

Wenn die KI Donald Trump verhaftet

KÜNSTLICHE INTELLIGENZ Forscher aus der Region warnen vor unkontrollierter Verbreitung von Fake News durch die immer perfekteren Computermodelle.

VON ALEXANDER JUNGKUNZ

NÜRNBERG - Warum denkt man an Goethes Ballade vom „Zauberlehrling“, wenn man die Debatte um Künstliche Intelligenz verfolgt? Weil es der Forschung gehen könnte wie jenem Jung-Magier, der die Abwesenheit seines Lehrmeisters nutzt. Der Zauberlehrling schickt bekanntlich den Besen zum Wasserholen - und kann ihn nicht mehr stoppen. Bis der alte Meister die erlösende Formel spricht und der Spuk endet.

Nur: So ein alter Meister ist bei der Künstlichen Intelligenz (KI) nicht in Sicht. Dabei klang eine erstaunliche Forderung eines breiten Bündnisses von KI-Fachleuten kürzlich wie der flehentliche Ruf nach so einem Einhalt gebietenden Zauberer: Eine illustre Runde namhafter Digitalexperten, darunter Elon Musk, Apple-Gründer Steve Wozniak und der Historiker Yuval Noah Harari, forderte ein sechsmonatiges Moratorium bei der KI-Weiterentwicklung. Darin ist von einem „außer Kontrolle geratenen Wettlauf“ und einer Entwicklung die Rede, die mittlerweile „niemand – nicht einmal ihre Schöpfer – verstehen, vorhersagen oder zuverlässig kontrollieren“ könne.

Der Geist ist aus der Flasche

Aber ist der Geist wieder in die Flasche zu bekommen, die er in Form von Sprachmodellen wie ChatGPT und anderen KI-Anwendungen längst verlassen hat? Der frühere Chef des Ethikrats hält dagegen. Peter Dabrock, Theologe und Ethik-Experte an der Uni Erlangen: „So viel Naivität will man diesen sicher mit allen Wassern gewaschenen Menschen nicht unterstellen, als dass sie glauben könnten, auf ihren Appell hin klappten alle Programmierer für sechs Monate ihre Laptops zu.“ Auch



Ute Schmid, Uni Bamberg

Ute Schmid, Leiterin des Lehrstuhls für Kognitive Systeme an der Universität Bamberg hält so einen Forschungsstopp für nicht realisierbar. Dennoch schloss sie sich den Unterzeichnern des Appells an: „Ich glaube, es ist höchste Zeit, dass wir Expertinnen und Experten aus der KI-Forschung, Technologie-Unternehmen und Akteurinnen und Akteure aus vielen Gesellschaftsbereichen mit Entscheidungsträgern aus der Politik zusammenbringen. Dabei wäre zu debattieren: Was erwarten wir in den jeweiligen Anwendungsbereichen von der KI in Arbeit, Bildung, Journalismus oder Medizin, was muss reguliert werden?“

Auch Peter Dabrock deutet das Moratorium als Aufruf für so eine konzertierte Aktion: Es sende das Signal an „Wissenschaft, Technik, Wirtschaft, Politik, Bildungsinstitutionen und Öffentlichkeit: Realisiert, was ihr da in der Hand habt, schmiedet mächtige Allianzen, um möglichst früh Gestaltungsmacht zu behalten!“ Der Erlanger Ethiker und die Bamberger KI-Expertin teilen eine Befürchtung: dass die inzwischen gut funktionierenden Sprachmodelle der KI eine „unkontrollierbare Überschwemmung mit Propaganda und Unwahrheiten“ ins Netz einspeisen (Ute Schmid). Ganz ähnlich



Echt? Natürlich nicht - aber wer nicht weiß, dass es keine Verhaftung Donald Trumps auf offener Straße gab, nimmt dieses von KI gestaltete Foto womöglich für bare Münze.

erwartet Dabrock „eine maliziöse Flutung von schwer überprüfbar Fake-Informationen via Schrift, Videos und Bildern“.

