einzelfall beurteilt werden.

[100] Die vorstehenden Ausführungen gelten auch für **Nutzungsbedingungen** von Online-Plattformen, allerdings nur insofern, als die Nutzungsbedingungen verbindlich vereinbart worden sind. Dies ist nach Schweizer Recht nicht schon der Fall, wenn sie auf der Online-Plattform publiziert sind. Damit sie zu einem verbindlichen Vertragsinhalt werden, ist ein Vertragsschluss nötig, wie z.B. im Rahmen einer Benutzerkontenregistrierung. Wer auf die Inhalte einer Online-Plattform ohne Registrierung oder sonst einen Vertragsschluss zugreifen kann, ist an die dort allenfalls publizierten Nutzungsbedingungen nicht gebunden.[170] Es gelten in diesem Fall nur – aber immerhin – die gesetzlichen Rahmenbedingungen für die Nutzung solcher Inhalte. Sind also die Inhalte – wie oft der Fall – urheberrechtlich geschützt, muss geprüft werden, welche der oben zitierten Grundlagen eine Verwendung für Trainingszwecke trotzdem erlaubt. Klar ist immerhin, dass dort, wo der Betreiber einer Website sich in Nutzungsbedingungen oder anderen Hinweisen gegen eine solche Verwendung ausspricht, wohl nicht von einer impliziten Zustimmung ausgegangen werden kann, jedenfalls wo davon auszugehen ist, dass er selbst auch der Rechteinhaber ist oder für diesen sprechen kann.[171]

C. Fazit

[101] Die in der Praxis schwierigste Hürde für das Training von grossen Sprachmodellen ist die vertragliche Regelung des Zugangs zu und der Verwendung von Datenmaterial, das für das Training der Modelle benutzt werden soll. Schwierig ist die Hürde deshalb, weil in jedem Einzelfall

170 Siehe zum Ganzen etwa: Martin Eckert, Digitale Daten als Wirtschaftsgut: Besitz und Eigentum an digitalen Daten, in: SJZ 112/2016, S. 265 ff., S. 272.

171 Siehe dazu auch Kapitel VIII.

geprüft werden muss, was erlaubt ist, und weil der Vertrag oft keine ausdrückliche Regelung enthält und damit ausgelegt oder sogar über den Weg der Lückenfüllung ergänzt werden muss.

X. Zusammenfassung

[102] Insgesamt zeigen die Ausführungen, dass es grundsätzlich möglich ist, ein grosses Sprachmodell unter Einhaltung des Schweizer Rechts mit Daten aus öffentlichen Quellen auch dann zu trainieren, wenn von den Rechteinhabern und den betroffenen Personen keine Zustimmung vorliegt. Die meisten Risiken in Bezug auf die Rechtmässigkeit bringt allerdings die Memorisierung von Trainingsinhalten mit sich, soweit sie nicht verhindert werden kann:

