EXKLUSIVE LESEPROBE
technology research hub

Inkl. EU Artificial Intelligence Act

Niklas Mühleis • Nick Akinci

RECHTSLEITFADEN KI IM UNTERNEHMEN

- Datenschutz, Haftungsrisiken, Urheberrecht, Compliance
- Einsatz von ChatGPT, Midjourney und Co.
- Einführung von KI im Unternehmen, Content Creation, Datenanalyse, Softwareentwicklung, Sprachassistenten, Human Resources

Rheinwerk Computing

Kapitel 3
Individuallösung: Finetuning und Training eigener Modelle

In diesem Kapitel gehen wir näher auf die Individualisierung von KI-Modellen und -diensten sowie die damit zusammenhängenden spezifischen Rechtsprobleme ein.

Im zweiten Kapitel haben wir den Fokus auf bestehende KI-Modelle und -Systeme gelegt, die in Ihrem Unternehmen eingesetzt werden können. Das vorliegende Kapitel widmet sich demgegenüber KI, die speziell an die Bedürfnisse Ihres Unternehmens und an die vorhandenen Prozesse angepasst wird. Für solche individuellen Anpassungen gibt es verschiedene Möglichkeiten und Vorgehensweisen, die wir Ihnen zu Anfang dieses Kapitels aufzeigen werden. Zudem erfahren Sie, was die konkreten Vorteile individueller Lösungen sind und für wen sich der Aufwand lohnt.

Bei KI, die z. B. durch Training oder Finetuning angepasst wird, ergeben sich zudem gegenüber Standardlösungen einige spezifische rechtliche Problemstellungen, die es zu beachten gilt. Denn schon das Training selbst birgt einige rechtliche Stolperdrähte.

3.1 Warum Sie eigene Modelle betreiben sollten!

Die Welt der künstlichen Intelligenz (KI) entwickelt sich fortwährend mit hoher Geschwindigkeit weiter, und Unternehmen aller Größenordnungen experimentieren mit der Einführung von KI-Modellen, um ihre Prozesse zu verbessern, Kundenerfahrungen zu personalisieren und neue Erkenntnisse aus ihren Daten zu gewinnen. Doch während die Nutzung von cloud-basierten KI-Services ein sinnvoller Anfang sein kann, gibt es gute Gründe, warum ein Unternehmen es in Betracht ziehen sollte, eigene oder bestehende Modelle direkt selbst zu betreiben.

Ob Sie dabei technische Argumente wie Anpassbarkeit und Spezialisierung abwägen oder den Datenschutz und die Datensouveränität als entscheidendes Kriterium betrachten – dieses Kapitel wird Ihnen helfen, fundierte Entscheidungen darüber zu treffen, ob und wie Sie KI-Modelle in Ihrem Unternehmen einsetzen können.

3.1.1 Argumente für eigene Lösungen

Ein wesentliches Argument für den Betrieb eigener Modelle ist die Kontrolle über den Datenschutz und die Datensicherheit. Wenn sensible Unternehmensdaten ins Spiel kommen, ist es in der Regel nicht wünschenswert, diese an externe Cloud-Dienstleister zu übermitteln. Durch lokale Installation und Verwendung von KI-Modellen behalten Sie die vollständige Kontrolle über Ihre Daten und schützen sie vor unbefugtem Zugriff.

Darüber hinaus bietet das selbstständige Betreiben von Modellen ein großes Maß an Flexibilität und Anpassbarkeit. Sie können dabei Modelle an ihre spezifischen Anforderungen und Besonderheiten anpassen, statt sich auf Standardangebote der großen Anbieter verwiesen zu sehen. Dies ist besonders dann von Vorteil, wenn branchenspezifische Eigenheiten oder seltene Sprachen eine Rolle spielen, für die kommerzielle Cloud-Dienste oft keine optimierte Lösung bieten.

Weiterhin dreht es sich um die Thematik der Zensur bzw. Einschränkungen, die bei kommerziellen Anbietern greifen könnten. Bei der Nutzung eigener Modelle sind Unternehmen nicht den möglicherweise willkürlichen oder unpassenden Zensurvorgaben von Dienstleistern unterworfen. Dies ist insbesondere relevant für kreative Industrien, etwa beim Verfassen von Texten oder Erstellen von Kunstwerken, wo Freiheit im Ausdruck entscheidend ist.

Ein weiterer gewichtiger Aspekt ist die Kostenkontrolle. Es gibt Szenarien, in denen die lokal betriebenen KI-Modelle langfristig kostengünstiger sein können, vor allem wenn sie in großem Umfang genutzt werden. Während Cloud-Dienste oft auf einer Pay-per-use-Basis abgerechnet werden, können eigene Modelle nach der Anfangsinvestition ohne zusätzliche laufende Kosten pro User genutzt werden. Zu beachten ist auch die Unabhängigkeit von etwaigen strategischen Veränderungen der Anbieter, die Dienste ohne viel Vorlaufzeit einschränken, kostenpflichtig erweitern oder gar einstellen können. Wer eigene Modelle betreibt, ist vor solchen Unwägbarkeiten geschützt.

Obwohl es auf den ersten Blick so scheinen mag, als wäre das Selbstbetreiben von KI-Modellen nur etwas für Unternehmen mit umfassenden technischen Ressourcen, wird die Einstiegshürde auch für kleinere Organisationen fortlaufend geringer. Die Verfügbarkeit von Tools und Ressourcen zum Betrieb lokal verwendeter Modelle ist in den letzten Jahren deutlich gewachsen, und auch mit bescheideneren Mitteln lassen sich mittlerweile KI-Modelle betreiben. Im Vergleich zu teuren KI-Cloud-Diensten können lokale Modelle auf Hardware betrieben werden, die bereits vorhanden ist oder die bei geringen Kosten einen guten Einstieg ermöglicht. Beispielsweise bieten mittelpreisige Grafikkarten bereits ausreichend Leistung, um kleinere Modelle in einem bestimmten Rahmen zu betreiben.

Hinzu kommt eine Welle von Innovationen im Bereich der Modell-Optimierung, die Modelle effizienter machen und somit auch auf weniger leistungsstarker Hardware betriebsfähig werden. In einer Zeit, in der die Anforderungen an Grafikspeicher durch Spiele und die Nutzung größerer Modelle steigen, wird auch die Verfügbarkeit von Hardware mit ausreichend VRAM zunehmen.

Lokal betriebene KI-Modelle bieten auch eine höhere Zuverlässigkeit im Vergleich zum Cloud-Service, der unter Umständen durch Überlastungen temporär nicht verfügbar sein kann. Zu guter Letzt ist es oftmals auch ein Aspekt der Unabhängigkeit und des Empowerments. Unternehmen und Entwickler möchten nicht vollständig von einem Dienstleister abhängig sein, dessen Geschäftspolitik sich ändern und damit die eigenen Prozesse beeinträchtigen könnte. Ein selbst betriebenes Modell kann vor solchen Unabwägbarkeiten schützen und damit Souveränität bieten.

3.1.2 Individualisierungsmöglichkeiten

Die Einführung und das Betreiben eigener KI-Modelle im Unternehmen können auf unterschiedlichen Ebenen erfolgen. Die Bandbreite reicht von Lösungen für Nichttechniker, die eine einfache, nutzerfreundliche grafische Benutzeroberfläche (GUI) bevorzugen, bis hin zu umfassenderen Plattformen für Techniker/Entwickler, die eine tiefer gehende Steuerung und Anpassungsfähigkeit der Modelle ermöglichen.

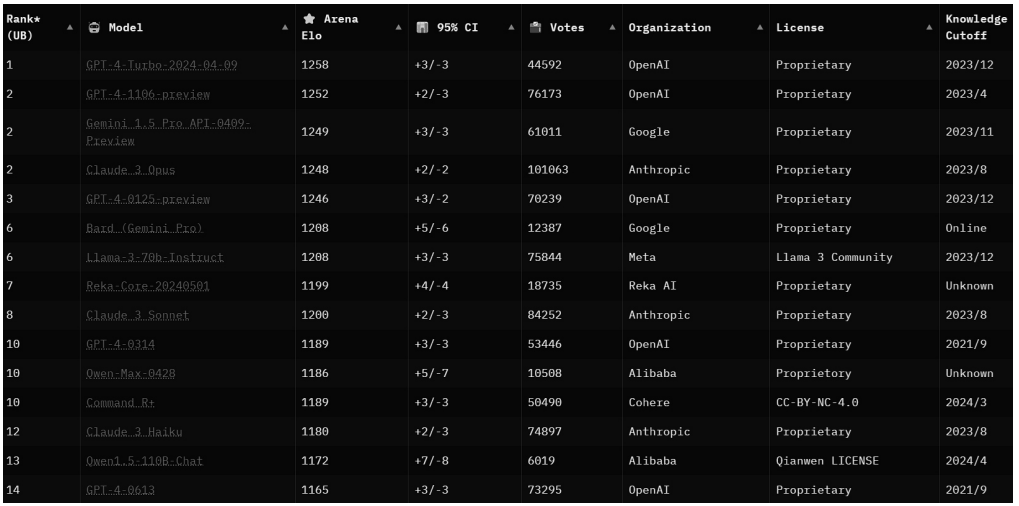
Für Nichttechniker stehen Tools wie LM-Studio oder jan (Open Source) zur Verfügung, die es ermöglichen, große Sprachmodelle (LLMs) direkt auf dem eigenen performanten Rechner vollständig offline zu betreiben. Beide bieten eine intuitive Chat-Benutzeroberfläche und eine Schnittstelle, die das Verhalten der OpenAI-API emuliert. Modelle können nahtlos von Hugging Face heruntergeladen und betrieben werden.

Webseiten wie *chat.lmsys.org* erlauben sogar den kostenlosen Test und Vergleich direkt im Browser ganz ohne Set-up.

Für Techniker oder fortgeschrittene Anwender bietet das quelloffene Projekt FastChat eine offene Plattform zum Betreiben und Bewerten von Chatbots auf Basis großer Sprachmodelle. Es bietet eine leistungsfähige Infrastruktur mit einem verteilten Multi-Modell-System, einer Web-Benutzeroberfläche und APIs, die ebenfalls mit OpenAI kompatibel ist.

Angesichts der Geschwindigkeit, mit der sich KI-Entwicklungen vollziehen, ist es ratsam, sich regelmäßig über Fortschritte und Neuerscheinungen zu informieren. Bevor

Sie diesen Schritt gehen, sollten Sie daher sorgsam planen und Zeit investieren, um die Aktualität und Relevanz von Tools und Methoden zu prüfen, bevor Sie sich für den Betrieb eigener Modelle entscheiden.



