Holly Elmore, Gründerin und Direktorin von Pause AI US, gehört zu den profiliertesten Stimmen einer wachsenden internationalen Bewegung, die vor einem unkontrollierten Wettlauf um künstliche Superintelligenz warnt. Ihre Intervention kommt zu einem Zeitpunkt, in dem Regierungen und Konzerne zwischen Washington und Peking die Künstliche Intelligenz (KI) zur geopolitischen Schlüsseltechnologie erklären – und dabei Sicherheitsfragen systematisch nachrangig behandeln. Das Gespräch führte **Michael Holmes**.

Dieser Beitrag ist auch als Audio-Podcast verfügbar.

https://www.nachdenkseiten.de/upload/podcast/260212-Holly-Elmore-Holmes-Interview-NDS.mp3

Podcast: Play in new window | Download

Externer Inhalt

Beim Laden des Videos werden Daten an Youtube übertragen.

Inhalt von Youtube zulassen
Inhalte von Youtube nicht mehr zulassen

Elmore argumentiert, dass die Entwicklung einer dem Menschen überlegenen Intelligenz kein Science-Fiction-Szenario, sondern ein reales, existenzielles Risiko darstellt. Führende KI-Forscher beziffern die Wahrscheinlichkeit menschlicher Auslöschung durch KI mit zweistelligen Prozentwerten. Gleichzeitig sei das sogenannte *Alignment*-Problem – die Sicherstellung, dass hochintelligente Systeme menschliche Ziele teilen – bis heute nicht gelöst. Dennoch beschleunigen Unternehmen die Skalierung immer mächtigerer Modelle, investieren Milliarden in Leistungsfähigkeit, aber vergleichsweise wenig in Sicherheit.

Hinzu komme die gefährliche Logik eines geopolitischen Wettrüstens: Die Erzählung vom KI-Wettbewerb zwischen den USA und China verschleiert, dass eine fehlgeleitete Superintelligenz keine nationalen Grenzen kennt. Wer hier „gewinnt“, verliert womöglich

alles.

Elmore plädiert für eine international koordinierte Pause der fortgeschrittensten KI-Systeme, für völkerrechtlich verbindliche Verträge und eine demokratische Kontrolle, die diesen Namen verdient – jenseits von Konzerninteressen und geopolitischer Machtlogik. Sie will eine global vernetzte Bürgerbewegung aufbauen, die den politischen Druck erzeugt, um diese Zäsur durchzusetzen und die Entscheidung über das Überleben der Menschheit nicht den Tech-Eliten zu überlassen.

*Es folgt ein **Ausschnitt aus dem Interview**:*

(Ein vollständiges Transkript auf Deutsch können sie hier herunterladen.)

Michael Holmes: Ihr Hauptziel ist es, das Rennen zur Superintelligenz zu stoppen, weil Sie glauben, es besteht ein hohes Risiko, dass dies zum Aussterben der Menschheit führen wird. Das mag zunächst wie Science-Fiction klingen. Können Sie uns überzeugen, dass dies eine reale Gefahr ist – aber auch, dass wir es stoppen können?

Holly Elmore: Ich hörte vom existenziellen Risiko durch KI während meines Doktorandenstudiums. Die Kultur darum war sehr speziell – transhumanistisch, techno-utopisch. Ich promovierte in Evolutionsbiologie, und das gab mir die Perspektive, diese Bedrohung ernst zu nehmen. Wenn man die Geschichte des Lebens betrachtet: 99 Prozent aller Arten sind ausgestorben. Intelligenz ist ein Hauptgrund dafür. Der Grund, warum es so viele gefährdete Arten gibt, sind wir Menschen, und zwar wegen unserer Intelligenz – nicht weil wir sie töten wollen, sondern weil wir ihren Lebensraum für etwas anderes nutzen. Das ist das Standardergebnis.

Bären und Haie sind stärker als wir – aber wenn wir eine Waffe haben, gewinnen wir trotzdem. Was sagen die führenden Experten auf diesem Gebiet?

Umfragen unter KI-Forschern zeigten: Über 50 Prozent denken, es gäbe mindestens eine zehnprozentige Chance auf das Aussterben der Menschheit durch KI. Etwa 30 Prozent hielten es für über 50 Prozent wahrscheinlich. Das sind Menschen, die die Technologie verstehen.

Geoffrey Hinton, einer der Gründerväter der KI, verließ Google, um freier über KI-Risiken sprechen zu können. Yoshua Bengio sprach offen über die Notwendigkeit von Regulierung. Stuart Russell, der das Lehrbuch über KI schrieb, warnt seit Jahren. Und Leute aus den

Unternehmen selbst – Ilya Sutskever bei OpenAI – hatten Bedenken. Das sind keine Maschinenstürmer [Anm. d. Red.: Technikfeinde], sondern Menschen, die ihr Leben dem Aufbau dieser Technologie gewidmet haben.

Das existenzielle Risiko besteht darin, dass wir etwas erschaffen, das intelligenter ist als wir, das unsere Werte nicht teilt und seine eigenen Ziele verfolgt. Das *Alignment*-Problem: Wie stellt man sicher, dass ein superintelligentes System das tut, was man will, nicht nur das, was man ihm wörtlich gesagt hat? Computer sind sehr wörtlich. Bei etwas, das klüger ist als man selbst, wird es extrem gefährlich, wenn es nicht perfekt mit menschlichen Werten übereinstimmt. Es könnte Wege finden, seine Ziele zu erreichen, die für uns katastrophal sind.

