VISCHER

Künstliche Intelligenz und Recht.

Wann darf ich KI mit öffentlichen Inhalten trainieren?

David Rosenthal, Partner, VISCHER AG
31. Oktober 2024

Es wird bereits gestritten …

Worüber spreche ich heute?

- Training von **grossen Sprachmodellen (LLM)** (wie GPT-4o)
  - Auch das ETH AI Center will ein solches entwickeln
- Training mit **öffentlichen Inhalten**
  - Crawling aus dem Internet (direkt oder ab Sammlungen)
  - Publizierte Inhalte (z.B. Bücher)
- **Training**, nicht Gebrauch eines Sprachmodells
  - Sammeln und Aufbereiten der Inhalte (auch "pre-processing")
  - Eigentliches "Training" (auch "pre-training")
  - Danach: Fine-Tuning, Alignment
  - Abgrenzen: Anbieten des LLM (roh oder als Lösung)
  - Abgrenzen: Verwendung des LLM (roh oder als Lösung)

Das anwendbare Recht

- **Urheberrecht**, Datenschutzrecht, Lauterkeitsrecht, Strafrecht, Markenschutzrecht, Vertragsrecht …
  - Fokus in dieser Präsentation: Urheberrecht
- **Training in der Schweiz**: Es gilt Schweizer Urheberrecht
  - Es findet alles in der Schweiz statt, kein Zugang aus dem Ausland
  - Mit dem Training ist die relevante Handlung abgeschlossen
  - Ausländisches Urheberrecht eher nicht anwendbar
- **Anders:** Anbieten und Verwenden des LLM
  - Ausländisches Urheberrecht kann zur Anwendung kommen
  - Was ist zu diesem Zeitpunkt von den ursprünglichen Werken noch im LLM enthalten?

Ein kleiner Exkurs …

- Sprachmodelle speichern nicht die Trainingsdaten als solche, sondern nur den **"Durchschnitt" der Erkenntnisse** daraus
- Mit jedem Text können die Milliarden von "Schrauben" des Modells etwas besser **justiert**
  - Damit es besser in der Lage ist, neue Texte zu generieren
  - Texte generiert es aber nur in Kombination mit einem Prompt
- Wichtig: Wortwörtliche und sinngemässe **Memorisierung**
  - Wortwörtlich: Text so oft, dass das Modell ihn sich gemerkt hat
  - Sinngemäss: Viele Texte beschreiben immer wieder denselben Sachverhalt, so dass das Modell ihn sich gemerkt hat
  - Beides kann sich auf ein urheberrechtlich relevantes Werk beziehen (z.B. Slogan, berühmte Geschichte)

Und Art. 53 Abs. 1 Bst. c EU AI Act?

- Verweis auf **Text & Data Mining-Regel** des EU-Urheberrechts
  - Pflicht der Provider von Allzweck-KI-Modellen: "put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790"
  - Erwägung 106: "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." (Beachte: AI Act will Urheberrecht nicht ändern)
- **Aber:** Das EU-Urheberrecht will auf Handlungen in der Schweiz gar nicht angewendet werden – auch die Berufung auf die TDM-Ausnahme wäre in der Schweiz gar nicht möglich …

Schweizer Urheberrecht: Analyse

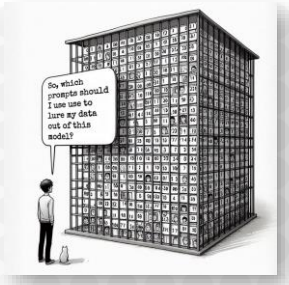
- **Liegt eine urheberrechtlich relevante Handlung vor?**
  - Wenn kein Werkgenuss ermöglicht wird?
  - Wenn das Werk zerlegt wird und nicht mehr als solches besteht?
  - Wenn es verblasst oder es einen inneren Abstand gibt?
- **Falls ja: Ist das Training trotzdem rechtmässig?**
  - Liegt eine Zustimmung des Rechteinhabers vor?
  - Greift eine der Schrankenbestimmungen?
- **Verletzung des Urheberrechts durch den Input für die KI**
- **Verletzung des Urheberrechts durch den Output der KI**

Trainieren = relevante Handlung?