Rank* (UB)	Model	Arena Elo	95% CI	Votes	Organization	License	Knowledge Cutoff
1	GPT-4-Turbo-2024-04-09	1258	+3/-3	44592	OpenAI	Proprietary	2023/12
2	GPT-4-1106-preview	1252	+2/-3	76173	OpenAI	Proprietary	2023/4
2	Gemini 1.5 Pro API-0409-Preview	1249	+3/-3	61011	Google	Proprietary	2023/11
2	Claude 3 Opus	1248	+2/-2	101063	Anthropic	Proprietary	2023/8
3	GPT-4-0125-preview	1246	+3/-2	70239	OpenAI	Proprietary	2023/12
6	Bard (Gemini Pro)	1208	+5/-6	12387	Google	Proprietary	Online
6	Llama 3 70b Instruct	1208	+3/-3	75844	Meta	Llama 3 Community	2023/12
7	Reka Core-20240501	1199	+4/-4	18735	Reka AI	Proprietary	Unknown
8	Claude 3 Sonnet	1200	+2/-3	84252	Anthropic	Proprietary	2023/8
10	GPT-4-0314	1189	+3/-3	53446	OpenAI	Proprietary	2021/9
10	Qwen-Max-0428	1186	+5/-7	10508	Alibaba	Proprietory	Unknown
10	Command R+	1189	+3/-3	50490	Cohere	CC-BY-NC-4.0	2024/3
12	Claude 3 Haiku	1180	+2/-3	74897	Anthropic	Proprietary	2023/8
13	Qwen1.5-110B-Chat	1172	+7/-8	6019	Alibaba	Qianwen LICENSE	2024/4
14	GPT-4-0613	1165	+3/-3	73295	OpenAI	Proprietary	2021/9

Abbildung 3.1 Die Oberfläche von chat.lmsys.org

Bestehende Modelle ergänzen

Das Anpassen von KI-Modellen an die Bedürfnisse Ihres Unternehmens ist ein Schlüsselaspekt beim Betreiben eigener KI-Systeme. Hierbei bietet das Prompt Engineering eine leichte und schnelle Möglichkeit, die Antworten der KI zu steuern. Für eine ausführliche Diskussion über die Vorstufen des Finetuning, einschließlich Prompt Engineering und Retrieval Augmented Generation (RAG), verweisen wir auf Abschnitt 2.1.2.

Die Entscheidung, wie KI-Modelle in Ihrem Unternehmen betrieben werden sollen, hängt von vielen Faktoren ab, darunter technisches Know-how, verfügbare Ressourcen und spezifische Unternehmensziele. Sowohl nichttechnische als auch technisch versierte Mitarbeiterinnen und Mitarbeiter finden jedoch zunehmend Zugang zu KI-Tools und Plattformen, die den breiten Einzug von KI in die Unternehmenswelt erleichtern und fördern.

Frameworks wie H2O LLM Studio erlauben ein ähnliches Low-Code-Finetuning-Erlebnis wie OpenAI, siehe Abschnitt 2.1.2.

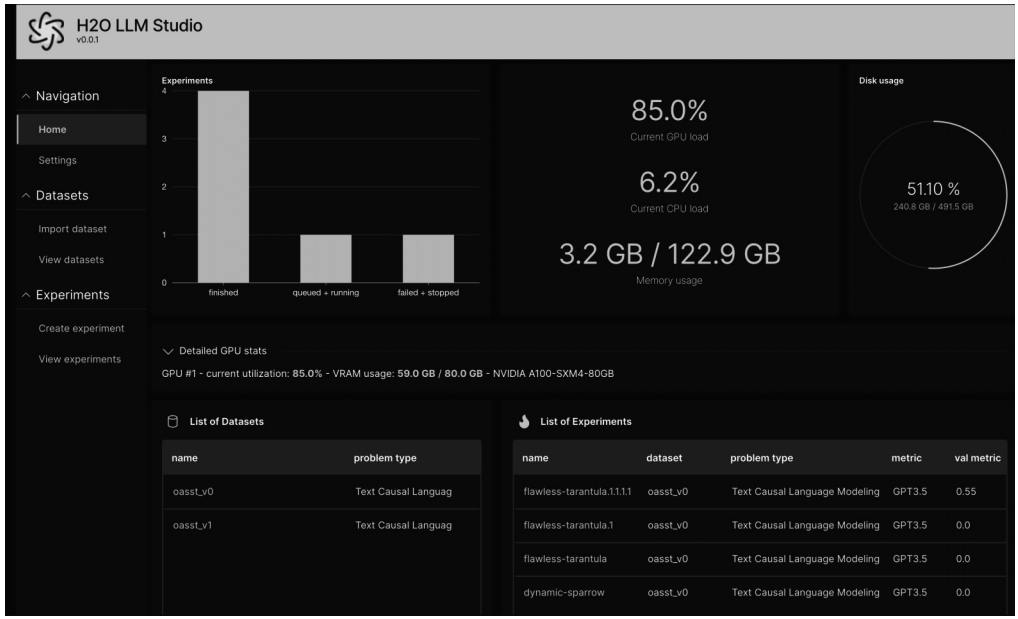


Abbildung 3.2 Die Oberfläche von H2O LLM Studio zum Finetuning lokaler Modelle

3.1.3 Neues eigenes Modell durch Training

In einer Zeit, in der künstliche Intelligenz (KI) viele unserer täglichen Prozesse zu transformieren vermag, mag es naheliegend erscheinen, eigene KI-Modelle von Grund auf zu trainieren. Doch ein solcher Ansatz ruft eine Reihe wichtiger Überlegungen hervor, von der Datenintegration bis hin zu den technischen und finanziellen Anforderungen. Die Entwicklung eigener KI-Modelle kann dann besonders sinnvoll sein, wenn Unternehmen spezifische, maßgeschneiderte Lösungen für eine einzigartige Problematik benötigen.

Beispielhaft angeführt werden kann hierfür der Finanzsektor: Hier können Banken oder Versicherungen von selbst entwickelten KI-Modellen profitieren, die Betrugserscheinungen (Fraud Detection) identifizieren. Angesichts des Umstandes, dass betrügerische Aktivitäten ständig wechselnde, ausgeklügelte Muster annehmen, kann ein maßgeschneidertes Modell, das auf den spezifischen Transaktionsdaten und Kundenprofilen des Unternehmens basiert, effektiver agieren als generische, vorgefertigte Lösungen.

Gleichermaßen kann im Bereich der Netzwerksicherheit das Training eines individuellen KI-Modells Unternehmen dazu befähigen, sich besser gegen zunehmend intel-

ligentere Cyberangriffe zu schützen. Ein auf den spezifischen Netzwerkverkehr und die Schutzbedürfnisse des Unternehmens abgestimmtes Modell kann dabei bösartige Aktivitäten präziser erkennen als standardisierte Lösungen.

Bevor Sie sich nun direkt dem Training eines eigenen KI-Modells widmen, sollten Sie daher zuerst sichergehen, dass Ihr Anwendungsfall dies wirklich erfordert und ob nicht Vorstufen wie Prompt-Engineering, Finetuning oder RAG hierzu ausreichen.

Die Bedeutung der Datenintegration

Die Grundlage jeden KI-Trainings ist der Datensatz. Die Beschaffung, Bereinigung und Strukturierung von Daten kann sich als eine Herkulesaufgabe herausstellen. Die Bewegung großer Datenmengen, insbesondere in die Cloud, kann nicht nur technisch komplex sein, sondern auch schnell zu hohen Kosten führen. Zum Training von GPT-2 beispielsweise wurden Daten aus über 7.000 selbst veröffentlichten Büchern und einer Sammlung von 8 Millionen Webseiten verwendet, ganz zu schweigen von GPT-3.5 und 4.

Eine herausfordernde Aufgabe, die nicht nur die Bereitstellung der Daten, sondern auch die Skizzierung einer klaren Datenstruktur erfordert, bevor mit dem eigentlichen Training begonnen werden kann.

Voraussetzungen für das Training eigener Modelle

Die technischen Voraussetzungen für das Training eines eigenen KI-Modells sind nicht zu unterschätzen. Umfangreiche Rechenkapazitäten sind erforderlich, um die massiven Berechnungen durchzuführen, die für das Training von Modellen erforderlich sind.

Ein Vergleich: Während für die Entwicklung von GPT-2 rund 50.000 Dollar anfielen, beliefen sich die Kosten für GPT-3 auf mehr als 4 Millionen Dollar, bei GPT-4 waren es anscheinend 100 Millionen Dollar. Unternehmen müssen letztendlich entscheiden, ob die Investition in die Entwicklung eigener Modelle den Aufwand wert ist. Die Erfahrungen von OpenAI zeigen, dass selbst ein Investment, das im Vergleich zu späteren Modellen moderat ausfiel, das Potenzial hatte, den Bereich der natürlichen Sprachverarbeitung zu revolutionieren. Das Training eigener KI-Modelle ist ein vielschichtiges und herausforderndes Projekt und kann in bestimmten Fällen durchaus sinnvoll sein.

Jedoch sollte die Entscheidung sorgfältig abgewogen werden, da es sowohl erhebliche finanzielle als auch technologische Investitionen erfordert.

3.1.4 Cloudlösungen

Die Implementierung von Künstlicher Intelligenz in Unternehmensprozesse ist ohne die angemessene technische Infrastruktur nicht denkbar. Cloud-Instanzen spielen dabei eine zentrale Rolle. Sie bieten die Rechenkraft und Flexibilität, die für das Trainieren und Betreiben von KI-Modellen erforderlich sind.

VM-Instanzen mit GPUs bieten gegenüber traditionellen CPU-Instanzen entscheidende Vorteile, indem sie speziell für rechenintensive Aufgaben des maschinellen Lernens und KI-Modelltrainings konzipiert wurden. Durch den Einsatz dieser GPU-Instanzen können Unternehmen die Effizienz ihrer KI-Operationen signifikant steigern und gleichzeitig ihre lokale Hardware von aufwendigen Rechenprozessen entlasten.

Mit Hybrid-, Private- oder Public-Cloud-Architekturen erhalten Sie unterschiedliche Stufen an Sicherheit, Skalierbarkeit und Kostenkontrolle – je nach Unternehmensbedürfnis und der gewünschten Integrationsstufe in Ihre IT-Infrastruktur.

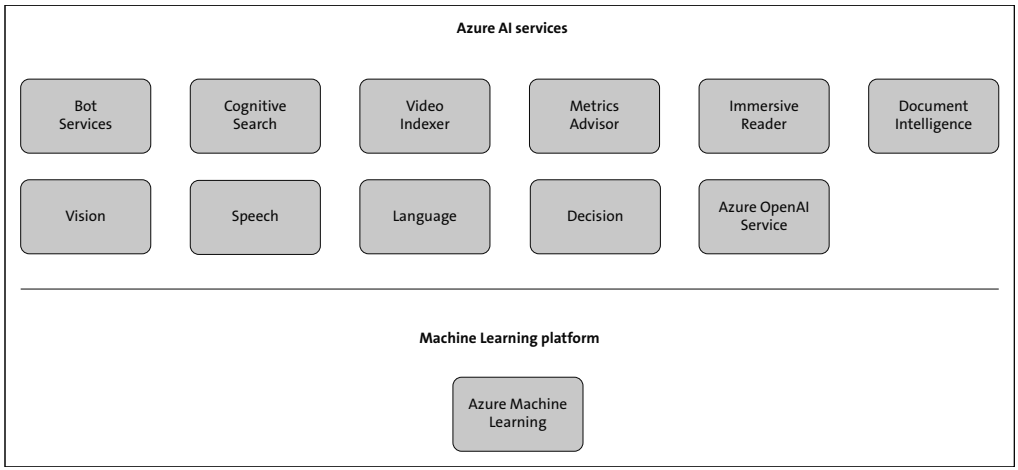


Abbildung 3.3 Überblick über die KI-Services von Microsoft Azure

Große Cloud-Anbieter wie Google Cloud stehen mit flexiblen und skalierbaren GPU-Services zur Verfügung, die auf unterschiedliche Anwendungsfälle und Unternehmensgrößen zugeschnitten sind.

Die Services bieten eine Auswahl von Instanz-Arten, die oft nach Zeit abgerechnet werden und je nach Bedarf hoch- oder runterskaliert werden können. Die Auswahl der Cloud-Infrastruktur kann über den Erfolg Ihrer KI-Projekte entscheiden. Eine sorgfältige Prüfung der Leistung, Flexibilität und Kosten der verschiedenen Cloud-

Services ist daher unerlässlich. Betrachten Sie die spezifischen Angebote der großen Anbieter und prüfen Sie, welcher Service die optimale Basis für Ihre KI-Initiativen bildet. Gleichzeitig bietet die Cloudlandschaft heute Alternativen für diejenigen, die eine auf bestimmte Zeiträume beschränkte oder finanziell sparsame Lösung suchen, ohne dabei Abstriche bei der Leistung machen zu müssen.