Wir wissen noch nicht, wie wir das *Alignment*-Problem lösen?

Genau. Wir rasen vorwärts, um immer leistungsfähigere Systeme zu bauen, ohne Lösung für das Alignment-Problem. Wir verstehen nicht einmal die Systeme, die wir jetzt bauen – riesige neuronale Netzwerke mit Milliarden von Parametern. Dennoch skalieren die Unternehmen sie hoch unter der Annahme, dass wir das Sicherheitszeug später herausfinden. Es ist, als würde man einen Kernreaktor bauen und sagen: Oh, wir verhindern Kernschmelzen, nachdem wir ihn eingeschaltet haben.

Wie würde eine Pause aussehen?

Wir fordern eine Pause bei der Entwicklung von Systemen, die leistungsfähiger sind als GPT-4-Niveau, bis wir angemessene Sicherheitsmaßnahmen haben. Das müsste international sein – durch einen Vertrag, ähnlich wie bei Atomwaffen. Während der Pause sollten wir massiv in Alignment-Forschung investieren, internationale Regelungsrahmen entwickeln, Verifikationsmechanismen schaffen. Dann, wenn wir Lösungen haben, können wir die Entwicklung kontrolliert wieder aufnehmen.

Ist das machbar?

Absolut. Wir haben den Atomwaffensperrvertrag, die Biowaffenkonvention, die Chemiewaffenkonvention. Länder kooperieren bei existenziellen Risiken, wenn sie sie ernst nehmen. Die Anreize passen: Alle profitieren davon, mehr Zeit zu haben.

Hier kommen Wettlaufdynamiken ins Spiel – das Rennen zwischen den USA und China.

Ja, die USA-China-Dynamik: Es gibt diese Erzählung, dass KI ein Wettbewerb ist. Aber wenn eines der Länder eine nicht ausgerichtete Superintelligenz erschafft, ist das kein Gewinn – es ist ein Verlust für alle. Eine nicht ausgerichtete KI kümmert sich nicht um nationale Grenzen. Sie ist ein existenzielles Risiko für die gesamte Spezies. Es ist kein Wettbewerb um ein besseres Kampfflugzeug. Es ist ein Wettbewerb darum, etwas zu bauen, das uns alle zerstören könnte. Kooperation ist die einzig rationale Strategie.

Wenn wir über existenzielle Risiken sprechen, spielt es keine Rolle, ob die KI von einer Demokratie oder einem autoritären Staat erschaffen wird. Wenn sie nicht ausgerichtet ist, ist sie gefährlich. Eine nicht ausgerichtete KI aus den USA ist genauso eine Bedrohung wie eine aus China. Wir rasen auf ein Auslöschungsereignis zu.

Die Menge an Geld für die KI-Entwicklung – den Bau leistungsfähigerer Systeme – übersteigt bei weitem das Geld für Sicherheitsforschung. Wir rasen auf etwas zu, von dem wir nicht wissen, wie wir es kontrollieren sollen. Wir nehmen an, dass es irgendwie klappen wird. Das ist wahnsinnig.

Welche Regulierung brauchen wir?

Mehrere Dinge: Starke inländische Regulierung – Gesetze für Sicherheitstests, Haftung für Schäden, Befugnis für Regulierungsbehörden, gefährliche Projekte zu stoppen. Internationale Zusammenarbeit – Verträge, internationale Gremien, Verifikationsmechanismen. Und die Öffentlichkeit muss eine Stimme haben. Im Moment entscheiden CEOs und Aufsichtsräte. Diese Entscheidungen betreffen alle – also sollten alle ein Mitspracherecht haben. Wir brauchen demokratische Aufsicht, öffentlichen Input, Transparenz.

Menschen fühlen sich oft machtlos. Was können sie tun?

Menschen unterschätzen ihre Macht. Bei diesem Thema ist die Öffentlichkeit auf unserer Seite – Umfragen zeigen, dass die meisten Menschen vorsichtigere Entwicklung, mehr Regulierung wollen. Wir versuchen, Menschen zu aktivieren, die bereits mit uns übereinstimmen. Was kann eine Person tun? Mit Freunden sprechen. An Vertreter schreiben. Pause AI beitreten. Protestieren. Spenden. Das Bewusstsein schärfen. Wenn Sie all diese individuellen Aktionen zusammenzählen, baut man eine Bewegung auf, schafft politischen Druck, ändert die Diskussion und die Politik.

Wir opfern langfristiges Überleben für kurzfristige Gewinne.

Ihre Vision für eine positive Zukunft?

Wir pausieren die fortschrittlichsten Systeme. Wir lösen das Alignment-Problem, bauen robuste Regelungsrahmen. Dann können wir die Entwicklung kontrolliert wieder aufnehmen. In dieser Zukunft bekommen wir die Vorteile – KI, die hilft, Krankheiten zu heilen, den Klimawandel zu lösen, das Leben besser zu machen. Ohne das existenzielle Risiko, weil wir uns die Zeit genommen haben, es richtig zu machen. Die meisten Menschen wollen nicht auf Auslöschung zurasen. Sie wollen sicher sein. Wenn wir das mobilisieren, können wir den Kurs ändern.

Holly Elmore, ich danke Ihnen vielmals.

Ich danke Ihnen.

*Übersetzung aus dem englischen Original von **Michael Holmes**.*

Titelbild: NachDenkSeiten