Experten diskutieren darüber, ob jeder Zugang zu den mächtigsten KI-Modellen der Welt haben sollte. ANTHONY KWAN / GETTY

Offene KI an der Weltspitze

Das KI-Modell von Deepseek kann jeder frei herunterladen, verändern und weitergeben. Was Open Source bei künstlicher Intelligenz bedeutet und welche Motivation dahintersteckt. VON ANNA WEBER

Hat China die USA in Sachen künstliche Intelligenz (KI) eingeholt? Seit das chinesische Startup Deepseek letzte Woche ein KI-Modell veröffentlicht hat, das es mit den besten amerikanischen KI-Modellen aufnehmen kann, beschäftigt diese Frage Medien, Politiker und Experten. Doch Deepseek symbolisiert nicht nur den Wettkampf der beiden Weltmächte. In ihm äussert sich noch ein weiterer Konflikt, einer, der noch tiefer liegt: offene gegen geschlossene KI. Deepseek stellt sein neustes KI-Modell frei zur Verfügung. Jeder kann es herunterladen, prüfen, verändern, weiterverwenden – im Gegensatz zu den Top-Modellen von Open AI. Diese sind nur für zahlende Kunden zugänglich, und wie sie aufgebaut wurden, bleibt geheim.

Für Befürworter offener KI ist das Unternehmen Deepseek ein Held. Sie wollen, dass KI allen Menschen frei zur Verfügung steht, damit alle von ihr profitieren und sie gemeinsam weiterentwickeln können. Kritiker sehen in offener KI eine Gefahr. Wenn sie frei verfügbar ist, gibt es keine Möglichkeit mehr, ihren sicheren Einsatz zu gewährleisten. So argumentiert etwa Open AI. Welche Folgen hat es, wenn KI frei verfügbar ist? Und warum ist der Ausdruck Open Source in Bezug auf KI so umstritten?

Eine Utopie wird geboren

Über die Frage, ob Computer-Code geteilt werden oder geheim gehalten werden sollte, wird gestritten, seit es Software gibt. In den frühen achtziger Jahren entstand die Idee von freier Software. Ihr Begründer ist der Programmierer und Aktivist Richard Stallman. 1983 formulierte er seine Vision: Jeder soll Software uneingeschränkt nutzen, den Quellcode einsehen, verändern und veränderte Versionen an andere weitergeben dürfen.

Die Idee der freien Software umweht ein Hauch von Utopie. Sie weckt das Bild einer Welt, in der zumindest im digitalen Raum alle Menschen zusammenarbeiten, die Resultate ihrer Arbeit solidarisch miteinander teilen und sich dem gemeinsamen Fortschritt verschreiben. Unabhängig von Politik, Nationalitäten und persönlichen Zielen.

Heute spricht man statt von freier Software meist von Open-Source-Software, also Software mit offen zugänglichem Quellcode. Aus den Ideen Stallmans hat sich eine weltweite Bewegung entwickelt: Vom Hobby-Programmierer, der in seiner Freizeit ein kleines Computerspiel entwickelt, bis zum Vollzeit-Entwickler bei einer grossen Softwarefirma tragen viele tausend Menschen zu Open Source Softwareprojekten bei.

Eine zentrale Rolle

Und diese Enthusiasten haben vieles geschaffen. Der Internet-Browser Firefox, das E-Mail-Programm Thunderbird und das Handy-Betriebssystem Android sind nur ein paar wenige, prominente Beispiele für Open-Source-Software. Diese Anwendungen seien nur die Spitze des Eisbergs, sagt Matthias Stürmer. Er forscht an der Berner Fachhochschule zu digitaler Nachhaltigkeit und Open-Source-Software und ist Präsident des Vereins CH Open, der sich für offene Software in der Schweiz einsetzt.

«Wenn heute eine Firma Software auf den Markt bringt, dann ist da ganz viel Open Source drin», sagt er. Denn es gebe viele Open-Source-Bausteine, die für verschiedene Softwareprojekte nützlich sein könnten. Laut der neusten Studie von CH Open geben 96 Prozent aller Schweizer Unternehmen an, in mindestens einem Einsatzgebiet Open-Source-Software zu verwenden. Fast die Hälfte aller Unternehmen nutzt Open Source sogar in mehr als 15 Einsatzgebieten.

Doch Open-Source-Software hat auch eine schwierige Seite. Hacker können den Code lesen und Schwachstellen gezielt angreifen. Diese Sicherheitslücken zu schliessen, liegt dann meist in der Verantwortung von Freiwilligen. Befürworter von Open Source halten dagegen, dass Schwachstellen dafür in öffentlich zugänglichem Code besonders schnell gefunden würden. Und selbst den Befürwortern ist klar, dass Open-Source-Software für schädliche Zwecke eingesetzt werden kann. Manche Entwickler setzen deshalb sogenannte «Do no harm»-Lizenzen ein. Diese schreiben fest, dass man die Software nicht benutzen darf, um Menschenrechte zu verletzen, die Umwelt zu zerstören oder Kriege zu gewinnen. Ob eine Lizenz ausreicht, das zu verhindern, bleibt fraglich.