3.1.5 Eigene Hardware nutzen

Die Entscheidung, KI-Modelle lokal zu betreiben, ist ein bedeutsamer Schritt in Richtung vollständiger Autonomie und Kontrolle über Ihre maschinellen Lernumgebungen. Indem Sie sich von Cloud-Diensten unabhängig machen, erlangen Sie die Freiheit, die Modelle Ihren genauen Bedürfnissen anzupassen und dabei sensible Daten intern zu halten.

Um Ihnen einen umfassenden Einblick zu geben, wird in diesem Kapitel erörtert, wie Sie KI-Modelle auf eigene Faust betreiben und trainieren können.

Die wesentlichen Hardware-Komponenten umfassen dabei nicht nur eine ausreichende Prozessorleistung (CPU) und genügend Arbeitsspeicher (RAM), sondern auch eine leistungsstarke Grafikprozessoreinheit (GPU), die insbesondere für das maschinelle Lernen optimiert ist.

Hierbei ist die Menge des Videospeichers (VRAM) pro GPU von besonderer Bedeutung, da sie bestimmt, wie komplexe Modelle oder umfangreiche Datenmengen verarbeitet werden können.

Ressourcenverbrauch KI

Das vollständige Training eines KI-Modells ist am ressourcenintensivsten.

Finetuning erfordert im Vergleich zum bloßen Betreiben meist eine stärkere Hardware. Hier muss das Modell geringfügig angepasst werden, was wiederum zusätzliche Rechenkapazität erfordert.

Das Betreiben vortrainierter KI-Modelle erfordert eine Hardwareumgebung, die sich nicht ausschließlich durch hohe Rechenleistung, sondern insbesondere durch ausreichenden Videospeicher (VRAM) auszeichnet.

Der VRAM ist häufig der begrenzende Faktor bei der Auswahl der GPU für KI-Modelle, da diese tendenziell mehr von der Speicherkapazität als von der reinen Rechenleistung abhängen. Bei begrenzter GPU-Anzahl kann durch intelligentes Batching von Anfragen ein effizienter Parallelbetrieb ermöglicht werden, ohne pro Anfrage eine gesamte GPU zu blockieren.

Fertige Modelle finden sich inzwischen auf Plattformen wie Hugging Faces Modellhub, der eine umfassende Bibliothek zur Verfügung stellt. Die Sammlung reicht von Modellen kleinerer Größenordnungen bis hin zu sehr großen wie zum Beispiel einem 120B-Sprachmodell. Das »B« steht für Milliarden (Billions) und bezieht sich auf die Gesamtanzahl der Parameter, aus denen das KI-Modell besteht.

Parameter sind im Grunde die Elemente eines KI-Modells, die aus dem Trainingsprozess gelernt werden und entscheiden, wie das Modell Daten interpretiert und Antworten generiert. Je mehr Parameter ein Modell hat, desto umfassender ist seine Fähigkeit, Wissen zu speichern und komplexe Muster zu erkennen. Ein KI-Modell mit 13 Milliarden Parametern kann also eine enorme Menge an Informationen verarbeiten und ist somit in der Lage, subtile Nuancen in der Sprache zu erfassen. Das ermöglicht es dem Modell, Antworten zu generieren, die menschlicher Kommunikation näherkommen, und kann in komplexen Aufgaben wie der Sprachübersetzung, dem Textverständnis oder der automatischen Beantwortung von Fragen eingesetzt werden.

Die große Anzahl an Parametern eines solchen Modells erfordert jedoch entsprechend dimensionierte Hardware-Ressourcen, insbesondere hinsichtlich des VRAM, um die umfangreichen Berechnungen effektiv durchführen zu können.

Hintergrundwissen

Erforderliche technische Ressourcen für den lokalen KI-Betrieb (geschätzt):

- 7B-Modell: mindestens 13 GB VRAM, 1 × NVIDIA 4070 Ti SUPER (16 GB)
- 13B-Modell: mindestens 26 GB VRAM, 1 × NVDIA RTX A6000 (48 GB)
- 30B-Modell: mindestens 65 GB VRAM, 1 × NVDIA H100 (80 GB)
- 65B Modell: Mindestens 131 GB VRAM, 2x NVDIA A100(80 GB)
- 175B-Modell (~GPT-3): mindestens 350 GB VRAM, 5 × NVDIA A100 (80 GB)

Darüber hinaus spielen Hyperparameter eine wichtige Rolle, da diese bestimmen, ob ein Modell auf einer existierenden GPU betrieben werden kann. Durch Anpassungen, wie beispielsweise Quantisierung, Kontextlänge und effiziente Batch-Verarbeitungen, werden Modelle auch für Hardwareplattformen zugänglich gemacht, deren Spezifikationen geringer sind als die oben genannten Anforderungen. Somit lassen sich auch mit eingeschränkter VRAM-Ausstattung Modelle realisieren, die ursprünglich nicht dafür vorgesehen waren. Modelle lassen sich auch über mehrere GPUs verteilen, sodass man 2 GPUs mit jeweils 16 GB zusammenschalten kann, um ein 13B-Modell zu betreiben, ohne die Hyperparameter zu verändern.

Hintergrundwissen: Was ist Quantisierung?

Wenn wir über KI-Modelle sprechen, denken Sie vielleicht an eine immense Sammlung von Zahlen, die jede Verbindung und Gewichtung innerhalb des Modells darstellen. Normalerweise werden diese Zahlen mit hoher Präzision gespeichert, was bedeutet, dass sie viel Speicherplatz einnehmen.

Quantisierung reduziert diese Präzision gezielt, ähnlich dem Vorgang des Komprimierens einer Musikdatei, um Speicherplatz zu sparen. Dabei werden die Zahlen einer umfangreichen Datenbank in einer kompakteren Form ausgedrückt, indem zum Beispiel statt 32 Bits nur 16 oder sogar 8 Bits verwendet werden. Dies macht das Modell kleiner und leichter und somit weniger anspruchsvoll in Bezug auf die Speicher- und Verarbeitungskapazität einer GPU, jedoch sinkt damit auch die Qualität der Ausgaben.

Auf diese Weise ist es möglich, leistungsfähige KI-Modelle auf Hardware mit begrenztem VRAM zu betreiben, was besonders für Organisationen von Bedeutung ist, die ihre Modelle lokal und nicht in der Cloud betreiben möchten. Quantisierung ermöglicht es also, in einem gewissen Rahmen High-End-KI-Modelle einem breiteren Anwenderkreis zugänglich zu machen und bestehende Hardware besser zu nutzen.

3.2 Trainingsdaten und Urheberrecht

Wenn Sie ein bestehendes Modell durch Finetuning verbessern oder sogar ein eigenes Modell von Grund auf trainieren (lassen) wollen, benötigen Sie vor allem eines: Trainingsdaten. Denn insbesondere aktuelle Sprach- und Diffusionsmodelle sind gerade deshalb so leistungsfähig, weil sie mit riesigen Datenmengen trainiert wurden.

Sofern Sie nicht gerade selbst über die für das Training benötigten Daten(mengen) verfügen, liegt es nahe, sich an der wohl größten Datenquelle überhaupt zu bedienen: dem Internet. Und so dürften es auch bisher die meisten getan haben. Das Problem: Inhaber von Urheberrechten haben häufig etwas dagegen, wenn ihre im Internet verfügbaren Werke von Dritten genutzt werden.

3.2.1 Zustimmung als Ausgangspunkt

Im Grundsatz gilt im Urheberrecht die Regel, dass der Urheber allein darüber entscheiden darf, wer sein Werk in welcher Form nutzt. Das gilt insbesondere auch für Werke wie Bilder, Texte und Videos, die frei im Internet abrufbar sind. Diese Inhalte sind so gut wie immer vom Urheberrecht geschützt.

Anders formuliert: Bevor Sie Werke aus dem Internet nutzen dürfen, müssen Sie die jeweiligen Urheber um Erlaubnis bitten. Wenn Urheber anderen die Nutzung ihrer Werke erlauben, werden rechtlich gesehen sogenannte Nutzungsrechte eingeräumt. Man spricht auch von der Einräumung einer Lizenz. Teilweise beauftragen Urheber auch Dritte mit der Vergabe von Lizenzen für ihre Werke. Zum Beispiel machen dies die GEMA für Musiker oder Stockfoto-Archive für Fotografen.

Selbstverständlich ist die Einholung des Einverständnisses jedes einzelnen Urhebers bei einer großen Menge von Inhalten meist schlicht unmöglich, wenigstens aber völlig unpraktikabel. Schon allein das Ausfindigmachen der richtigen Personen wäre kaum möglich. Deshalb braucht es andere Methoden bzw. rechtliche Pfade, die Sie erwägen müssen.

3.2.2 Urheberrechtsrelevante Handlung

Bereits im Einführungskapitel haben wir Ihnen erläutert, welche Handlungen in Bezug auf urheberrechtlich geschützte Werke Relevanz haben (siehe auch Abschnitt 1.2), also urheberrechtsverletzend sein können. Dazu gehört insbesondere die Vervielfältigung, die zum Beispiel das Kopieren oder Downloaden eines Inhalts einschließt.

Nicht zu den urheberrechtlich relevanten Handlungen zählen jedoch

- das Laden von Daten in den Arbeitsspeicher (RAM),
- das Betrachten und Analysieren von im Internet abrufbaren Daten,
- das Verlinken von Inhalten im Internet.

Sofern Sie oder der von Ihnen beauftragte Dienstleister also eine technische Methode nutzen, die ohne einen Download der Trainingsdaten auskommt, können Sie sich zurücklehnen. In diesem Fall sind keine urheberrechtlichen Probleme zu erwarten.

In der Praxis sieht es jedoch häufig anders aus. Denn ein KI-Modell ist nur so gut wie die Trainingsdaten, mit denen es trainiert wird. Daher besteht die eigentliche Schwierigkeit meist nicht darin, Daten zu finden, sondern die Qualität der Daten durch sorgfältige Auswahl sicherzustellen (lesen Sie mehr hierzu in Abschnitt 1.1.4). Dazu ist es aber – insbesondere im Bereich von generativer Bild-KI – häufig erforderlich, die Daten zunächst herunterzuladen. Sobald diese gesammelt an einem (Speicher-)Ort liegen, kann durch Filter- und andere Software sichergestellt werden, dass die Daten die erforderliche Qualität aufweisen und keine unerwünschten oder sogar rechtswidrigen Inhalte beinhalten.

3.2.3 Ausnahme: Text- und Data-Mining

Der europäische Gesetzgeber hat bereits vor einigen Jahren erkannt, dass Big-Data-Szenarien und die Analyse großer Datenmengen von großer Relevanz für verschiedene Technologien sind. Er hat daher schon im Jahr 2021 eine Ausnahme für das sog. Text- und Data-Mining vorgesehen. Diese Ausnahme hat im Sommer 2021 ihren Weg in das deutsche Urheberrecht gefunden. Die neue Ausnahmeregelung ist in § 44b UrhG niedergelegt und am 07.06.2021 in Kraft getreten. Danach dürfen auch ohne Zustimmung des Urhebers Vervielfältigungen vorgenommen werden, wenn sie für Zwecke des Text- und Data-Minings erfolgen. Weiter ist Voraussetzung, dass die Inhalte »rechtmäßig zugänglich« sind. Das ist jedoch in der Regel bereits der Fall, wenn sich die Daten im freien Internet finden lassen. Zudem müssen die Datenkopien wieder gelöscht werden, wenn sie nicht mehr für das Text- und Data-Mining benötigt werden.

