

DIE WEGLEITUNG FÜR DEN UMGANG MIT KÜNSTLICHER INTELLIGENZ DES SAV

ROLAND KÖCHLI

Lic. iur., Rechtsanwalt, Altenburger Ltd legal + tax, Küsnacht

Stichworte: KI-Guidelines, künstliche Intelligenz, KI-Gesetzgebung

I. Einleitung

Nachdem kaum mehr eine Software offeriert wird, die keine KI integriert hat, war es an der Zeit eine Wegleitung zum sicheren Gebrauch von KI-Systemen durch die Anwaltschaft zu verfassen.

Das grösste Thema bei der Verwendung von Computersystemen, ist jeweils die Datenhaltung und der Datenverkehr. Aufgrund des Anwaltsgeheimnisses und der Datenschutzgesetzgebung ist immer zu klären, wo Daten hingehen und allenfalls gespeichert werden. Diesem Thema sind wir bereits in der Wegleitung zu Cloud und Outsourcing nachgegangen und die zu beachtenden Regeln können auf der SAV-Website (<https://www.sav-fsa.ch>) abgerufen werden. Die nachstehende Wegleitung widmet dem Thema Anwaltsgeheimnis und Datenschutz deshalb nur ein paar Absätze und richtet den Fokus auf die Punkte, die bei der praktischen Anwendung von KI-Systemen beachtet werden sollten. Zusätzlich gibt die Wegleitung einen Kurzüberblick über die bestehende und geplante Gesetzgebung im Zusammenhang mit KI-Systemen.

Die Wegleitung für den Umgang mit künstlicher Intelligenz wurde vom Vorstand des SAV am 14. 6. 2024 verabschiedet und kann unter <https://www.sav-fsa.ch> abgerufen werden.

II. SAV-Wegleitung für den Umgang mit künstlicher Intelligenz

1. Einleitung

Künstliche Intelligenz (KI) ist derzeit in aller Munde und soll gemäss Herstellern von Software in vielen Anwendungen stecken. Auch die Anwaltschaft nutzt vermehrt KI-Systeme, insbesondere im Zusammenhang mit Übersetzungssoftware oder mit der Analyse von grossen Datenmengen bei internen Untersuchungen für die Klientschaft oder bei Due-Diligence-Prozessen. Weiter kommt aber auch die sogenannte generative KI, d. h. eine KI, die Inhalte wie Bilder oder Texte selbst generiert, immer öfter zum Einsatz (so u. a. um Texte zusammenzufassen, zu korrigieren oder zu verbessern).

Die nachfolgenden Guidelines fokussieren auf generative KI. Die Empfehlungen gelten aber für alle Anwendungen, in denen KI zum Einsatz kommt.

Wie alle neuen Technologien eröffnen auch KI-Systeme neue Möglichkeiten, insbesondere Effizienzgewinne,

sie bergen aber auch Risiken. Die vorliegenden Guidelines sollen als Leitfaden dienen, um den Einsatz von KI in der Anwaltspraxis auf verantwortungsvolle Weise umzusetzen. Insbesondere sollten Kanzleien für den Gebrauch von KI-Systemen eine interne Weisung erlassen und die Regeln diesbezüglich festlegen.

KI-Systeme durchlaufen derzeit eine rasante Entwicklung. Die vorliegenden Ausführungen können also sehr schnell überholt sein. Es wird denn auch darauf verzichtet, spezifische KI-Anwendungen hervorzuheben.

2. Risiken im Umgang und sichere Nutzung

A) Anwaltsgeheimnis, Datenschutz und andere Geheimhaltungsverpflichtungen

Bei der Auswahl und Nutzung jeglicher Software und damit auch von KI-Anwendungen ist vorgängig zu klären, was mit den Inputdaten geschieht, insbesondere wer auf die Daten Zugriff hat und wo diese (zwischen-)gespeichert werden. Zu beachten sind nicht nur das anwaltliche Berufsgeheimnis, sondern auch das Datenschutzgesetz (DSG) sowie allfällige weitere Geheimhaltungsverpflichtungen. Mitglieder des Schweizerischen Anwaltsverbandes müssen zudem auch die Landesregeln berücksichtigen.

