Liebe Mitglieder des Netzwerkes der ehemaligen SSAB, liebe Interessierte

Gerne blicke ich auf die interessante Tagung vom 21.3.2024 zurück. Vielen Dank allen, die teilnahmen, online oder vor Ort. Besonders freuten mich die Gespräche mit Mitgliedern aus dem Netzwerk der ehemaligen SSAB, die ich am Rand der Tagung in Zürich führen konnte.

In den kommenden E-Mails leuchte ich weiterhin interessante KI-Aspekte aus.

**Werden KI-Sprachmodelle wegen KI-generierter Daten schlechter?**
Das Internet füllt sich immer mehr mit KI-generierten Texten. Was bedeutet dies für das Trainieren von KI-Sprachmodellen? Zwei Publikationen, die diese Frage thematisieren, kommen zu folgender Schlussfolgerung: Wenn ein Trainingsdatensatz eines KI-Sprachmodells zu viele von der KI generierte Texte enthält, werden die Ergebnisse mit dem wiederholten Training immer schlechter. Der Titel der ersten Publikation ist vielsagend: The Curse of Recursion: Training on Generated Data Makes Models Forget. (Fluch der Rekursion. Training mit generierten Daten lässt Modelle vergessen, beziehungsweise macht sie funktionsunfähig). Hinter der Abhandlung steht ein Autorenteam vor allem britischer Universitäten. Zu einer ähnlichen Schlussfolgerung kommt eine zweite Publikation aus England, die Modelle mit Bildern trainierte. Die Ergebnisse sind veröffentlicht unter dem Titel: Towards Understanding the Interplay of Generative Artificial Intelligence and the Internet.

**Ist das wirklich möglich?**
Die von der KI generierten Texte oder Bilder lassen sich ja kaum von jenen unterscheiden, die Menschen erstellten. Das Problem liegt gerade darin, dass es bisher keine sichere Methode gibt, auch nicht mit einer Art von «Wasserzeichen», um die von der KI generierten Texte von anderen zu unterscheiden. Was geht hier vor – gibt es eine Art Rückkopplungseffekt? Ben Lutkevich erklärt hier (mit Erklärvideo, 1', auch direkt auf YouTube) das Phänomen auf WhatIs?, einer Site des US-amerikanischen Unternehmens TechTarget. Lutkevich gibt am Schluss den Ratschlag, für das Training vorläufig auf Daten vor 2023 zurückzugreifen, bevor das Internet durch KI-generierte Daten «verschmutzt» wurde (Data pollution).

**Tech-Giganten sichern sich hochwertige, von Menschen erstellte Daten**
Die Tech-Giganten sind offenbar bereit, auf Druck für die Nutzung von Daten zu bezahlen. OpenAI, wie andere Tech-Giganten, hatten für das Training ihrer KI-Sprachmodelle vorerst ungefragt die im Internet verfügbaren Daten genutzt. Damit stellen sich Fragen des Urheberrechts und einer angemessenen Entschädigung für die Nutzung der Daten. Von besonderem Interesse sind die hochwertigen Daten von Medienhäusern. Bereits im Dezember 2023 äusserte sich OpenAI im britischen House of Lords (4. Punkt) dahingehend, dass es unmöglich sei, KI-Sprachmodelle ohne urheberrechtlich geschütztes Material zu trainieren, weshalb OpenAI Partnerschaften mit Medienhäusern eingehe, die für beide Seiten vorteilhaft seien. Tatsächlich schlossen OpenAI und die US-Nachrichtenagentur Associated Press (AP) bereits im Sommer 2023 eine Partnerschaft ab, die OpenAI den Zugang ins AP-Archiv zurück bis 1985 gewährt. Im Dezember 2023 kam es zu einer Partnerschaft des deutschen Medienkonzerns Axel-Springer mit OpenAI. Wie viel OpenAI bezahlte, wurde nicht bekannt.

Offenbar konnten sich die New York Times und OpenAI nicht einigen. Wie die britische Nachrichtenagentur Reuters am 27.2.2024 mitteilte, hat das Medienhaus Klage eingereicht. Seitdem sind die beiden im Rechtsstreit. Eine erfolgreiche Klage dürfte weitere, kleinere Medienhäuser ermutigen, dem Beispiel zu folgen. Tatsache ist, dass die KI-Anwendungen für die Tech-Giganten ein Riesengeschäft sind. Gemäss New York Times vom 16.2.2024 verdreifachte sich der Firmenwert von OpenAI innert zehn Monaten auf 80 Milliarden Dollar.

**Auch andere Tech-Giganten in der Bredouille**
Aber auch andere Tech-Giganten kommen unter Druck. Gemäss der ARD-Tagesschau vom 28.2.2024 verklagten 32 Medienunternehmen aus 17 europäischen Ländern Google auf 2,3 Milliarden Euro Schadenersatz für Verluste bei Werbeeinnahmen. Google machte 2022 mit Online-Werbung einen Umsatz von 224,5 Milliarden.

Mit freundlichen Grüssen
Für das Netzwerk der ehemaligen SSAB: Hanna Muralt Müller

***Neues Datenschutzrecht: Falls Sie diese E-Mail nicht mehr erhalten möchten, melden Sie sich bitte bei mir!***