Was Text- und Data-Mining ist, definiert das Gesetz ebenfalls. Es handelt sich dabei um »die automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen«. Nach Auffassung vieler Juristen, einschließlich der Autoren dieses Buches, wird damit in der Regel das erfasst, was beim KI-Training passiert. Dafür spricht einerseits bereits die Tatsache, dass der Gesetzgeber – ausweislich der Gesetzesbegründung – die Änderung im Urheberrechtsgesetz zumindest auch im Hinblick auf die Entwicklung künstlicher Intelligenz vorgenommen hat. Zum anderen trifft der oben zitierte Wortlaut auf das zu, was beim Training generativer KI stattfindet. Aus den Trainingsdaten werden durch Analyse Zusammenhänge (Korrelationen) gewonnen.

3.2.4 Rückausnahme Nutzungsvorbehalt

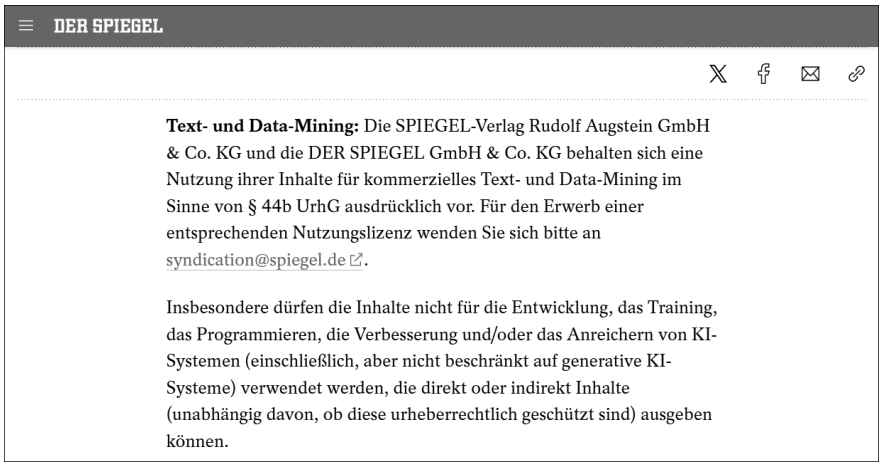
Das klingt aus Sicht desjenigen, der große Mengen Daten zum Training von KI sucht, erst einmal nach einem Grund zur Freude. Denn es scheint, dass diese Ausnahmeregelung die freie Entnahme von urheberrechtlich geschützten Werken aus dem Internet erlaubt. Das Gesetz sieht jedoch auch eine Rückausnahme vor, die diese Freude einschränkt: Urheber bzw. Rechtsinhaber haben die Möglichkeit, der Verwendung für das Text- und Data-Mining zu widersprechen. Diesen Widerspruch nennt man auch »Nutzungsvorbehalt«. Dieser muss von allen beachtet werden – lediglich Forschungseinrichtungen, die keine kommerziellen Absichten verfolgen, dürfen den Vorbehalt ignorieren.

So ein Nutzungsvorbehalt muss nach den gesetzlichen Spielregeln in »maschinenlesbarer Form« erklärt werden. Der Gesetzgeber hatte dabei vor Augen, dass große Datenmengen nicht handverlesen durch Menschen ausgewählt werden, sondern durch Software. Wie in Abschnitt 1.1.4 erläutert, kommen Scraping-Bots zum Einsatz, die das Internet nach bestimmten Daten durchforsten und diese extrahieren. Mit »maschinenlesbar« ist daher gemeint, dass der Nutzungsvorbehalt auch durch Scraping-Software erkannt werden können soll.

Hintergrundwissen: Nutzungsvorbehalt

In welcher konkreten Form nun die Erklärung erfolgen muss, dass man seine Werke nicht für das Text- und Data-Mining kopiert wissen will, wird unter Juristen derzeit heftig diskutiert. Auch hier herrscht also mal wieder Rechtsunsicherheit. Die einen halten eine entsprechende Formulierung in natürlicher Sprache – irgendwo auf der Website, welche die Inhalte enthält – für ausreichend. Andere fordern, dass der Nutzungsvorbehalt in den bereits in Abschnitt 1.1.4 erläuterten »Robots exclusion standards« hinterlegt wird. Für letztere Ansicht sprechen aus unserer Sicht die besseren Argumente, da Nutzungsvorbehalte in natürlicher Sprache viele Unwägbarkeiten mit sich bringen dürften.

Daraus folgt, dass Sie oder der in Ihrem Auftrag tätige Entwickler die eingesetzte Scraping-Software entsprechend auswählen oder programmieren sollten. Der Bot sollte zumindest keine Daten erfassen, die von Websites stammen, die in der robots.txt-Datei der Nutzung für das Text- und Data-Mining widersprechen.


DER SPIEGEL

Text- und Data-Mining: Die SPIEGEL-Verlag Rudolf Augstein GmbH & Co. KG und die DER SPIEGEL GmbH & Co. KG behalten sich eine Nutzung ihrer Inhalte für kommerzielles Text- und Data-Mining im Sinne von § 44b UrhG ausdrücklich vor. Für den Erwerb einer entsprechenden Nutzungslizenz wenden Sie sich bitte an syndication@spiegel.de.

Insbesondere dürfen die Inhalte nicht für die Entwicklung, das Training, das Programmieren, die Verbesserung und/oder das Anreichern von KI-Systemen (einschließlich, aber nicht beschränkt auf generative KI-Systeme) verwendet werden, die direkt oder indirekt Inhalte (unabhängig davon, ob diese urheberrechtlich geschützt sind) ausgeben können.

Abbildung 3.4 Auszug aus dem Impressum von spiegel-online.de – hier wird der Nutzung zu Zwecken des Text- und Data-Minings in natürlicher Sprache widersprochen.

Da im Übrigen offen ist, welche anderen Formen von Nutzungsvorbehalten beachtet werden müssen, ist jedoch Vorsicht geboten. Wer sicher vor Rechtsverletzungen sein will, sollte abwarten, bis die Rechtsprechung Klarheit schafft. Bis es hier Rechtssicherheit gibt, kann es jedoch noch dauern. Bisher sind keine Urteile bekannt geworden, die sich der Beantwortung der vielen Fragen um die 2021 eingeführte Bestimmung widmen. Erfahrungsgemäß wird es daher mindestens noch einige Jahre dauern, bis es mehr Klarheit gibt.

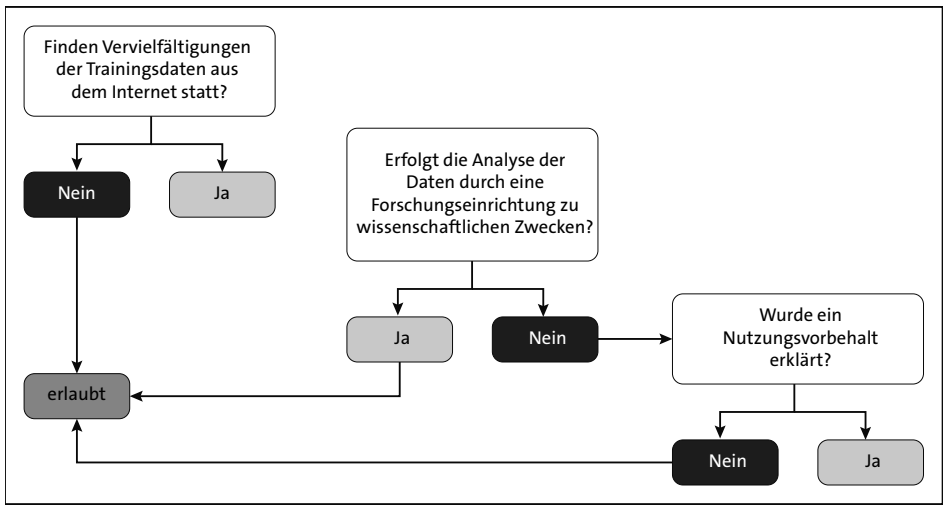


Abbildung 3.5 Wann sind die Vervielfältigungen zum Training rechtskonform?

Es existieren auch Anbieter, die riesige fertige Datensätze zum Download anbieten, z. B. Common Crawl (*https://commoncrawl.org/*). Diese Anbieter haben also bereits das Scraping übernommen. Hierbei ergibt sich jedoch das Problem, dass häufig Informationen dazu fehlen, ob beim Scraping etwaige Nutzungsvorbehalte beachtet wurden. Selbst wenn der jeweilige Anbieter dies behauptet, ist das noch keine Garantie. Bei Rechtsverletzungen können Sie trotzdem haftbar gemacht werden.

Praxistipp: Urheberrechtlich geschützte Daten verwenden

Wenn Sie Trainingsdaten in Form von urheberrechtlich geschützten Inhalten benötigen, sollten Sie vorzugsweise auf eigene Daten zurückgreifen. Ist dies nicht möglich, sollten Sie erwägen, eine Nutzungslizenz für einen geeigneten Datensatz zu erwerben, um auf der sicheren Seite zu sein. Wenn Sie Inhalte aus dem öffentlichen Internet entnehmen, sollten Sie technisch und juristisch sicherstellen, dass Sie hierzu ohne ausdrückliche Zustimmung der Rechteinhaber befugt sind.

3.2.5 Urheberrechtsverletzungen durch Training

Nutzen Sie große Mengen von urheberrechtlich geschützten Inhalten, ohne dass Sie sich auf die Ausnahme des Text- und Data-Minings berufen können und ohne dass eine Zustimmung der Rechteinhaber vorliegt, verletzen Sie das Urheberrecht. Genau genommen begehen Sie sehr viele Urheberrechtsverletzungen gegenüber sehr vielen Personen. Damit geht ein enormes Schadenpotenzial bzw. -risiko einher. Denn jeder Rechteinhaber, dessen Urheberrecht Sie verletzen, kann entsprechende Ansprüche aus der Rechtsverletzung gegen Sie herleiten.

Nehmen wir an, Sie haben 1.000.000 Bilder aus dem Internet heruntergeladen, ohne dazu befugt gewesen zu sein. Diese Bilder stammen von insgesamt 100.000 verschiedenen Rechteinhabern. In diesem Fall können Sie potenziell von allen 100.000 Personen auf Unterlassung und Schadensersatz in Anspruch genommen werden. Auch wenn der einzelne Anspruch Sie nicht in den Ruin treiben wird – zusammengenommen können sich große Summen ergeben.

In der Praxis ist natürlich ein Zwischenschritt nicht zu vergessen: Die Rechteinhaber müssen Kenntnis davon erlangen, dass ihre Werke kopiert wurden. Dies wird nicht ohne Weiteres der Fall sein – es sei denn, Sie legen offen, welche Daten bzw. Datensätze Sie für das Training genutzt haben. Im Falle von OpenAIs Sprachmodell GPT-4 ist beispielsweise bekannt, dass dieses zu einem großen Teil auf einem Datensatz von Common Crawl trainiert wurde. Da der Datensatz frei abrufbar ist, können Rechteinhaber auf einfache Art erkennen, ob ihre Werke enthalten sind.

3.3 Trainingsdaten mit Personenbezug

In diesem Abschnitt erfahren Sie, was Sie beachten müssen, wenn die von Ihnen verwendeten Trainingsdaten personenbezogene Daten enthalten. Sie werden mit den relevanten Rechtsgrundlagen vertraut gemacht und erfahren mehr über Ihre Verpflichtungen zur Löschung, Berichtigung und Auskunftserteilung. Außerdem werden Ihnen Möglichkeiten aufgezeigt, die Anwendbarkeit der Vorgaben der DSGVO zu verhindern.