- Falls nicht, dann ist das Urheberrecht **nicht verletzt**
- Technische Vervielfältigungsvorgänge finden statt
- Aber dienen diese wirklich dazu, den **menschlichen Genuss** des Werks zu ermöglichen?
  - In der Regel nicht (Ausnahme ggf. bei Memorisierung)
  - Folge: Keine relevante Handlung
- Führen sie zur Erstellung einer **Kopie des Werks im Modell**?
  - Nein, denn es liegt nur noch in seine Bestandteile aufgelöst vor, auch wenn diese durch unsichtbare Fäden verbunden sein mögen (sie sind vom Menschen nicht mehr wahrnehmbar)
  - Folge: Die Zerlegung des Werks ist keine relevante Handlung (jedoch können solche dem vorausgegangen sein)

Trainieren = relevante Handlung?

- Und wenn dem nicht gefolgt wird, etwa weil geltend gemacht wird, dass eine Memorisierung vorliegt?
  - Der Abruf setzt einen passenden **Prompt** voraus, denn nur damit kann aus einer Memorisierung ein wahrnehmbarer Inhalt werden
  - Der Inhalt eines Werks ist im LLM mit den Inhalten vieler anderer Werke vermischt, wie in einer **Suppe von Puzzle-Teilchen**
  - Urheberrechtliche Folgerung: Das Werk "verblasst" in der Menge; sein Vorkommen im Modell ist damit urheberrechtlich irrelevant
  - Es besteht auch ein "innerer Abstand" zwischen dem Originalwerk und der Memorisierung, weshalb letztere urheberrechtlich frei ist
- Ferner: Wenn der menschliche Werkgenuss urheberrechtlich frei ist, muss nicht auch der "Werkgenuss" durch ein LLM frei sein, da der Vorgang vergleichbar ist?



Zustimmung?

- Liegt eine **Zustimmung des Rechteinhabers** vor?
  - Fall A: Eine ausdrückliche Lizenz oder sonstige Erlaubnis liegt vor
  - Fall B: Es wird nichts gesagt, aber ein Werk wird frei zugänglich im Internet publiziert
  - Fall C: Ein Werk wird frei zugänglich im Internet publiziert, aber mit dem Hinweis, dass es nicht für Trainings benutzt werden darf
  - Fall D: Ein Werk wird als Buch sonst publiziert (nicht online)
  - Womit musste der Rechteinhaber rechnen?
  - Muss jemand, der im Internet Werke publiziert damit rechnen, dass sie gelesen und inhaltlich ausgewertet werden, einschliesslich der dabei vorkommenden Vervielfältigungen?
  - Falls ja: Mangels Widerspruch liegt eine implizite Zustimmung vor

Aus einer Lizenzvereinbarung

- use the Subscribed Products or parts thereof with a closed artificial intelligence tool including to train an algorithm, test, process, analyse, generate output and/or develop any form of artificial intelligence tool in as much as this will not result in the Subscribed Products, parts thereof or data directly derived from it becoming available to third parties.

  An artificial intelligence tool is deemed to be closed, either:

  (1) if it is self-hosted in an on-premises environment or in an environment hosted externally solely for use by Participating Institutions or Authorised Users; or

  (2) if third-party providers of the artificial intelligence tool renounce to use the provided input (specifically the Subscribed Products, parts thereof or data derived from it) to train or expand the capabilities of the tool.

