Nur noch bunte Hasen:

Wird die KI bald kollabieren?

**Schon mehr als die Hälfte der Texte im Internet sind nicht mehr von Menschen geschrieben.Aber wenn man neue Generationen von KI-Modellen mit ihrem eigenen Output trainiert, werden sie rasch unbrauchbar. Menschliche Erzeugnisse gewinnen deshalb an Wert.**

eingesetztes-Bild.tiff

*Von Karl Gaulhofer*

Ach wie schön! In der englischen Grafschaft Somerset gibt es in Dutzenden kleinen Dörfern mittelalterliche Kirchtürme, so hoch und reich verziert wie bei Kathedralen. Wie kommen die dort hin? Wir lesen aus einem Wikipedia-Eintrag zu den „Somerset Towers“, dass es zweiTheorien gibt: Vielleicht hat sie ein einziger regionaler Baumeister entworfen, vielleicht aber auch berühmte Architekten, die . . . Hier bricht der Text leider ab. Also fragen wir Chat GPT oder ein anderes großes Sprachmodell, wie er weitergehen könnte. Und müssen lesen: „Hier gibt es eine der weltweit größten Populationen von Hasen mit schwarzen Schwänzen, Hasen mit weißen Schwänzen, Hasen mit blauen Schwänzen . . .“ – und gelbe und rote. Was soll derUnsinn? Was haben Kirchtürme mit einer bizarren Ansammlung bunter Hasen zu tun? Nichts, außer dass sie in dieser Kombination – als Experiment aus einer Studie – ein Menetekel für einen Kollaps des Internets durch den Fluch der KI sind.

Mit Stand Oktober sind bereits mehr als die Hälfte (52 Prozent) aller Texte im Internet zum größeren Teil von einer künstlichen Intelligenz verfasst, nach einer Recherche von Graphite,einer Agentur für Digitalmarketing. Nach einer anderen Studie vom April enthalten knapp dreiViertel aller neu erstellten Webseiten zumindest teilweise KI-Texte. Die Folge: Die selbstlernenden Sprachmodelle, von denen die Hersteller stolz verkünden, dass sie mit allen im Netz verfügbaren Texten trainiert worden sind, werden in künftigen Versionen häufiger mit Content gefüttert werden, den sie selbst erstellt haben. Sie werden durch diese Nabelbeschau nicht besser, sondern schlechter, bis sie womöglich kollabieren und nur noch sinnloses Zeug lallen. Aber warum?

Das Beispiel stammt aus einer „Nature“-Studie vom Juli vorigen Jahres. Die Autoren, Forscher der **Universität Oxford**, warnten darin als Erste vor einem „model collapse“. Sie trainierten ein Sprachmodell mit Wikipedia-Artikeln und stellten ihm dann die Aufgabe, für kurzeTextabschnitte nach der wahrscheinlichsten Fortsetzung zu suchen. Wären die Modelle völlig fehlerfrei, hätten sie die den gesamten Korpus eins zu eins repliziert. Aber sie machten Fehler, und als man sie mit dem eigenen kontaminierten Output trainierte, häuften sich diese. Zudem wurden die Texte immer repetitiver. In Generation neun landete man bei den Hasen.

Der Vergleich zur Inzucht

Das erinnert an das Spiel „Stille Post“. Beliebter, weil bedrohlicher klingend, sind Vergleiche aus der Biologie. Man denke an die „Inzuchtdepression“: Eine Population, in der man sich nur noch mit nahen Verwandten kreuzt, verliert ihre genetische Vielfalt, wird anfälliger fürErbkrankheiten und droht irgendwann auszusterben. Klingt suggestiv, aber wie lässt sich das auf die KI umlegen? Sprachmodelle errechnen Wahrscheinlichkeiten für das jeweils nächsteWort in einem Text. Diese folgen typischerweise einer Normalverteilung, also einerGlockenkurve: innen die häufigen Wörter, außen die seltenen. Wenn man eine KI mit ihrem eigenen Output füttert, gehen zunächst die Fälle mit geringerer Häufigkeit verloren, die auf den auslaufenden Seiten der Kurve liegen. Alles spitzt sich am Scheitelpunkt zu, die Varianz nimmt ab, die Vielfalt der Wirklichkeit wird nicht mehr abgebildet. Auf den Prompt „Zeichne einen Hund“ liefert eine Grafik-KI nur noch Golden Retriever, weil die besonders beliebt sind, und „vergisst“, dass es auch Pudel und Dackel gibt.

Das kann fatale Folgen haben, wenn man etwa KI zur Diagnose in der Medizin einsetzt, denn hier sind die seltenen Fälle gerade die interessanten und gefährlichen. Verständlich ist auch, dass die KI durch die Konzentration auf die häufigsten Fälle immer einförmiger antwortet, derZusammenhang verloren geht. Aber warum lallt die KI nicht „Kirchtürme, Kirchtürme,Kirchtürme“, sondern „Hasen, Hasen, Hasen“? Das hat mit den Fehlern zu tun, die sie macht, die sie auf eine falsche Fährte locken und zu etwas führen, was mit dem ursprünglichenKontext nichts mehr zu tun hat.

So viel Mensch muss sein

Wie wahrscheinlich ist ein Kollaps? Darüber wird in der KI-Forschung heftig debattiert. Dabei ist man sich im Prinzip einig: Wird ein selbstlernendes Modell ausschließlich mit eigenemOutput trainiert, kollabiert es. Die Frage ist: Wie viele menschliche Artefakte muss man gezielt beimischen, damit es leistungsfähig bleibt? Jedenfalls bleibt hier ein praktisches Problem: Es wird (derzeit noch) immer schwieriger, menschliche von KI-Erzeugnissen zu unterscheiden.Die einzige sichere Lösung wäre eine Art „Wasserzeichen“ für KI-Inhalte. Aber selbst wenn sich alle großen Hersteller, wie Open AI, Google Gemini oder Meta, darauf einigen könnten, blieben immer noch die „Schurkenstaaten“, die durch KI Lügen verbreiten wollen und sich deshalb nie an eine Kennzeichnungspflicht halten würden.

Fest steht: Von echten Menschen erzeugte Texte, Bilder und Melodien werden wertvoller. DieKI-Hersteller beeilen sich gerade, Lizenzverträge abzuschließen, um original menschlicheErzeugnisse legal nutzen zu können. Denn der Preis steigt: Im September zahlte das KI-Unternehmen Anthropic einer Gruppe von Schriftstellern 1,5 Milliarden Dollar, um einer Urheberrechtsklage zu entgehen. Die Vergleichssumme deckt 500.000 Bücher ab, das einzelne Exemplar wurde also mit 3000 Euro bewertet. Stolze Summen. Die auch uns verunsicherten Menschenkindern ein wenig Stolz und Selbstbewusstsein zurückgeben: Ohne unser Schaffen würden unsere künstlichen Avatare schon bald völlig verblöden. Blaue Hasen, rote Hasen, gelbe Hasen . . .

*Quelle: Die Presse vom 7.11.2025, Foto: Getty*