> «Langfristig werden offene und geschlossene KI koexistieren.»
>
> **Lewis Tunstall**
> Datenwissenschafter

Bei KI sind die Risiken sogar noch grösser. Moderne KI-Modelle können Code für Cyberattacken schreiben, Bilder von Kindesmissbrauch generieren oder Menschen in den Selbstmord treiben. Grosse Unternehmen wie Open AI bemühen sich, solchen Missbrauch zu minimieren. Doch offene Modelle sind meist weniger stark kontrolliert. So erstaunt es nicht, dass der Chatbot von Deepseek laut ersten Tests der KI-Sicherheitsfirma Enkrypt AI deutlich mehr toxische und schädliche Inhalte generiert als Chat-GPT.

Für Vertreter offener KI wiegen diese Risiken weniger schwer als jene, die aus einer Machtkonzentration durch verschlossene KI entstehen. Sie haben mit Deepseek und anderer Open-Source-KI genau das umgekehrte Problem: Sie ist ihnen nicht offen genug.

Das Startup Deepseek etwa hat ein trainiertes KI-Modell veröffentlicht und erlaubt, dass jeder es herunterladen, selbst ausführen und verändern darf. Ausserdem haben die Programmierer in einem Forschungspapier genau beschrieben, wie sie die KI entwickelt haben. Doch KI ist komplexer als ein normales Softwareprogramm. Wie ein Chatbot genau reagiert, wenn jemand eine Anfrage eintippt, hängt von vielen weiteren Faktoren ab: den Daten, mit denen das Modell trainiert wurde, dem Programmcode, der das Training gesteuert hat, dem Programmcode, der die Anfrage an das Modell weitergibt. Nichts davon hat Deepseek veröffentlicht.

Diese Wissenslücke möchte Lewis Tunstall schliessen. Er arbeitet bei der Plattform Hugging Face, auf der Entwickler ihre offenen KI-Modelle zur Verfügung stellen, und leitet eine Initiative von Entwicklern, die den Chatbot von Deepseek nachbauen möchten. Alle Daten und den gesamten Programmcode, den sie für ihren Nachbau verwenden, wollen sie veröffentlichen. Denn laut Tunstall gibt es bis anhin nur sehr wenige KI-Modelle, die tatsächlich auf allen relevanten Ebenen transparent und zugänglich sind. Das gilt nicht nur für Deepseek, sondern auch für Meta, das französische KI-Unternehmen Mistral AI und Elon Musks Firma XAI. Sie stellen ihre trainierten KI-Modelle zur Verfügung, halten alle weiteren Informationen aber geheim.

Gut fürs Image

Besonders kritisch ist das bei den Trainingsdaten. Denn sie bestimmen, welche Vorurteile in einem Modell stecken. Verzerrte Daten können zum Beispiel dafür sorgen, dass ein Chatbot annimmt, bei Ärzten handle es sich immer um Männer. Auch antisemitische, rassistische und homophobe Gedanken finden über die Daten ihren Weg in das KI-Modell. Dennoch werden die meisten Firmen ihre Daten wohl nicht herausgeben. «Die Daten sind das Geheimrezept, das Unternehmen einen Wettbewerbsvorteil verschafft», sagt Tunstall. Zudem stammen die Daten meist aus dem Internet, wo sie ohne Einwilligung der Urheber gesammelt wurden – ein Umstand, der bereits zu Klagen führte.

Was motiviert ein Unternehmen überhaupt dazu, sein KI-Modell zu veröffentlichen? Schliesslich stecken in der Entwicklung und dem Training eines grossen Modells enorm viel Zeit, Arbeit und Geld. Lewis Tunstall nennt zwei Gründe. «Es kann darum gehen, ein Ökosystem aufzubauen», sagt er. Stellt ein Unternehmen sein KI-Modell zur Verfügung, bauen andere darauf auf, entwickeln es weiter, und das Unternehmen profitiert wiederum von ihren Fortschritten. Das sei die Strategie von Meta, sagt Tunstall. Ausserdem sei es gut für das Image. Das sehe man gerade bei Deepseek. «Wir kannten die Firma schon seit Jahren, aber mit dieser Veröffentlichung haben sie alle beeindruckt und an Glaubwürdigkeit und Anerkennung gewonnen.»

Unabhängig von Ängsten und Idealismus steht der Welt mit Deepseeks Chatbot nun ein sehr mächtiges KI-Modell zur Verfügung. In den zwei Wochen seit der Veröffentlichung wurde es millionenfach heruntergeladen. Und die Weiterentwicklung läuft schon auf Hochtouren. Über tausend abgewandelte Versionen haben Entwickler bereits hochgeladen. Laut Tunstall beschleunigt das Modell von Deepseek die Weiterentwicklung von KI gleich doppelt. Einmal direkt, weil Forschern, Unternehmen und Entwicklern Zugriff auf ein Modell mit Spitzenleistung ermöglicht wird. Und einmal indirekt, weil es den Konkurrenzdruck auf die grossen Tech-Firmen erhöht.

Dass offene KI-Modelle geschlossene verdrängen werden, glaubt Tunstall jedoch nicht. «Langfristig werden offene und geschlossene KI koexistieren», sagt er. Denn beide hätten für verschiedene Anwendungsbereiche Vor- und Nachteile. Genau wie es auch bei klassischer Software sowohl für freie als auch für proprietäre Programme einen Platz gibt.