Grundsätzlich gibt es folgende Möglichkeiten:

1. Die Software wird im kanzleieigenen Netzwerk installiert und betrieben (sogenannte On-Premise-Lösung), und es ist sichergestellt, dass keinerlei Daten dieses interne Netzwerk verlassen oder ausserhalb der Infrastruktur der Anwaltskanzlei gespeichert werden.
2. Einhaltung der Regeln über das Outsourcing, wenn Anwendungen über einen Provider bezogen und allenfalls auf dessen Infrastruktur genutzt werden. Diesbezüglich kann auf die SAV-Wegleitung für IT-Outsourcing und Cloud-Computing verwiesen werden (abrufbar unter <https://digital.sav-fsa.ch/digitale-kanzlei-nutzung-von-clouddiensten>).
3. Eine Einverständnis- und Verzichtserklärung der aufgeklärten Klientschaft in Bezug auf Berufsgeheimnis und DSG kann eingeholt werden.

Ansonsten dürfen vertrauliche Informationen, Unternehmensgeheimnisse (Strategien, Finanzdaten usw.), Informationen oder persönliche Daten von Mitarbeitenden, Klienten oder Geschäftspartnern oder anderen Personen (egal in welchem Format, d. h. auch einschliesslich Fotos, Videos,

usw.) oder Inhalte, die durch Rechte des geistigen Eigentums, insbesondere Urheberrechte, geschützt sind, nicht in KI-Systeme eingegeben werden.

B) (Unabhängige) Überprüfung der Resultate

KI-Systeme sind weder allwissend noch arbeiten sie immer perfekt, und die Ergebnisse von KI-Systemen können falsch, unzureichend oder unvollständig sein. Es ist demzufolge von grösster Wichtigkeit, die Resultate, den sogenannten Output, unabhängig und kritisch zu überprüfen und allenfalls zu korrigieren bzw. zu ergänzen.

Dabei ist es wichtig zu wissen, dass eine KI nicht dazu in der Lage ist, die von ihr generierten Resultate zu überprüfen. Man kann also zum Beispiel nicht einfach ein KI-System fragen, ob der gelieferte Output der Wahrheit entspricht, da KI-Systeme dazu nicht in der Lage sind.

Fehler oder fehlerhafte Outputs von KI-Anwendungen können insbesondere folgenden Ursprung haben:

- Halluzinationen, d. h. die KI «erfindet» den Output.
- Falsche oder fehlende Informationen, weil dem KI-System die Datengrundlage fehlt. Ein KI-System kann beispielsweise auf einer in der Vergangenheit liegender Datengrundlage trainiert werden und demzufolge nach diesem Datum liegende Ereignisse nicht kennen. KI-Systeme neigen zudem desto eher zu Halluzinationen, je weniger Quellmaterial für eine bestimmte Fragestellung in das Training eingeflossen ist.
- Sycophancy: Ein KI-Modell passt seine Antworten so an, dass sie mit der Sichtweise des Benutzers übereinstimmen, selbst wenn diese Sichtweise objektiv verzerrend ist.

Zu beachten ist auch, dass KI-Systeme voreingenommen sein können. Solche sogenannte Biases können aufgrund des Datensatzes, der für das Training des KI-Systems verwendet wurde, die Art und Weise des Trainings an sich sowie Modellierungsentscheide der Programmierer entstehen. Das ist primär bei der Verwendung von KI-Systemen zur Datenanalyse ein Thema, muss aber auch bei generativen KI-Systemen im Hinterkopf behalten werden.