- Kommt es zu einer wortwörtlichen oder inhaltlichen Memorisierung von Werken Dritter, greifen im **Urheberrecht** gewisse der üblicherweise angeführten rechtlichen Grundlagen wie die Forschungsschranke oder der betriebliche Eigengebrauch, die eine Verwendung selbst öffentlicher Trainingsinhalte im Sinne einer Ausnahme erlauben, unter Umständen nicht mehr. Hierbei ist die wortwörtliche Memorisierung (z.B. ein Slogan) problematischer als die bloss inhaltliche. Doch auch in diesen Fällen erscheint die Rechtmässigkeit des Trainings eines Sprachmodells begründbar: Die Aufnahme von solchem Sprach- und Sachwissen im Modell ist unseres Erachtens mit guten Gründen nicht eine urheberrechtlich relevante Handlung, weil das Originalwerk im Modell selbst im Falle einer Memorisierung entweder gar nicht mehr als urheberrechtlich relevantes Werk vorkommt, oder aber zumindest die individuellen Züge des Originalwerks bei dem, was im Modell im Ergebnis im Falle einer Memorisierung vorliegt, angesichts aller anderen, im Modell abgebildeten Informationen völlig in den Hintergrund treten, und die Modellinformationen den nötigen inneren Abstand zu den Originalwerken aufweisen, sodass die entsprechenden Rechte nicht mehr verletzt werden. Zur Urheberrechtsverletzung kann es nur noch kommen, falls und wenn mit dem passenden Prompt ein Output generiert wird, der in den Schutzbereich eines Originalwerks fällt. Hierfür ist aber nicht derjenige verantwortlich, der das Modell trainiert, sondern dessen Verwender; die Nutzung eines Modells und dessen Output ist eine vom Training zu trennende Handlung. Für das **Lauterkeitsrecht** gelten ähnliche Überlegungen, jedenfalls im Hinblick auf die Bestimmungen zum Schutz vor der Übernahme fremder Arbeitsergebnisse.
- Für das **Datenschutzrecht** spielt nicht nur die wortgetreue Memorisierung eine Rolle, sondern ebenso die sinngemässe Memorisierung, bei welcher zwar nicht die Trainingsdaten selbst memorisiert werden, sondern die darin enthaltenen Informationen. Auf diese Weise können Personendaten in ein Sprachmodell gelangen. Diese Memorisierung muss allerdings nicht vollständig verhindert werden. Werden öffentliche Informationen von Personen des öffentlichen Lebens memorisiert, wird dies regelmässig gerechtfertigt bzw. verhältnismässig sein.

[103] Weiter ist darauf zu achten, dass keine etwaigen im Zusammenhang mit der Beschaffung oder Nutzung der Trainingsdaten vereinbarten **Vertragsbedingungen** verletzt werden und dass nichts verwendet wird, was aus einer Verletzung einer Geheimhaltungsvereinbarung stammt. Als solche Vertragsbedingungen gelten auch die (im Rahmen einer Registrierung oder dergleichen vereinbarten) Nutzungsbedingungen von Online-Plattformen, auch wenn diese öffentlich

zugänglich sind. Rein einseitig geäusserte Crawler-Verbote sind hingegen für Dritte nicht verbindlich, d.h. sie entfalten keine über die gesetzlichen Regelungen hinausgehende Verbotswirkungen.

[104] Schliesslich weisen wir darauf hin, dass die hier dargelegten Überlegungen, Interpretationen und Herleitungen aus rechtlicher Sicht weitestgehend Neuland darstellen. Mit dem Restrisiko, dass die Gerichte es anders sehen, muss für das Training von grossen Sprachmodellen bis zu entsprechenden Präzedenzfällen, die es womöglich nie oder erst in Jahren geben wird, gelebt werden. Wir gehen dabei davon aus, dass Unternehmen bzw. Stellen, deren Geschäft erheblich vom Schutz ihrer Inhalte abhängen (z.B. Medienunternehmen), auch bei unklarer Rechtslage versuchen werden, ihre Interessen auf dem Rechtsweg durchzusetzen. Auch mit politischen Vorstössen insbesondere zur Anpassung des Urheberrechts ist zu rechnen, so etwa bei der Frage, ob in der Schweiz ein dem EU-Recht nachempfundenes «Opt-out»-Recht für die Forschungsschranke eingeführt werden soll. Andere werden darauf drängen, den Technologie- und Forschungsstandort Schweiz zu stärken, indem hier noch bessere Rahmenbedingungen als in der EU geschaffen werden, als sie unseres Erachtens ohnehin schon bestehen. Das Schweizer Recht ermöglicht nach unserer Einschätzung jedenfalls mehr Möglichkeiten für die Entwicklung von künstlicher Intelligenz, als sie in der EU derzeit bestehen.