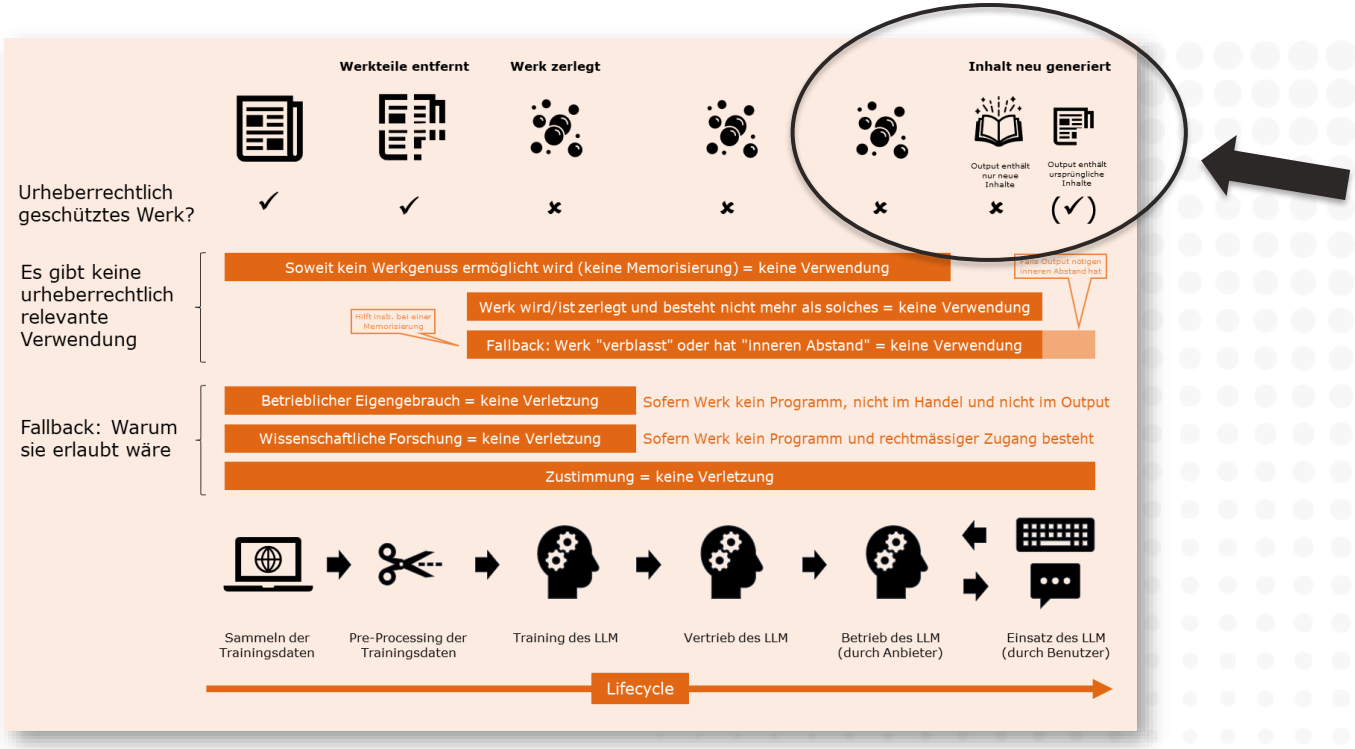
Schrankenbestimmungen?

- **Zitatrecht**: Eher Nein
  - Zweck passt eher nicht, und Quellenhinweis fehlt oft auch
- **Vorübergehende Vervielfältigung**: Nein
  - Nicht für die Übertragung in einem Netz
  - Eigenständige wirtschaftliche Bedeutung
- **Betriebsinterne Information oder Dokumentation**: Möglich
  - Mitarbeiter liest einen Aufsatz, um einen Klienten zu beraten
  - LLM liest einen Aufsatz, um sein Wissen zu verbessern
  - Weitgehend vollständige Kopie eines im Handel erhältlichen Werks ist aber nicht erlaubt, weshalb besondere Techniken nötig sind (z.B. Goldfish-Loss)

Schrankenbestimmungen?