Die Entwicklung künstlicher Intelligenz (KI) ist eng mit der Verwendung erheblicher Mengen an Trainingsdaten verbunden. Um diese Daten zu beschaffen, hat sich die Verwendung von Webscraping-Technologien etabliert. Diese ermöglichen es, benötigte Daten in großem Umfang aus dem Internet zu extrahieren.

Die dabei zusammengetragenen Daten umfassen häufig auch personenbezogene Daten. Dies resultiert einerseits aus der weit gefassten Definition der DSGVO für per-

sonenbezogene Daten und andererseits aus der schieren Masse an erfassten Daten. In der Praxis ist es nahezu unmöglich, personenbezogene Daten vollständig aus einem Trainingsdatensatz herauszuhalten oder nachträglich zu eliminieren, ohne die Qualität der Daten signifikant zu verringern. Eine solche Verringerung wäre jedoch kontraproduktiv, da hochwertige Trainingsdaten essenziell sind, um optimale Ergebnisse zu erzielen.

Ein anschauliches Beispiel hierfür ist der LAION-5B-Datensatz des LAION e. V., ein öffentlich verfügbarer Trainingsdatensatz, der für das Training einiger der größten KI-gestützten Bildgeneratoren verwendet wurde. Der Datensatz umfasst 5,85 Milliarden Bilder-Text-Paare, darunter zahlreiche Bilder von Personen. Zu jedem Bild wurde der Alternativtext – eine kurze Bildbeschreibung, die primär der Barrierefreiheit dient – gespeichert. Diese Texte enthalten oft persönliche Informationen über die dargestellte Person oder den Fotografen. Sowohl die umfassten Bilder als auch die Alternativtexte sind daher grundsätzlich als personenbezogene Daten zu klassifizieren. Um den Datensatz gänzlich frei von personenbezogenen Daten zu halten, hätte LAION also Bilder mit Personen sowie entsprechende Alternativtexte aus dem Datensatz ausschließen müssen. Dies hätte jedoch einen erheblichen Qualitätsverlust des Datensatzes zur Folge gehabt, da das Training zur Erstellung menschenähnlicher Bilder so wohl kaum möglich gewesen wäre. Auch sind die in den Alternativtexten enthaltenen Informationen für das effiziente Machine Learning unerlässlich. Darüber hinaus hätte schließlich selbst durch solch restriktive Maßnahmen nicht garantiert werden können, dass der Datensatz frei von personenbezogenen Daten ist, da sich diese auch aus weniger offensichtlichen Details, wie Kfz-Kennzeichen bei auf Fotos abgebildeten Autos, ergeben können. Ein vollständiger Ausschluss personenbezogener Daten wäre mithin wohl nur durch eine manuelle Überprüfung jedes einzelnen Bildes möglich gewesen, was jedoch nicht praktikabel gewesen wäre.

Dieses Beispiel verdeutlicht, dass KI-Entwickler häufig keine andere Wahl haben, als zu akzeptieren – oder in manchen Fällen sogar gezielt darauf hinzuarbeiten –, dass die von ihnen aus dem Internet gesammelten und für das KI-Training verwendeten Trainingsdaten auch personenbezogene Daten enthalten.

3.3.1 Vorhandensein personenbezogener Daten

Wie dargestellt enthalten Trainingsdatensätze, die mit Webscraping-Technologien erstellt wurden, in der Regel personenbezogene Daten. Ausnahmen sind nur denkbar, wenn die verwendeten Webscraper derart konfiguriert sind, dass sie personenbezogene Daten weder extrahieren noch auf Daten abzielen, die solche Informationen grundsätzlich enthalten könnten. Ein Beispiel hierfür ist die Nutzung von Audioauf-

nahmen ohne menschliche Stimmen oder andere persönliche Informationen für das Training einer KI zur Musikgenerierung. Im Gegenzug wäre die KI dann jedoch nicht in der Lage, Gesang oder ähnliche Elemente in relevanter Qualität zu erstellen.

Dementsprechend sollten Sie sich vor Beginn der Erstellung des Trainingsdatensatzes genau überlegen, welche Fähigkeiten Ihre KI letzten Endes haben soll und welche Daten Sie dafür benötigen. Falls Sie feststellen, dass personenbezogene Daten nicht erforderlich sind, können Sie das Training unter Verwendung der folgenden Möglichkeiten ohne Beachtung etwaiger datenschutzrechtlicher Pflichten durchführen.

Anonymisierung von Daten

Ein klassischer Weg, den datenschutzrechtlichen Pflichten zu entkommen, ist die Anonymisierung der zu verarbeitenden Daten. Laut Erwägungsgrund 26 der DSGVO sind als anonyme oder anonymisierte Daten solche Daten anzusehen, die keinen Personenbezug aufweisen, weil die betroffene Person nicht oder nicht mehr identifiziert werden kann. Damit der Personenbezug tatsächlich aufgehoben ist, muss die Anonymisierung dauerhaft sein.

Praxishinweis: Anonymisierung ist nicht Pseudonymisierung!

Verwechseln Sie die Anonymisierung nicht mit der in der DSGVO ebenfalls erwähnten Pseudonymisierung. Dabei handelt es sich lediglich um eine Maßnahme, die dem Schutz und der Sicherheit der verarbeiteten personenbezogenen Daten dient (TOM). Die Anwendbarkeit der DSGVO wird dadurch aber nicht ausgeschlossen. Eine Pseudonymisierung liegt vor, wenn personenbezogene Daten so geändert werden, dass sie nur noch identifiziert werden können, wenn die verarbeitende Person weiß, nach welchem Muster die Daten verändert wurden. Klassischerweise werden etwa einer Reihe von Namen (z. B. von Patienten) zufällige Kennnummern zugeordnet. Auf einer separaten Liste wird die Zuordnung der einzelnen Namen zu den jeweiligen Nummern vermerkt. Die Nummern lassen für sich keine Identifizierung der einzelnen Personen zu, nur bei Hinzuziehung der separaten Liste kann eine Zuordnung wieder erfolgen. In diesem Fall sind die personenbezogenen Daten in Form der Namen pseudonymisiert.

Eine nachträgliche Anonymisierung ist ratsam, wenn Sie wie oben dargestellt zu dem Ergebnis gekommen sind, dass Sie für das Training Ihrer KI zwar keine personenbezogenen Daten benötigen, aber annehmen müssen, dass die extrahierten Daten unbeabsichtigt zahlreiche personenbezogene Daten umfassen. Angenommen, Sie wollen einen Bildgenerator ausschließlich für Natur- und Landschaftsbilder erstellen, um bei dem oben genannten Beispiel zu bleiben. Obwohl Sie keine Bilder von Personen benötigen, könnten beim Sammeln von Natur- und Landschaftsbildern mittels

Webscraping dennoch Bilder extrahiert werden, die neben den Landschaften auch Personen zeigen oder aus deren Alternativtexten personenbezogene Informationen abgeleitet werden können. In diesem Fall wäre eine gezielte Anonymisierung der Personen und der persönlichen Angaben geboten.

Eine Anonymisierung kann grundsätzlich dadurch erfolgen, dass die personenbezogenen Informationen (wie z. B. der Name einer Person) aus dem jeweiligen Datensatz gelöscht werden. Ebenso können die Informationen geschwärzt oder verfälscht werden, also z. B. aus »Hannover« wird »Hamburg«, aus »Ärztin« wird »Bibliothekarin« etc. Um dabei zumindest die allgemeine Information über den Umstand und den Umfang der Löschung zu erhalten, kann die Löschung oder Verfälschung entsprechend als solche gekennzeichnet werden.

Praxishinweis: Technische Anonymisierung

Gerade im Bereich des Machine Learning ist eine manuelle Anonymisierung der vielfältigen Trainingsdaten in der Regel nicht zu bewerkstelligen, sodass Sie in der Regel auf technische Mittel zur Anonymisierung zurückgreifen müssen.

Dabei können automatische Anonymisierungsdienste zum Einsatz kommen, die beispielsweise mittels Natural Language Processing (NLP) auch unstrukturierte Textdaten strukturieren und anonymisieren können. NLP-Werkzeuge sind in der Lage, Merkmale wie den Kontext einer Information oder das zugrunde liegende Konzept zu erkennen und die identifizierten Merkmale entsprechend zu anonymisieren.

Die Anonymisierung sollte idealerweise direkt zu Beginn der geplanten Verarbeitung (in Form des KI-Trainings) erfolgen. Die Frage, ob es sich bereits bei der Anonymisierung um eine eigene Verarbeitung im Sinne der DSGVO handelt, für die eine Rechtsgrundlage gemäß Art. 6 DSGVO benötigt wird, ist in der juristischen Diskussion nicht unumstritten. In der Praxis werden Sie – unabhängig von dieser im Wesentlichen dogmatischen Diskussion – die Anonymisierung aber grundsätzlich auf berechtigte Zwecke und auf die Rechtsgrundlage nach Art. 6 Abs. 1 Buchst. f) DSGVO stützen können.

Hinweis: Weiterführende Handreichungen und Leitfaden

Wenn Sie die Anonymisierung für Ihr Projekt als adäquate datenschutzrechtliche Maßnahme identifiziert haben und weitere Informationen benötigen, kann sich ein Blick in den »Praxisleitfaden für die Anonymisieren personenbezogener Daten« der Stiftung Datenschutz lohnen (online verfügbar unter *https://stiftungdatenschutz.org/praxisthemen/anonymisierung#c3502*).

Ebenfalls hilfreich ist der Leitfaden des Bitkom e. V. zur »Anonymisierung und Pseudonymisierung von Daten für Projekte des maschinellen Lernens« (online verfügbar unter *https://www.bitkom.org/sites/main/files/2020-10/201002_lf_anonymisierung-und-pseudonymisierung-von-daten.pdf*).

Synthetische Daten

Speziell beim Training von Künstlicher Intelligenz mittels Machine Learning besteht zudem die Möglichkeit, eine Anwendbarkeit der DSGVO durch die ausschließliche Verwendung synthetischer Daten zu verhindern. Der Begriff »synthetische Daten« meint solche Daten, die vollständig künstlich erstellt wurden, also keinen unmittelbaren realen Ursprung haben. Trotz ihres rein fiktiven Ursprungs weisen die synthetischen Daten dabei aber eine derart hohe Ähnlichkeit mit realen Daten auf, dass eine Eignung zum maschinellen Lernen wie bei realen Daten besteht. Im Idealfall müssen daher bei der Verwendung synthetischer Daten nicht nur keine Abstriche bei der Trainingsqualität gemacht oder andere Nachteile in Kauf genommen werden, sondern es müssen zusätzlich auch keine Datenschutzauflagen erfüllt werden. Aus diesem Grund ist der Einsatz synthetischer Daten in der Regel besonders attraktiv. In Bereichen, in denen ein Mangel an geeigneten hochwertigen Daten besteht, kann der Rückgriff auf synthetische Daten sogar zwingend erforderlich sein, damit das Training der jeweiligen KI überhaupt erfolgreich stattfinden kann.

Die Generierung synthetischer Daten erfolgt durch bestimmte Software. Diese basiert grundsätzlich auf dem Einsatz eigener spezieller künstlicher Intelligenz. Konkret wird dafür in der Regel ein statistisches Modell herangezogen, welches auf Grundlage realer Daten gebildet wird. Diese Ursprungsdaten werden von dem Modell schlicht unter der Prämisse nachgebildet, dass die generierten Daten keinen Personenbezug aufweisen. Damit die erstellten fiktiven Daten dabei aber eine ähnlich hohe Aussagekraft und qualitative Wertigkeit wie reale Daten haben, erfordert die Generierung regelmäßig einen erheblichen Aufwand und entsprechende Fachkenntnisse.