C) Haftung

Als Auftragnehmer haftet der Rechtsanwalt bzw. die Rechtsanwältin für eine allfällige Schlechterfüllung seines oder ihres Auftrags. Sie können sich nicht darauf berufen, die KI habe einen Fehler gemacht.

Beim Einsatz von KI-Systemen in Absprache mit der Klientschaft (z. B. beim Durchforsten grosser Datenmengen) empfiehlt es sich, im Vorfeld die Verantwortung für das Resultat anzusprechen und im Rahmen des rechtlich Zulässigen eine Haftungsbeschränkung zu vereinbaren.

D) Urheberrecht

Aktuell wird die Frage der Nutzung von urheberrechtlich geschütztem Material zum Training von Large Language Models (LLM) oder von Bild- bzw. Videogeneratoren kontrovers diskutiert. Umstritten ist, ob solche Trainingsdaten ohne Zustimmung von Rechteinhaberinnen deren Urheberrechte verletzt bzw. ob die Nutzung zu Trainings-

zwecken überhaupt eine urheberrechtsrelevante Nutzung darstellt. Diese Fragen stehen allerdings weniger bei der reinen Nutzung von KI-Anwendungen im Vordergrund; sie können aber relevant sein, wenn Kanzleien ihre eigenen LLM trainieren oder bestehende LLM ergänzen.

Ohne entsprechendes Nutzungsrecht können Inputdaten für das Training von KI-Systemen oder zur Generierung von Outputs allenfalls Urheberrechte Dritter verletzen.

Der Output gilt, sofern es sich um eine reine KI-Schöpfung handelt, grundsätzlich nicht als geistige Schöpfung im Sinne des Urheberrechts. Wird ein KI-System nur für die Ideensammlung, für einen ersten Entwurf o. Ä. eingesetzt, kann in der Folge dennoch eine geistige Schöpfung mit individuellem Charakter entstehen, und es kann trotz Nutzung von KI-Systemen ein urheberrechtlich geschütztes Werk entstehen. Nicht gänzlich ausgeschlossen ist es auch, dass ein bestimmter KI-generierter Output Urheberrechte Dritter verletzt, indem der Output einem urheberrechtlich geschützten Werk zu stark ähnlich bleibt. Dies ist aber sehr unwahrscheinlich.

E) Hinweispflicht

Einzelne Anbieter haben in ihren Nutzungsbedingungen Bestimmungen, welche die Nutzer verpflichten, den Einsatz von KI offenzulegen. Vor dem Einsatz einer KI-Anwendung sollten daher, nicht zuletzt aus diesem Grund, die Nutzungsbedingungen des betreffenden Anbieters geprüft werden.

Abgesehen davon könnte eine Hinweispflicht allenfalls auch dann bestehen, wenn der Klient als Auftraggeber die höchstpersönliche Ausführung des Auftrags durch die Anwältin oder den Anwalt verlangt oder zu Recht erwartet.

III. Regulierung von KI-Systemen

In der Schweiz hat der Bundesrat im November 2023 das UVEK beauftragt, bis Ende 2024 mögliche Ansätze zur Regulierung von KI aufzuzeigen. Die Analyse soll auf bestehendem Schweizer Recht aufbauen und mögliche Regulierungsansätze für die Schweiz aufzeigen, die mit dem AI Act der EU und der KI-Konvention des Europarats (siehe dazu gleich nachstehend) kompatibel sind. Geprüft wird der Regulierungsbedarf mit besonderem Augenmerk auf die Einhaltung der Grundrechte. Berücksichtigt werden auch die technischen Standards sowie die finanziellen und institutionellen Auswirkungen der unterschiedlichen Regulierungsansätze. Der Bundesrat plant, 2025 einen konkreten Auftrag für eine Regulierungsvorlage KI zu erteilen.