Diese Befürchtung begründet sich durch die Art und Weise, wie die Technologie arbeitet. Ute Schmid: „Es wird viel leichter, gezielt Propaganda oder Fake News zu verbreiten – es reicht, ChatGPT oder einem anderen Sprachmodell einen entsprechenden Auftrag zu geben, dann generiert die KI Texte oder auch Bilder und Filme dazu. Die werden dann wiederum in den Fundus von ChatGPT eingespeist – so kann ich leicht eine Lawine von Desinformation losstreuen. Stellen wir uns das vor beim Blick auf Trump und dem Marsch aufs Capitol oder auf Fake-News-Kampagnen gegen Impfempfehlungen – dann ist das eine Vollkatastrophe.“

Je mehr Fehlinformation also in der KI eingespeist ist, desto mehr bedient sie sich auch daraus. Weil sie ja nicht weiß, was sie inhaltlich liefert, sondern nur, welche Sätze und Wort-Kombinationen am erwartbarsten sind als Antwort auf die ihr gestellten Fragen und Aufgaben. Und sich da nur an der Quantität orientiert, nicht an der Qualität von Texten – die sie gar nicht beurteilen kann. Ute Schmid sieht in all den Selbstversuchen, bei denen Schüler, Studierende und viele andere ChatGPT ausprobierten und über schräge bis falsche Ergebnisse staunten, auch eine erfreuliche Erkennt-

nis: „Anfangs ließen sich viele Menschen von den Aufgaben bluffen. Inzwischen haben viele verstanden, dass ChatGPT aufgrund seiner Funktionsweise nicht dafür gemacht ist, faktenreue Auskünfte zu geben. Wenn wir von Chat GPT eine Antwort kriegen, dann können wir nicht beurteilen, ob das wahr ist oder nicht, Fakt oder Meinung.“ Das müsse erst gecheckt werden – von Menschen.

Täuschend echte Fotos

Seit ein paar Wochen kursieren beeindruckende Ergebnisse im Netz, wie KI künstliche Fotos generieren kann. Populär wurde Papst Benedikt in einer weißen Daunenjacke. Oder ein Bild, das die vermeintliche Festnahme von Donald Trump auf offener Straße zeigt. Oder Angela Merkel und Obama am Strand. Alles Deep Fakes. Momentan lässt sich an kleinen Fehlern noch erkennen, dass es Fälschungen sind – je mehr die KI lernt, desto besser werden wohl bald aber auch diese Fotos.

Was hilft da? Ute Schmid setzt auf ein vermeintlich altmodisches Mittel: „Qualitätsjournalismus, der prüft und gewichtet. Seriösen Medien vertraue ich, dass diese keine Fake News liefern und falls fehlerhafte Information berichtet wird, dies im Nachgang offengelegt und richtig gestellt wird. Wenn Qualitätsmedien nicht davon berichten, dass sich Merkel und Obama am Strand trafen, dann kann ich mir relativ sicher sein, dass

das Foto ein Deep Fake ist.“ Dabrock befürchtet da allerdings eine wachsende Kluft zwischen denen, die sich Qualitätsmedien – deren Angebote nicht umsonst zu haben sind – leisten können und wollen, und anderen, die dafür nicht zahlen können und/oder auch mit den nicht kontrollierbaren Ergebnissen aus dem Netz zufrieden sind. Zudem drohe „das Verschwinden des Autors“, wenn Texte von der KI stammen und wenig zuverlässig sind. Das ist problematisch, weil der Autor für Zurechenbarkeit und seine Verantwortung steht.