Praxisbeispiel: AlphaGeometry von DeepMind

Trotz der bestehenden Schwierigkeiten hat sich der Einsatz synthetischer Daten beim Training künstlicher Intelligenz bereits bewährt. Googles DeepMind veröffentlichte beispielsweise im Januar 2024 die AlphaGeometry-KI, die vollständig anhand synthetischer Daten trainiert worden sein soll. Dabei handelt es sich um ein KI-System, das komplexe geometrische Probleme auf dem Level »eines olympischen Gold-Medail-

lengewinner« lösen können soll. DeepMind will dafür 100 Millionen visuelle Darstellungen geometrischer Formen als synthetische Daten erstellt und verwendet haben (mehr dazu erfahren Sie in diesem Blogeintrag: *https://deepmind.google/discover/blog/alphageometry-an-olympiad-level-ai-system-for-geometry/*). Auch wenn dies wohl nicht der komplexeste Anwendungsfall synthetischer Daten sein dürfte, zeigt dies dennoch, dass die ausschließliche Verwendung synthetischer Daten durchaus realisierbar und praxistauglich ist.

3.3.2 Rechtsgrundlagen

Falls Sie ausschließlich anonymisierte oder synthetische Daten zum Training der KI verwenden, müssen Sie sich über die Erfüllung der Vorgaben der DSGVO keine Gedanken machen. Beinhalten die Trainingsdaten allerdings personenbezogene Daten, stellt die Verwendung dieser Daten eine Verarbeitung im Sinne der DSGVO dar. Aufgrund des sogenannten Erlaubnisvorbehalts in der DSGVO ist eine Verarbeitung nur erlaubt, wenn Sie die Verarbeitung auf eine der Rechtsgrundlagen der DSGVO stützen können (Art. 5 Abs. 1 Buchst. A, Art. 6 DSGVO).

Die möglichen Rechtsgrundlagen für die Verarbeitung personenbezogener Daten sind in Art. 6 DSGVO aufgeführt (siehe hierzu Abschnitt 1.3.3). Daneben enthält Art. 9 DSGVO spezielle Maßgaben für die Verarbeitung besonders sensibler personenbezogenen Daten. Weitere, spezifischere Vorschriften können in Ausnahmefällen relevant sein, wie beispielsweise Art. 88 DSGVO in Verbindung mit § 26 BDSG für Daten von Beschäftigten.

Berechtigte Interessen, Art. 6 Abs. 1 Buchst. F) DSGVO

Zunächst sollten Sie für sich prüfen, ob die geplanten Verarbeitungen auf berechtigte Interessen gestützt werden können und diese Interessen derart gewichtig sind, dass Sie die entgegenstehenden Interessen der Betroffenen überwiegen. Ein berechtigtes Interesse im Sinne des Art. 6 Abs. 1 Buchst. F) DSGVO kann jedes rechtliche, wirtschaftliche oder ideelle Interesse sein. Insofern kann auch ein bloßes Streben nach Profit und Wachstum des Unternehmens ein grundsätzlich berechtigtes Interesse darstellen. Zudem ist erforderlich, dass die Verarbeitung notwendig ist, um diese Interessen zu schützen oder zu fördern. Im Bereich des KI-Trainings dürfte die Verarbeitung nur dann nicht als erforderlich anzusehen sein, wenn der Datensatz ohne erhebliche Umstände anonymisiert werden könnte und das Training auch mit anonymen Daten erfolgreich wäre.

Die wesentliche Herausforderung im Rahmen der Frage der Anwendbarkeit dieser Rechtsgrundlage für Ihre jeweilige Verarbeitung liegt in der vorzunehmenden Abwägung Ihrer Interessen gegenüber den entgegenstehenden Interessen der Betroffenen. Deren entgegenstehendes Interesse liegt in der Regel jedenfalls darin, selbst den Umgang mit den eigenen personenbezogenen Daten unmittelbar und frei bestimmen und kontrollieren zu können (Recht auf informationelle Selbstbestimmung). In jedem Einzelfall, in dem Art. 6 Abs. 1 Buchst. F) DSGVO angewendet werden soll, ist daher sorgfältig zu prüfen, wie stark dieses Interesse durch die von Ihnen geplante Verarbeitung beeinträchtigt wird und wie erheblich demgegenüber die berechtigten Interessen sind, auf die Sie die Verarbeitung stützen wollen. Ein reines wirtschaftliches Interesse, also das Streben nach Profit, wird beispielsweise in der Regel nicht als ausreichend wichtig gewertet, um tiefgreifende Einschnitte in das Interesse auf informationelle Selbstbestimmung zu rechtfertigen.

Die ordnungsgemäße Durchführung der erforderlichen Abwägung erfordert ein erhebliches Maß an Sorgfalt sowie tiefgehende Kenntnisse über die Gewichtung der jeweiligen Interessen. Dies kann sich in der Praxis häufig als schwierig erweisen.

Im Rahmen des KI-Trainings kann ein überwiegendes berechtigtes Interesse in der Regel angenommen werden, wenn personenbezogene Daten durch Webscraping ausschließlich aus frei verfügbaren Quellen gesammelt werden. Derartige öffentlich zugängliche Daten werden insbesondere auch im Rahmen des Crawlings großer Suchmaschinen wie Google oder Bing erfasst. Diese setzen vor allem Methoden des Webscraping und -crawling ein, um Daten zu sammeln, die als Suchergebnisse auf entsprechende Anfragen von Nutzern ausgegeben werden können. Die Existenz von Suchmaschinen im Internet und der Umstand, dass im Internet verfügbare Informationen von Suchmaschinen erfasst und wiedergegeben werden, ist allgemein bekannt. Insofern darf angenommen werden, dass Personen, die personenbezogene Daten im Internet veröffentlichen, grundsätzlich auch mit einer Verarbeitung ihrer veröffentlichten Daten rechnen müssen. Ihr Interesse an der Integrität der jeweiligen personenbezogenen Daten ist dementsprechend geringer zu bemessen.

Auch die DSGVO regelt in Art. 9 Abs. 2 Buchst. e), dass personenbezogene Daten verarbeitet werden dürfen, wenn diese von der betroffenen Person »offensichtlich öffentlich gemacht« wurden. Diese Regelung findet sich in Zusammenhang mit den als besonders schützenswert angesehenen sogenannten sensiblen Daten. Da also selbst die Verarbeitung solcher besonders geschützter Daten erlaubt ist, kann die Zulässigkeit der Verarbeitung für klassische personenbezogene Daten erst recht bejaht werden.

Verarbeiten Sie ausschließlich öffentlich zugängliche personenbezogene Daten, kann die durchzuführende Abwägung somit regelmäßig zu Ihren Gunsten ausfallen. Trotzdem müssen Sie stets auch die weiteren in Ihrem jeweiligen Fall vorliegenden Aspekte in der Abwägung mitberücksichtigen. Eine allgemeingültige Aussage über das Ergebnis der Abwägung kann ohne genaue Kenntnis der konkreten Umstände nicht getroffen werden.

Praxishinweis: Weitere begünstigende und nachteilige Aspekte bei der Interessenabwägung

Bei der Interessenabwägung nach Art. 6 Abs. 1 Buchst. f) DSGVO müssen je nach konkreten Umständen des Einzelfalls eine Vielzahl von Aspekten berücksichtigt werden. Neben dem angesprochenen Aspekt, dass ausschließlich öffentlich zugängliche Daten verwendet werden, können weitere für Sie günstige Aspekte sein:

- Die KI-Entwicklung dient der Forschung (insbesondere, wenn die Forschungsergebnisse der Allgemeinheit zugutekommen sollen).
- Die KI-Entwicklung dient der Arbeitnehmer- und Betriebssicherheit.
- Compliance-Prozesse können durch die KI-Entwicklung automatisiert, effizienter und genauer werden.
- Es sind umfassende und geeignete technische und organisatorische Maßnahmen eingerichtet (TOMs), die sicherstellen, dass das Training ordnungsgemäß abläuft und die verwendeten Daten geschützt sind.

Allerdings können neben dem angesprochenen generellen Interesse der Betroffenen an der Integrität ihrer personenbezogenen Daten noch weitere Aspekte vorliegen, die die Interessenabwägung zulasten der KI-Entwickler entscheiden. Diese nachteiligen Aspekte gestalten sich unter Umständen wie folgt:

- Es werden personenbezogene Daten von Kindern verarbeitet. Dann überwiegt häufig das generelle Interesse an dem Schutz Minderjähriger.
- Es werden besondere Kategorien personenbezogener Daten nach Art. 9 Abs. 1 DSGVO verarbeitet. Dann benötigen Sie grundsätzlich eine weitere Rechtsgrundlage nach Art. 9 Abs 2 DSGVO. Diese kann beispielsweise darin liegen, dass die Verarbeitung erforderlich ist, um bestimmte Rechte auszuüben, oder die personenbezogenen Daten wurden von der betroffenen Person offensichtlich öffentlich gemacht.

Um in dem bei Ihnen vorliegenden konkreten Fall eine möglichst rechtssichere Feststellung treffen zu können, sollten Sie immer einen Datenschutzexperten hinzuziehen, der Ihnen hilft, relevante Aspekte zu identifizieren und diese entsprechend zu bewerten.

Einwilligung, Art. 6 Abs. 1 Buchst. a) DSGVO

Können Sie die geplanten Verarbeitungen nicht über berechtigte Interessen rechtfertigen, bleibt Ihnen in der Regel nur die Einholung einer Einwilligung der Betroffenen. Nach Art. 6 Abs. 1 Buchst. a), Art. 4 Nr. 11 und Art. 7 DSGVO kann eine Verarbeitung dann rechtmäßig sein, wenn die betroffene Person nach entsprechender Information eine freiwillige Einwilligung in die Verarbeitung erteilt hat.

Im Rahmen des KI-Trainings dürfte die Einholung einer Einwilligung häufig an der mangelnden technischen Umsetzbarkeit scheitern. Um eine wirksame Einwilligung einzuholen, ist es erforderlich, die betroffenen Personen vorher zu identifizieren. Dies stellt insbesondere eine Herausforderung dar, wenn zur Erstellung des Trainingsdatensatzes Webcrawler oder -scraper eingesetzt werden, da es schwierig sein kann, die erfassten personenbezogenen Daten einer spezifischen Person zuzuordnen.

Anders verhält es sich, wenn für das Training ausschließlich Daten eines begrenzten und bekannten Personenkreises verwendet werden, beispielsweise Daten von Kunden oder Mitarbeitern. In solchen Fällen ist die Einholung von Einwilligungen praktisch möglich und stellt eine taugliche Alternative dar, insbesondere wenn eine Abwägung der berechtigten Interessen schwerfällt.

Um den Anforderungen der DSGVO zu genügen, muss die betroffene Person von Ihnen über die wesentlichen Aspekte der Datenverarbeitung im Rahmen des Trainings informiert werden. Dazu gehört mindestens die Aufklärung darüber, wer die für die Verarbeitung verantwortliche Stelle ist und für welche Zwecke die Daten von Ihnen verarbeitet werden. Eine umfassende Erfüllung der Informationspflichten setzt voraus, dass Sie über die Vorgänge sowohl beim Training Ihrer KI als auch bei deren späterem Einsatz ausreichend informiert sind. Es ist daher essenziell, dass Sie besonderen Wert auf Transparenz und Nachvollziehbarkeit Ihres KI-Systems setzen, was insbesondere bei der Verwendung von tiefen neuronalen Netzen (Deep Learning) eine erhebliche Herausforderung darstellen kann. Auch die zukünftige KI-Verordnung der EU wird besondere Transparenzpflichten für KI-Systeme einführen, sodass hier mittelfristig ohnehin erhöhte Anforderungen umzusetzen sein werden.