[105] Folgende Darstellung illustriert die Schritte zur Prüfung einer neuen Quelle für das Training eines grossen Sprachmodells. Es sind nicht alle Voraussetzungen bzw. Annahmen abgedeckt, sondern jene, die nach unserer Ansicht besonders ins Gewicht fallen (Abbildung 3, vgl. Anhang für volle Grösse):

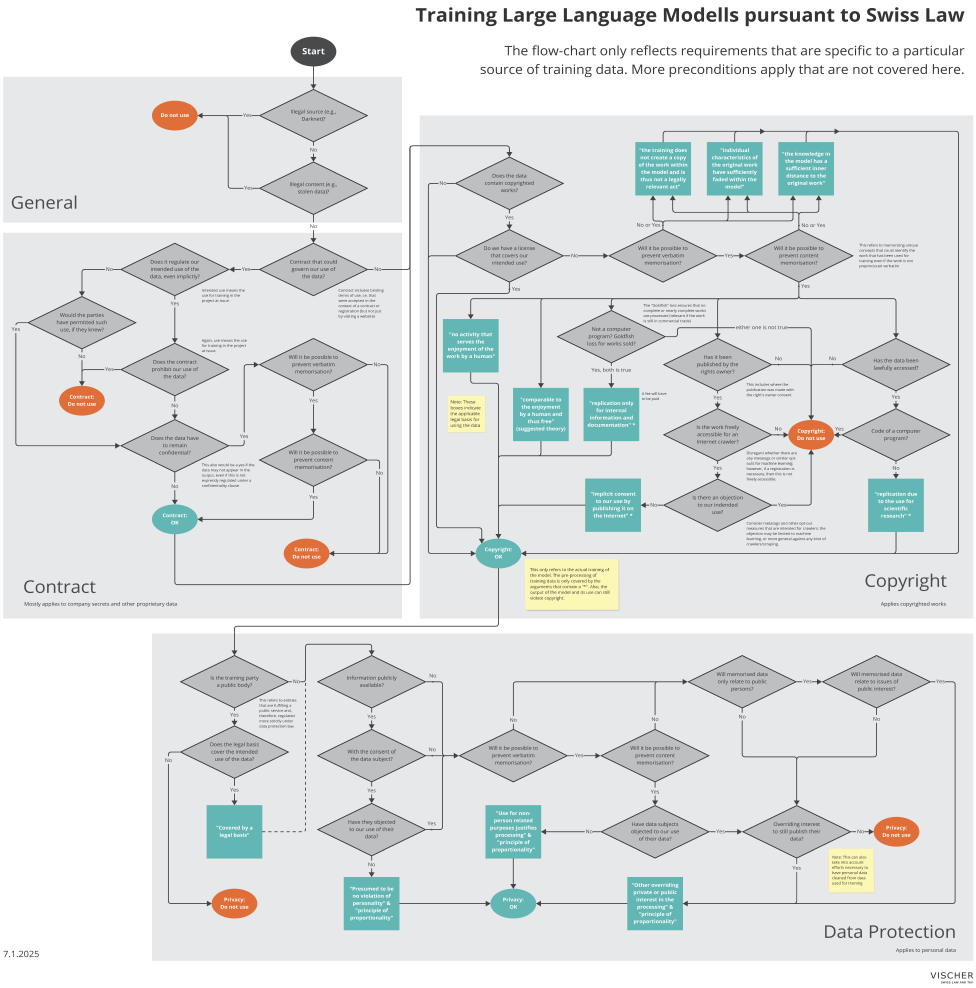

Abbildung 3: Die Prüfschritte beim Trainieren eines grossen Sprachmodells unter Schweizer Recht (verkleinerte Darstellung)

PDF der Abbildung 3

David Rosenthal ist Partner der Kanzlei VISCHER und Lehrbeauftragter der Universität Basel und ETH Zürich.

Livio Veraldi ist Junior Associate bei VISCHER.

Die Autoren danken Nicole Ritter, Giulia Odermatt und Valeria Locher (alle VISCHER) sowie Dr. Imanol Schlag (ETH AI Center), Prof. Martin Jaggi (EPFL) und Prof. Dr. Florent Thouvenin (Universität Zürich) für die geschätzte Unterstützung in der Entstehung dieses Beitrags. Sie danken ferner den verschiedenen technischen und rechtlichen Fachleuten für ihre wertvollen Anregungen zu diesem Beitrag.

Dieser Beitrag basiert auf einem Gutachten der Autoren für das ETH AI Center und stellt ihre persönliche Ansicht und nicht notwendigerweise jene des ETH AI Center und der ETH Zürich dar.