- **Wissenschaftsschranke**: Ja
  - "Zum Zweck der wissenschaftlichen Forschung ist es zulässig, ein Werk zu vervielfältigen, wenn die Vervielfältigung durch die Anwendung eines technischen Verfahrens bedingt ist und zu den zu vervielfältigenden Werken ein rechtmässiger Zugang besteht."
  - Wissenschaftliche Forschung = systematische, methodische Suche nach neuen Erkenntnissen, gleich welcher Disziplin
  - Passt gut zum Training von LLM (darf sogar kommerziell sein)
  - Kein Opt-out wie im EU-Recht, aber Zugang muss rechtmässig sein (vertragliche Abreden können unwirksam sein [strittig])
  - Gilt nicht für Computerprogramme (wichtig!)

Zwischenergebnis



Verletzung durch den Output?

- Erst der **Prompt des Benutzers** und die Reaktion des Modells bestimmen darüber, ob eine Urheberrechtsverletzung vorliegt
  - OK: "Liefere mir eine Analyse zu Harry Potter"
  - Nicht OK: "Verfasse mir Band 8 von Harry Potter."
- Ist der Entwickler des Modells dafür **mitverantwortlich**?
  - Nicht eine Frage des Trainings, sondern ob die Bereitstellung des Modells (oder passender SaaS) als eine **Teilnahme** an der Urheberrechtsverletzung (z.B. Gehilfenschaft) gilt
  - Analoge Anwendung der Rechtsprechung zu **Suchmaschinen**?
    - HGer Zürich HG220030-O (21.8.24): Suchmaschine nicht für die besonderen Suchbegriffe der Benutzer mitverantwortlich
    - BGE 145 III 72: Ein genug enger Zusammenhang ist nötig

Verarbeitung von Werken *durch* Sprachmodelle

- **Beispiel:** Professorin XY will lizenziertes Lehrbuch in ein generatives KI-Tool einspielen, damit die Studierenden in der Folge mit dem Tool Fragen zum Inhalt des Buchs stellen können
- Mögliche Rechtsgrundlagen im Schweizer Urheberrecht?
  - Allenfalls gar **keine urheberrechtliche Handlung**
  - **Lizenzbedingungen**, welche diese Nutzung zulassen
  - Verwendung der **Lehrperson für Unterricht in Klasse***
  - Vervielfältigung für **interne Information** oder **Dokumentation**, sofern der Output genügend zusammenfassend und umformuliert ist, dass er selbst nicht dem Werk entspricht*
  - **Wissenschaftschranke**: Was wäre die Forschungserkenntnis?

* Das Buch dürfte der KI nur auszugweise gefüttert werden

Crawler-"Verbote" rechtswirksam?

- **Unterscheiden:**
  - Widerspruchserklärung
  - Technische Massnahmen, die ein "Absaugen" verhindern
  - Login mit Nutzungsbedingungen
- **Widerspruch** zerstört die Vermutung der Zustimmung und im EU-Recht verhindert er die wirtschaftliche TDM-Schranke
  - Unklar ist, ob Widerspruch maschinenlesbar sein muss oder nicht
  - Er steht aber den anderen Rechtsgrundlagen **nicht entgegen**
- **Anders:** Vereinbarte Nutzungsbedingungen
  - Gegen die Wissenschaftsschranke können sie unter Umständen aber ebenfalls wirkungslos sein, da rechtmässiger Zugang genügt