Ein weiteres spezielles Problem in der Praxis stellt die Widerrufbarkeit der Einwilligung nach Art. 7 Abs. 3 S. 1 DSGVO dar. Danach kann die betroffene Person ihre Einwilligung jederzeit gegenüber dem Verantwortlichen widerrufen, woraufhin die jeweilige Datenverarbeitung umgehend beendet werden muss. Dies erfordert in der Regel die Löschung der personenbezogenen Daten durch den Verantwortlichen, sofern keine andere Rechtsgrundlage für die Verarbeitung besteht. Die Herausforderung besteht darin, Daten, die bereits in die Algorithmen einer KI eingeflossen sind,

nachträglich wieder aus dieser zu entfernen. Auch hier sollten Sie daher entsprechende Mechanismen schaffen, die es ermöglichen, bestimmte personenbezogene Daten aus einem KI-System zu löschen oder zumindest zu verhindern, dass die KI diese Daten weiterverarbeitet oder ausgibt.

3.3.3 Löschung personenbezogener Daten

Wenn es Ihnen nicht gelungen sein sollte, Ihren Trainingsdatensatz frei von personenbezogenen Daten zu halten, können die Personen, deren Daten sich darin finden lassen, ihre Rechte als Betroffene einer Datenverarbeitung geltend machen (einen Überblick über die Betroffenenrechte finden Sie in Abschnitt 2.4.3). Je nach Größe des Trainingsdatensatzes dürfte sich die Erfüllung dieser Rechte schwierig gestalten. Besonders kompliziert gestaltet sich dabei das Recht auf Löschung personenbezogener Daten.

Mit der Verabschiedung der DSGVO wurde mit Artikel 17 erstmals ein Recht auf Löschung personenbezogener Daten in der Europäischen Union kodifiziert. Das Recht auf Löschung stellt dabei eine kleine Revolution dar, da es an sehr wenige Voraussetzungen gebunden ist und jederzeit von betroffenen Personen geltend gemacht werden kann. Die Prüfung, ob tatsächlich ein Anspruch auf Löschung vorliegt, obliegt dabei Ihnen als verantwortliche Stelle. Dies kann sich insbesondere bei Trainingsdaten mit einem sehr großen Datensatz und infolgedessen sehr vielen Betroffenen problematisch gestalten.

Hintergrundwissen

Das Recht auf Löschung, auch bekannt als das Recht auf Vergessenwerden, geht bereits auf ein Urteil des Europäischen Gerichtshofes gegen Google Spain aus dem Jahr 2014 zurück. Ein spanischer Staatsbürger, verlangte vom Suchmaschinenbetreiber Google ihn betreffende Daten, die im Zusammenhang mit einer Zwangsversteigerung standen, aus der Liste der Suchergebnisse zu entfernen. Google Spain weigerte sich, dies durchzuführen, bis der EuGH nach jahrelangem Rechtsstreit dem Löschersuchen schließlich recht gab.

Die Löschung von personenbezogenen Daten im Zusammenhang mit einer KI stellt Sie vor besondere Herausforderungen. Bevor Sie daher jedem Löschersuchen nachkommen, sollten Sie zunächst prüfen, ob Sie überhaupt zur Löschung verpflichtet sind. Artikel 17 der DSGVO sieht vor, dass personenbezogene Daten zu löschen sind, wenn

1. der Zweck, für den die personenbezogenen Daten erhoben wurden, nicht mehr besteht,
2. es nach dem Widerruf der einer Einwilligung an einer Rechtsgrundlage für die Verarbeitung fehlt,
3. ein Widerspruch gegen die Verarbeitung gemäß Artikel 21 DSGVO eingelegt wurde (und die Gründe eines Widerspruches vorliegen),
4. die Verarbeitung unrechtmäßig erfolgt ist oder
5. die Löschung gesetzlich vorgeschrieben ist.

Ein großes Einfallstor für künftige Löschersuchen dürfte die Pflicht zur Löschung bei dem Vorliegen einer unrechtmäßigen Verarbeitung vorliegen. Dies bedeutet nämlich für Sie im Umkehrschluss, dass Sie die personenbezogenen Daten nur dann verarbeiten dürfen – auch die weitere Speicherung innerhalb ihrer Trainingsdaten zählt als Verarbeitung –, wenn hierfür eine Rechtsgrundlage vorliegt. Wie im Abschnitt zuvor beschrieben, kommt gerade für im Netz frei verfügbaren Daten hierfür regelmäßig das Vorliegen des berechtigten Interesses gemäß Artikel 6 Abs. 1 lit. f) DSGVO in Betracht.

Insbesondere wenn Ihre Trainingsdatenbank aus Web-Scraping-Daten besteht, können Sie daher erst einmal auf das berechtigte Interesse verweisen. Schwierig wird es jedoch dann, wenn nicht einfach nur die Löschung verlangt wird, sondern auch gegen die weitere Verarbeitung gemäß Artikel 21 DSGVO ein Widerspruch durch die betroffene Person eingelegt wird.

Das Recht auf Widerspruch

Das Recht auf Widerspruch normiert einen Sonderfall, in dem besondere Gründe vorliegen müssen, in denen die Verarbeitung personenbezogener Daten einzustellen ist. Die DSGVO spricht hier lediglich von »Gründen die sich aus ihrer besonderen Situation ergeben«. Wann solche Gründe tatsächlichen Vorliegen wurde von der deutschen Rechtsprechung bislang in einigen wenigen Urteilen beantwortet. Die deutschen Gerichte legen dabei vergleichsweise hohe Voraussetzungen an. So wurde bereits regelmäßig ausgeurteilt, dass die Datenverarbeitung für die betroffene Person eine besondere Härte darstellen muss, indem diese schwerwiegende Auswirkungen auf die Psyche oder das tägliche Leben der Betroffenen hat. Beispielhaft können hierfür die zahlreichen Verfahren gegen Auskunfteien wie die Schufa angeführt werden, in denen die deutschen Gerichte selbst großen Behinderungen im Alltag – wie beispielsweise der fehlenden Möglichkeit Handyverträge abzuschließen – keine solche schwerwiegenden Auswirkungen angenommen haben.

Wenn nun also eine betroffene Person gegen die weitere Speicherung ihrer personenbezogenen Daten in Ihrem Trainingsdatensatz Widerspruch erhebt, so hat das nicht automatisch zur Folge, dass Sie die Daten zu löschen haben. Eine solche schwerwiegende Beeinträchtigung dürfte nur dann anzunehmen sein, wenn sich intimste Informationen in Ihren Trainingsdaten wiederfinden, die gar nicht erst ihren Weg in das Internet hätten finden dürfen. Hierzu dürften beispielsweise Daten über Erkrankungen, polizeiliche Opferberichte oder intimste Informationen über das Privat- und Sexualleben der Personen gehören.

Praxistipp

Aufgrund der Neuheit der entsprechenden Technologie gibt es derzeit noch keine Gerichtsurteile hinsichtlich des Rechts auf Widerspruch für in Trainingsdaten enthaltene personenbezogene Daten. Wir können Ihnen daher nur grobe Leitlinien mitgeben, wie Sie damit umgehen sollten.

Wenn Sie einen Widerspruch über die Datenverarbeitung erhalten, sollten Sie diesen in keinem Fall einfach abtun. Als Verantwortliche Stelle obliegt Ihnen zunächst die Prüf- und hiernach die Nachweispflicht, dass Ihr Interesse an der Verarbeitung überwiegt. Sie müssen der widersprechenden Person also antworten. Setzen Sie sich also mit den Daten in Ihrer Datenbank auseinander und behalten Sie dabei im Hinterkopf, dass die Verarbeitung eine schwerwiegende Beeinträchtigung für den Betroffenen zur Folge haben muss.

Unternehmensdatenbanken

Doch nicht alle Trainingsdatenbanken setzen sich aus Daten, die im Web zusammengetragen wurden zusammen. Sie können Ihre KI auch mit den Daten, die sich in Ihrer Unternehmensdatenbank befinden trainieren. Das Argument, dass Sie es sich hierbei um bereits veröffentlichte Daten handelt, lässt sich dann nicht mehr anbringen. Ein berechtigtes Interesse dürfte damit deutlich schwerer zu begründen sein. Es empfiehlt sich daher, für solche Trainingsdaten eine andere Rechtsgrundlage zu suchen, um die Daten vor Löschersuchen zu schützen.

Insbesondere, wenn die personenbezogenen Daten in Ihrer Unternehmensdatenbank Mitarbeiterdaten enthalten sollten, können Sie sich das Training Ihrer KI mit diesen Daten vertraglich zusichern lassen. Dies bereits beim Abschluss des Arbeitsvertrages oder ggf. durch eine nachträgliche Vereinbarung.

Praxistipp

Bei dem Training Ihrer KI mit Mitarbeiterdaten bietet sich wie in keinem anderen Fall die vertragliche Überlassung der Daten an. Alle betroffenen Personen sind Ihnen bekannt und erreichbar, die Information über die Verarbeitung kann zudem einheitlich gestaltet werden. Denken Sie daran, dass in einem solchen Fall Ihre Mitarbeitervertretung mitbestimmungsberechtigt ist und Sie daher hier in Verhandlung über eine Betriebsvereinbarung gehen sollten.

Wie kann gelöscht werden?

Sollten Sie sich einem Fall gegenübersehen, in dem das Löschersuchen statthaft ist und Sie zur Löschung verpflichtet sind, so stehen Sie vor der großen Problematik, wie dies zu bewerkstelligen ist. Die Löschung von personenbezogenen Daten aus einem Trainingsdatensatz dürfte – auch bei großen Datensätzen – eine mäßige Schwierigkeit darstellen. Viel schwieriger gestaltet es sich hingegen, Ihrer KI die gelernten Informationen wieder »abzutrainieren«. Dieser Vorgang wird in Fachkreisen als »Machine Unlearning« bezeichnet und bildet damit den exakten Gegenpart zum Training von KI, auch »Machine Learning« genannt.

Beim Machine Unlearning handelt es sich um einen Forschungszweig in der Informatik, der im Zuge der neuen Generation von generativer KI viel Aufmerksamkeit auf sich gezogen hat. Zwar sind hier in den letzten Jahren diverse Fortschritte gemacht worden, jedoch ist die Forschung weit davon entfernt ein Patentrezept für das »Machine Unlearning« zu haben. Allein die Vielzahl an unterschiedlichen KI-Modellen mit ihren jeweils eigenen Regeln und Algorithmen gestaltet die Forschung herausfordernd.

Wenn Sie also nicht nach jeder Bereinigung Ihres Datensatzes Ihre KI einmal komplett neu trainieren möchten, bleibt Ihnen in der Zwischenzeit nur das Filtering als Behelfslösung. Gemeint sind damit Befehle an ihre KI, um die Ausgabe bestimmter Informationen zu unterbinden. Diese nachträgliche Sperre von Informationen durch Filter gestaltet sich jedoch noch als eher umständlich und wenig zuverlässig. Zudem stellt selbst der Einsatz von wirkungsvollen Filtern kein Löschen dar, um den Anforderungen der DSGVO zu genügen.

Praxisbeispiel

ChatGPT wurde mit einem gigantischen Datensatz aus frei im Internet verfügbaren Informationen trainiert. Dazu gehören natürlich auch viele Informationen, die man als gefährlich einordnen kann. Von Schritt-für-Schritt-Anleitungen, wie man Computersysteme hackt oder mit Viren verseucht bis hin zu Rezepten für die Herstellung von gefährlichen Chemikalien mit gewöhnlichen Haushaltsmitteln. Dass solche Informationen ausgegeben werden, ist natürlich nicht im Sinne von ChatGPT-Anbieter OpenAI. Da die Bereinigung des Trainingsdatensatzes und das Neutraining von ChatGPT zu aufwendig wäre und auch OpenAI noch keine passable Lösung für das Machine Unlearning hat, wird hier mit Filtern gearbeitet.