Am 12. 7. 2024 hat die EU eine der ersten umfassenden Regulierungen zum Thema KI im Amtsblatt publiziert (KI-Gesetz oder AI Act). Der AI Act wird wie folgt Wirkung entfalten:

- 1. 8. 2024: Inkrafttreten des AI Acts.
- 2. 2. 2025: Verbot gewisser Praktiken bezüglich KI-Systemen mit inakzeptablem Risiko.
- 2. 8. 2025: Anwendbarkeit der Regeln zu Mehrzweck KI-Systeme und des Sanktionssystems.

- 2. 8. 2026: Restliche Bestimmungen des AI Acts treten in Kraft (mit Ausnahme der Nachstehenden).
- 2. 8. 2027: Restliche Bestimmungen bezüglich der Klassifizierung von KI-Systemen, die nicht in Anhang 3 des AI Acts als Hochrisiko-KI-Systeme qualifiziert werden und die diesbezüglichen Bestimmungen kommen zur Anwendung.

Der AI Act hat sowohl in Bezug auf die betroffenen Personen als auch auf die territoriale Reichweite einen breiten Anwendungsbereich. Er gilt nicht nur für Entwickler von KI-Systemen und Personen, die diese auf den Markt bringen oder in Betrieb nehmen (im Gesetz als «Anbieter» bezeichnet), sondern auch für alle Personen, die im Rahmen ihrer beruflichen Tätigkeit ein KI-System nutzen (als «Anwender» bezeichnet). Der AI Act erfasst Anbieter und Anwender innerhalb der EU, aber in gewissem Umfang auch solche ausserhalb der EU, die auf den EU-Markt zugreifen. Wer als Anbieter oder Anwender qualifiziert, tut gut daran zu prüfen, ob der AI Act auf sie Anwendung findet.

Der AI Act verfolgt einen risikobasierten Ansatz, und KI-Systeme werden danach in vier Kategorien eingeteilt:

1. KI-Systeme mit inakzeptablem Risiko. Diese sind verboten. Dazu gehören insbesondere biometrische Echtzeit-Fernidentifizierungssysteme («Facial Recognition») sofern diese nicht der Strafverfolgung dienen und unter

strengen Auflagen angewandt werden (z. B. vorgängige Genehmigung des Einsatzes durch eine Justizbehörde). Weiter fällt darunter Social Scoring, d. h. die Klassifizierung der Vertrauenswürdigkeit natürlicher Personen auf der Grundlage ihres sozialen Verhaltens oder bekannter oder vorhergesagter persönlicher Eigenschaften oder Persönlichkeitsmerkmale, sofern dies zu einer un gerechtfertigten oder unverhältnismässigen Schlechterstellung oder Benachteiligung dieser Personen führt.

2. KI-Systeme mit hohem Risiko sind unter strengen Auflagen (Risikomanagement, Daten-Governance, technische Dokumentation, Aufzeichnungspflichten, Transparenzanforderungen, menschliche Aufsicht sowie Anforderungen an Genauigkeit, Robustheit und Cybersicherheit, usw.) erlaubt. Dazu gehören z. B. KI-Systeme im Zusammenhang mit Bewerbungen, Beförderungen oder Kündigungen von Arbeitsverhältnissen, mit der Prüfung von Ansprüchen auf Sozialhilfe, mit der Bewertung von Schülern, mit Zulassungsprüfungen zur Universität oder mit Prognosen über die Rückfälligkeit von Straftätern.
3. KI-Systeme mit beschränkten Risiken sind unter limitierten Transparenzanforderungen (z. B. Information des Nutzers, dass er es mit einem KI-System zu tun hat, oder Kennzeichnung von mit künstlicher Intelligenz generierten Inhalten) sie zulässig.

Hilfe für unsere Bergbevölkerung schafft neue Perspektiven



Die Schweizer Patenschaft für Berggemeinden fördert die Solidarität mit der Schweizer Bergbevölkerung und mindert das Gefälle zwischen wohlhabenden und wirtschaftlich benachteiligten Regionen.