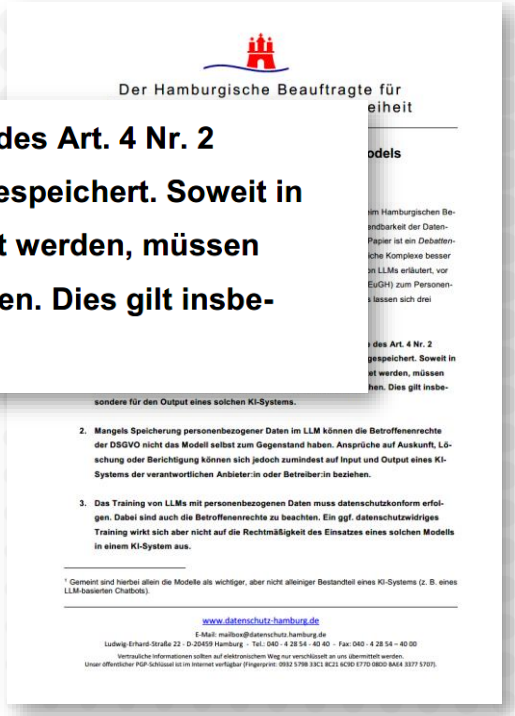
Und der Datenschutz?

- Mit Memorisierung muss gerechnet werden, ist aber in der Regel **unproblematisch** ("Geburtstag von Donald Trump")
  - **Zweckbindung:** Ohne Memorisierung ist es ein kompatibler Zweck, mit Memorisierung müssen die Personen damit rechnen, weil sie die Folge der öffentlichen Verbreitung der Daten ist
  - **Verhältnismässigkeit:** Je mehr, desto besser für das LLM; umstritten ist die Frage, ob Trainingsinhalte aufzubewahren sind
  - **Transparenz:** Es muss damit gerechnet werden, trotzdem ist eine Datenschutzerklärung zu empfehlen
  - **Richtigkeit:** Das Modell nimmt das auf, was es sieht, d.h. es ist im Hinblick auf den Zweck richtig; Halluzinationen sind irrelevant
  - **Rechtfertigung:** Ausser bei öffentlichen Personen greift oft der Rechtfertigungsgrund der nicht personenbezogenen Bearbeitung

Datenschutzbehörden haben noch ihre Mühe …

- Hamburg: LLM enthalten keine Personendaten

1. **Die bloße Speicherung eines LLMs stellt keine Verarbeitung im Sinne des Art. 4 Nr. 2 DSGVO dar. Denn in LLMs werden keine personenbezogenen Daten gespeichert. Soweit in einem LLM-gestützten KI-System personenbezogene Daten verarbeitet werden, müssen die Verarbeitungsvorgänge den Anforderungen der DSGVO entsprechen. Dies gilt insbesondere für den Output eines solchen KI-Systems.**


Der Hamburgische Beauftragte für

sondere für den Output eines solchen KI-Systems.

2. Mangels Speicherung personenbezogener Daten im LLM können die Betroffenenrechte der DSGVO nicht das Modell selbst zum Gegenstand haben. Ansprüche auf Auskunft, Löschung oder Berichtigung können sich jedoch zumindest auf Input und Output eines KI-Systems der verantwortlichen Anbieter:in oder Betreiber:in beziehen.

3. Das Training von LLMs mit personenbezogenen Daten muss datenschutzkonform erfolgen. Dabei sind auch die Betroffenenrechte zu beachten. Ein ggf. datenschutzwidriges Training wirkt sich aber nicht auf die Rechtmäßigkeit des Einsatzes eines solchen Modells in einem KI-System aus.