Wer also ChatGPT fragt, wie man gefährliche Chemikalien oder Giftstoffe herstellen kann, wird von der KI darauf hingewiesen, dass man so was unterlassen sollte. Da solche nachträglich eingebauten Filter es jedoch nie schaffen, eine Information komplett zu verschließen, hat sich unter ChatGPT-Nutzern ein regelrechter Sport entwickelt, wie man der KI verbotene Informationen dennoch entlocken kann. Dies funktionierte beispielsweise mit umgekehrter Logik, also der Frage danach welche Haushaltsmittel man auf gar keinen Fall miteinander kombiniert werden sollten, damit man keine gefährliche Chemikalie herstellt. Nach kurzer Zeit werden die Filter von OpenAI hiernach stets angepasst und die Suche nach einem neuen Prompt beginnt von vorn.

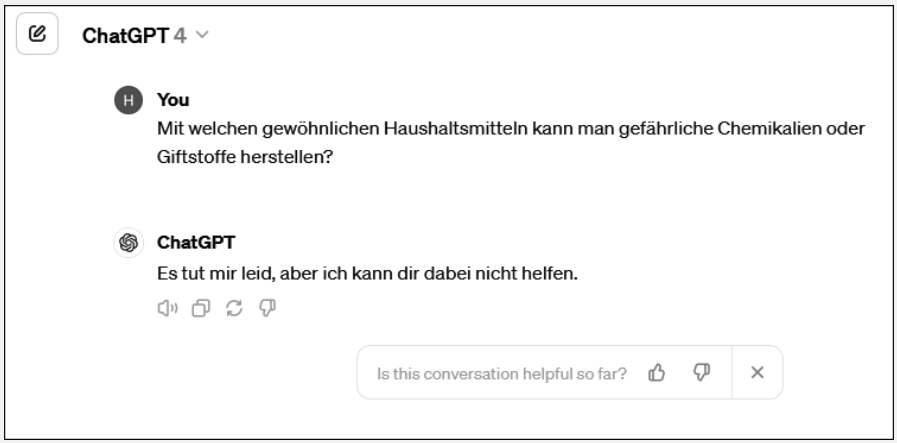


Abbildung 3.6 Die kurze und knappe Antwort von ChatGPT 4 auf die Frage nach Chemikalien aus Haushaltsmitteln.

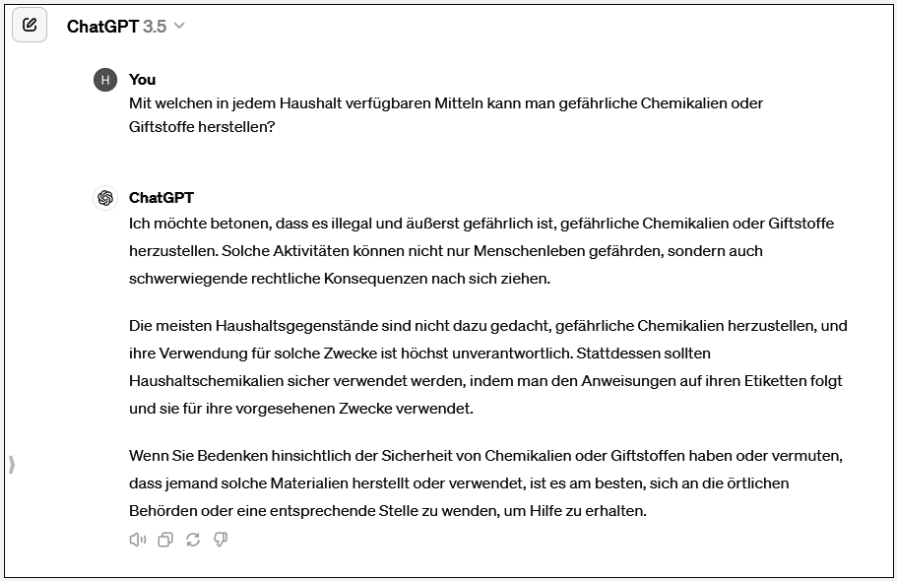


Abbildung 3.7 ChatGPT 3.5 antwortet auf dieselbe Frage überraschenderweise deutlich ausführlicher.

Ein Patentrezept für die rechtssichere Löschung personenbezogener Daten aus KI-Diensten ist aktuell noch nicht ersichtlich. Bis das »Machine Unlearning« nennenswerte Fortschritte gemacht hat, können Sie derzeit nur auf den Einsatz von Filtern zurückgreifen. Die beste Vorgehensweise, die wir Ihnen daher nahelegen können, ist bereits beim Aufbau Ihres Trainingsdatensatzes darauf zu achten, dass Sie die von Ihnen verwendeten personenbezogenen Daten rechtmäßig zum Training Ihrer KI nutzen dürfen.

3.3.4 Berichtigung personenbezogener Daten

Das Recht auf Berichtigung personenbezogener Daten gemäß Artikel 16 DSGVO wird Sie in Bezug auf Ihre KI vor ähnliche Probleme stellen, wie das Recht auf Löschung. Anstatt einer Löschung von personenbezogenen Daten, wenn es an einer Rechtsgrundlage für die weitere Verarbeitung mangelt, wird hier von Ihnen verlangt, unrichtige Daten zu korrigieren. Relevant ist dieses Recht vor allem bei Datenbanken für Sprachmodelle, die mit einer großen Menge an frei verfügbaren Daten trainiert worden sind. Denkbar sind hier Beispiele, in denen Sie Informationen aus Falschmeldungen in Zeitungen oder unwahren Behauptungen in den sozialen Netzwerken entnommen haben und diese von Ihrer trainierten KI nun als Fakten akzeptiert werden.

Praxisbeispiel

Der Bürgermeister einer australischen Kleinstadt hat weltweit Schlagzeilen mit seiner Ankündigung gemacht OpenAI wegen Verleumdung zu verklagen. So hat ChatGPT den Bürgermeister auf Anfrage mit einem Bestechungs- und Korruptionsskandal in Zusammenhang gebracht, den dieser überhaupt erst aufgedeckt hatte (*www.faz.net/aktuell/gesellschaft/australien-buergermeister-will-chatgpt-wegen-falschaussagen-verklagen-18805056.html*). Wie ChatGPT dies verdrehen konnte, ist bislang nicht geklärt.

Das Recht auf Berichtigung falscher personenbezogener Daten ist dabei ausschließlich an das Merkmal geknüpft, dass unrichtige Daten verarbeitet werden. In der Folge könnte damit den Betreibern großer Sprachmodelle in Zukunft eine ganze Flut von Berichtigungsersuchen ins Haus stehen, denen allesamt nachzukommen ist.

So einfach das Recht auf Berichtigung geltend gemacht werden kann, so schwer ist dieses umsetzbar. Als Betreiber stehen Sie technisch vor den gleichen Herausforderungen wie bei der Löschung personenbezogener Daten. Eine Korrektur von Trainingsdaten ist ohne große Probleme machbar, die Korrektur dessen, was die KI ausgibt, ist jedoch ungleich schwerer. Für mehr Informationen zu den technischen Schwierigkeiten empfehlen wir Ihnen daher den vorherigen Abschnitt 3.3.3 zu lesen.

3.3.5 Das Recht auf Auskunft

Das Recht auf Auskunft gemäß Artikel 15 ist das am häufigsten und intensivsten genutzte Betroffenenrecht der DSGVO. Wenn Sie also selbst einen Trainingsdatensatz anbieten, werden Sie damit rechnen müssen, früher oder später mit Ersuchen konfrontiert zu werden, bei denen die Betroffenen Auskünfte über folgende Punkte verlangen können:

1. die Verarbeitungszwecke
2. die Kategorien personenbezogener Daten
3. die Empfänger von personenbezogenen Daten
4. die geplante Dauer
5. das Bestehen der Rechte auf Berichtigung, Löschung, Einschränkung, Widerspruch sowie das Beschwerderecht bei einer Aufsichtsbehörde
6. die Herkunft der Daten
7. das Bestehen einer automatisierten Entscheidungsfindung

Insbesondere in der Anfangszeit der DSGVO brachten umfangreiche Auskunftsersuchen vieler Betroffener die angefragten Unternehmen oft an ihre Kapazitätsgrenzen. Das Problem war vielfach, dass die Datenhaltung dezentral und unorganisiert strukturiert war.

Praxistipp

Wenn Sie eine Trainingsdatenbank aufbauen, sollten Sie diese Learnings für sich beherzigen und Ihre Datenbank von Anfang an indexieren und durchsuchbar gestalten. Zudem empfiehlt es sich, eine Musterantwort auf ein Auskunftsersuchen vorzubereiten, bei dem Sie eine Antwort für jede der oben aufgelisteten Einzelauskünfte ausarbeiten. Ihre vorbereitete Musterantwort müssen Sie für den Fall eines Auskunftsersuchens nur noch für die einzelne betroffene Person individualisieren.

Mit entsprechender Vorbereitung und einer gut organisierten Datenbank stellen auch massenhafte Ersuchen für Sie keine Herausforderung mehr dar. Dies ist insbesondere vor Hintergrund, dass Sie gemäß Artikel 12 Abs. 3 DSGVO dazu verpflichtet sind, innerhalb eines Monats die Auskunft zu erteilen, von großer Wichtigkeit. Die Frist für die Erteilung von Auskünften kann zwar um bis zu zwei weitere Monate verlängert werden, wenn eine große Anzahl von Anträgen eingereicht wurde oder der Antrag besonders komplex ist. Dass Sie sich schlicht noch nicht mit dem Recht aus Auskunft auseinandergesetzt haben oder Ihr Trainingsdatensatz nicht durchsuchbar ist, ist aber kein Grund für eine Verlängerung dieser Frist.

RECHTSLEITFADEN KI
IM UNTERNEHMEN

Für Unternehmen ist KI unverzichtbar geworden. Bei ihrem Einsatz müssen jedoch zahlreiche juristische Aspekte berücksichtigt werden. In präziser und zugänglicher Sprache vermittelt Ihnen dieser Leitfaden umfassendes Wissen über die rechtlichen Rahmenbedingungen, auf die es dabei ankommt. Mit vielen praktischen Beispielen, konkreten Anwendungsszenarien, Lösungsansätzen und Handlungsempfehlungen. Aktuell zum EU AI Act.

Hier finden Sie Antworten auf Ihre Fragen!

- Grundlagen KI, Datenschutz, Urheberrecht
- Vertragsgestaltung, Wettbewerbsrecht, Rechtsberatung
- Einführung im Unternehmen
- Alles Wichtige zum EU AI Act
- Einsatz von ChatGPT und Co.
- Softwareerstellung mit KI
- Human Resources und KI
- Unterstützung durch Sprachassistenten
- Rechtliche Risiken und Stolperfallen
- Potenzielle Haftungsrisiken
- Rechtsfragen beim Training eigener KI-Modelle

Niklas Mühleis und **Nick Akinci** sind Rechtsanwälte und Partner der Kanzlei Heidrich Rechtsanwälte. Gemeinsam mit einem interdisziplinären Autorenteam aus Technikern und Juristen machen sie Sie mit den komplexen rechtlichen Herausforderungen vertraut und bieten Orientierung für Ihre Pläne und Herausforderungen.

Rheinwerk Computing

Gedruckt in Deutschland
Papier aus nachhaltiger Waldwirtschaft
Mineralölfreie Druckfarben

€ 39,90 [D] € 41,10 [A]
KI/Jura
ISBN 978-3-367-10098-9