Der Ethiker sieht eine gewaltige Zäsur – das Ende unseres gewohnten Umgangs mit Texten. Wenn die KI einigermaßen taugliche Texte liefert, würden manche darauf verzichten, Eigenes zu verfassen – und diese zentrale Kulturtechnik dann auch verlernen, mit noch unabsehbaren Folgen gerade für das Bildungssystem. Dabrock: „Alle Träger von Bildungs- und Kompetenzvermittlung müssen realisieren, dass diese Sprachmodelle einen Epochenbruch darstellen und wir umgehend lernen müssen, diese große Zeitenwende kritisch und konstruktiv zu gestalten.“

Weder Dabrock noch Ute Schmid plädieren für einen – ohnehin kaum möglichen – Stopp von KI, im Gegenteil. Die Bamberger Expertin sagt: „Es ist beeindruckend, was wir mit KI leisten können. Das kann ein großer Gewinn in vielen Bereichen sein –

Medizin, Bildung, Industrie, zur Entlastung von Menschen bei Routine-Arbeiten.“ Deutschland müsse da dran bleiben an der Entwicklung. „Wir brauchen Anschluss an die Technologie. Sonst sind wir Kunden anderer Länder und Konzerne, das sollten wir nicht aus der Hand geben.“ Zen-



Peter Dabrock, Uni Erlangen

tral sei nun aber die breite Debatte mit allen gesellschaftlichen Gruppen über Chancen, Risiken und Regeln von und für KI. Ute Schmid: „Wenn wir jetzt die Weichen falsch stellen, kann das schlimme Konsequenzen haben.“ Sie wolle „nicht zu schwarz malen“, vergleicht die aktuelle Zeitenwende aber mit der Phase, als die Atomwissenschaftler erleben mussten, dass ihre Forschung nicht nur „guten“ Zwecken dienen kann, sondern auch zur Atombombe führte.

Politische Instabilität als Gefahr

Diesen Vergleich zieht auch einer der führenden Experten im Bereich Deep Learning, Yoshua Bengio. Unkontrolliert könnten KI-Anwendungen Menschen falsch beraten, was zu problematischen Fehlentscheidungen führen könnte – vom Nichtgewähren einer Sozialdienstleistung bis zur medizinischen Fehlentscheidung. Unreguliert könnten KI-Systeme wie ChatGPT im schlimmsten Fall sogar zu politischer Instabilität beitragen. Laut Ute Schmid brauche es, um dies zu verhindern, Forschung an KI-Methoden, die maschinelles Lernen mit wissensbasierten Ansätzen kombinieren, sowie sinnvolle Konzepte zur Regulation – um dem Zauberlehrling einen Lehrmeister zur Seite zu stellen.

GESUNDHEITSMINISTER SIEHT CHANCEN, ABER AUCH RISIKEN

Lauterbach warnt vor Gefahr von Fehldiagnosen durch KI

Bundesgesundheitsminister **Karl Lauterbach** sieht beim Einsatz von künstlich intelligenten Chatbots wie ChatGPT im **Gesundheitssystem** Chancen, warnt aber auch vor Gefahren wie etwa Fehldiagnosen. „Es wird bald Programme geben, bei denen ein Patient Symptome, Befunde und bisherige Behandlungen mündlich erklärt und dann von der KI

eine Einschätzung seiner Krankheit und sogar mögliche **Therapievorschl**äge bekommt“, sagte der SPD-Politiker.

Allerdings können **Chatbots** auch erfundene Informationen verbreiten, ohne dass dies auf den ersten Blick erkennbar ist. Lauterbach forderte, die Anwendung von KI-Systemen wie ChatGPT im

Gesundheitsbereich zu regulieren. „Sie müssen geprüft werden und zuverlässig sein.“ Zudem müsse sichergestellt sein, dass die Daten nicht missbraucht werden könnten.

Grundsätzlich sehe er den **Einsatz von KI** im Gesundheitssystem aber positiv. Chatbots könnten künftig Therapien theoretisch durch-

spielen und Fragen nach der Wirksamkeit von **Medikamenten** für einen konkreten Patienten beantworten.

KI könne bereits jetzt manchmal besser sein als ein geübter **Facharzt**, meinte Karl Lauterbach in einem Interview. „Die besten Ergebnisse erzielt aber die **Kombination** aus künstlicher Intelligenz und einem Arzt.“ **dpa**