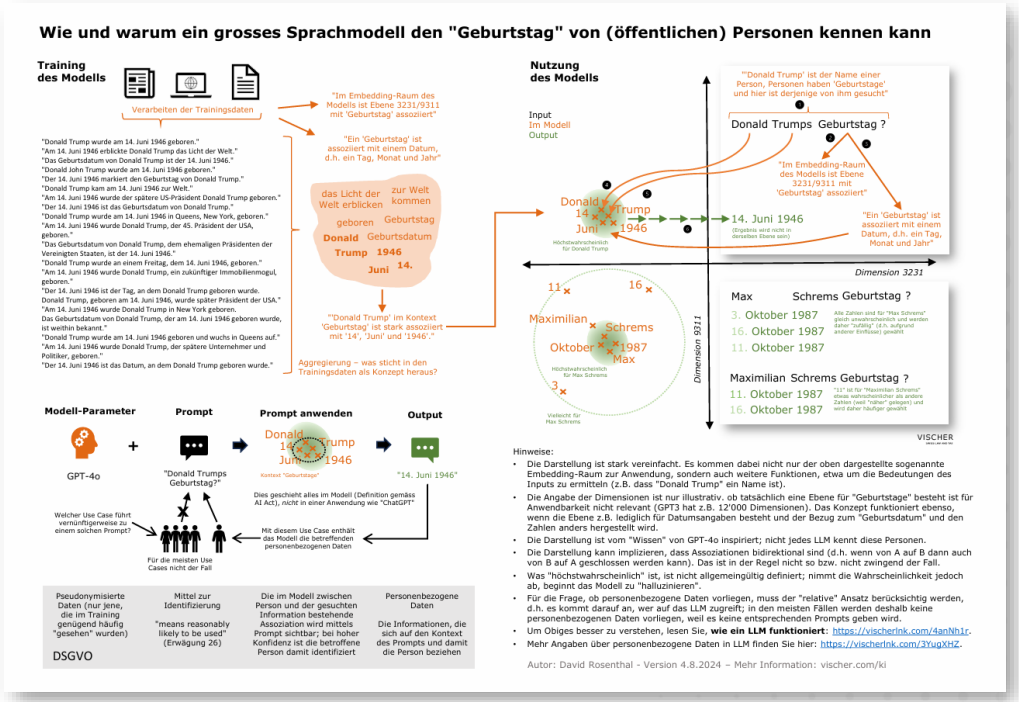
[1] Gemeint sind hierbei allein die Modelle als wichtiger, aber nicht alleiniger Bestandteil eines KI-Systems (z. B. eines LLM-basierten Chatbots).

www.datenschutz-hamburg.de

http://www.datenschutz-hamburg.de/

Nur weil die Personendaten mit blossem Auge nicht erkennbar und Outputs probabilistisch erzeugt sind, ist es falsch zu folgern, dass ein LLM solche nie enthalten kann (weicht vom Urheberrecht ab, weil es hier um von der Form losgelöste Informationen geht)

Wer es wissen will …



Wie funktioniert ein grosses Sprachmodell und was ist wirklich darin gespeichert?

Wie funktioniert ein neuronales Netzwerk?

vischerlnk.com/4anNh1r

vischerlnk.com/3zAoqeb

Fazit

- Auch öffentliche Inhalte sind **nicht einfach frei**
  - Es gibt diverse gesetzliche Bestimmungen, die zu beachten sind beim Training von KI-Modellen
  - Der **Datenschutz** macht für einmal weniger "Probleme" …
- Das **Urheberrecht** bietet schon heute Ansätze, um jedenfalls das Training von grossen Sprachmodellen möglich zu machen
  - Keine urheberrechtlich relevante Handlung bzw. aufgrund einer relevanten Schrankenbestimmung erlaubt
- **Abgrenzen:** Verarbeitung von Inhalten durch Sprachmodelle
- Legislative oder judikative **Klarstellungen** im Urheberrecht zugunsten des Trainings von KI-Modellen wären sinnvoll – auch für den **Standort Schweiz**

VISCHER

Danke für Ihre Aufmerksamkeit!

Fragen: david.rosenthal@vischer.com

Zürich
Schützengasse 1
Postfach
8021 Zürich, Schweiz
T +41 58 211 34 00

www.vischer.com

Basel
Aeschenvorstadt 4
Postfach
4010 Basel, Schweiz
T +41 58 211 33 00

Genf
Rue du Cloître 2-4
Postfach
1211 Genf 3, Schweiz
T +41 58 211 35 00

Mehr Unterlagen:
www.vischer.com/ki
www.rosenthal.ch