Mit projektbezogener Hilfe an unterstützungswürdige Gemeinden, Korporationen usw. hilft sie, dass unsere Bergregionen bewohnbar bleiben und aktiv bewirtschaftet und gepflegt werden. Die objektbezogenen Spenden werden ohne Spesenabzug den Begünstigten weitergereicht. Bei grösseren Vermächtnissen bietet sich die Möglichkeit, einen Fonds mit klar umschriebener Zweckbestimmung zu eröffnen.

PATENSCHAFT
BERGGEMEINDEN

Asylstrasse 74
8032 Zürich
Telefon

berggemeinden.ch
info@patenschaft.ch
044 382 30 80

Postkonto 80-16445-0
IBAN CH51 0900 0000 8001 6445 0



4. KI-Systeme mit keinem/niedrigem Risiko, z. B. im Zusammenhang mit Videospiele, sind zulässig und bleiben weitgehend unreguliert.

Im AI Act ist die Schaffung eines europäischen Ausschusses für künstliche Intelligenz vorgesehen, und auch auf nationaler Ebene sollen Behörden geschaffen werden. Diese werden fehlbare Unternehmen auch sanktionieren können. Nicht im AI Act enthalten sind Möglichkeiten zur individuellen Rechtsdurchsetzung.

Neben der EU hat auch der Europarat an einer KI-Konvention gearbeitet. Am 17. 5. 2024 erliess der Mi-

nisterrat die erst in englischer und französischer Sprache verfügbare «Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law» (KI-Konvention). Der Fokus der KI-Konvention ist die Sicherstellung des Einsatzes von KI in einer Weise, die die Grundrechte, die Demokratie und die Grundsätze der Rechtsstaatlichkeit achtet. Sie steht zur Unterzeichnung und Ratifizierung den Mitgliedstaaten und zudem jenen anderer Staaten offen, die sich an ihrer Ausarbeitung beteiligt haben. Die KI-Konvention wird in Kraft treten, sobald ihr fünf Staaten, wovon mindestens drei Mitgliedstaaten zugestimmt haben.

Anhang: Glossar

AI oder Artificial Intelligence	Siehe KI oder künstliche Intelligenz
AI Act oder KI-Gesetz	EU-Gesetz über die Regulierung von KI. (https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32024R1689)
Black Box Syndrome	Beschreibt die Unmöglichkeit bzw. die besondere Schwierigkeit, die Herangehensweise und den Lösungsweg von KI-Systemen zu überprüfen. Meist sind diese für User nicht und selbst für die Programmierer einer KI schwer überprüfbar.
Bias	Voreingenommenheit eines KI-Systems aufgrund des Trainingsdatensatzes, der Art und Weise des Trainings oder Modellierungsentscheiden der Programmierer
GPT	Generative Pre-trained Transformer
Halluzination	Angebliche (falsche) Fakten, die nicht auf realen Daten oder Ereignissen basieren, aber als solche dargestellt werden
Input	Daten, die in ein KI-System eingespeist werden
KI oder künstliche Intelligenz	Künstliche Intelligenz bezeichnet die Fähigkeit von Maschinen, Aufgaben auszuführen, die normalerweise menschliche Intelligenz erfordern würden. Dazu gehören Problemlösung, Lernen, Spracherkennung, Entscheidungsfindung und vieles mehr. KI-Algorithmen und -Systeme können Daten analysieren, Muster erkennen und darauf basierend Vorhersagen treffen oder Entscheidungen fällen.
LLM oder Large Language Model	Maschinelles Lernmodell, das Aufgaben im Bereich Natural Language Processing ausführen kann
NLP oder Natural Language Processing	Verarbeitung natürlicher Sprache durch ein KI-System
Output	Resultat einer generativen KI
Prompt	Input in eine generative KI mit den Instruktionen an das KI-System zur Erzeugung des Outputs