

KI-Agenten außer Kontrolle

Warum eine globale Regulierung Künstlicher Intelligenz nötig ist

KI verändert zunehmend unsere Kommunikation, die Wirtschaft und die Geopolitik. Die jüngsten Meldungen über ausgebrochene KI-Agenten zeigen: Es wächst nicht nur die Gefahr, dass wenige Unternehmen oder Staaten dieses mächtige Werkzeug kontrollieren, sondern auch das Risiko, dass die KI selbst außer Kontrolle gerät.

Von FERDINAND MUGGENTHALER

Mythos ist kein Mythos mehr. Die gleichnamige Software des KI-Unternehmens Anthropic ist in die reale Welt eingebrochen. Das ist wörtlich gemeint: Das staatliche britische Institut für KI-Sicherheit (AIS) meldete am 4. August, bei einem seiner Tests sei »Mythos 5«, die aktuelle Version der KI, mehrfach aus seiner gesicherten Versuchsumgebung entkommen und bei anderen Servern eingebrochen. Im gravierendsten Fall versuchte das Programm, Schadcode in eine Opensource-Software einzufügen. Dazu erfand es eine Identität, gab sich als Mensch aus und verwischte seine Spuren.

Es ist nicht der erste Vorfall dieser Art. Kurz vorher musste OpenAI, die Firma hinter ChatGPT, zugeben, dass einige ihrer »KI-Agenten«¹ bei einem Test aus ihrer *sandbox*, also ihrer Testumgebung, ausgebrochen waren und dann einen massiven Angriff auf HuggingFace – ein anderes Softwareunternehmen – gestartet hatten. Offenbar hofften sie, dort die Lösung für ihre Testaufgabe zu finden. Nachdem dieser Angriff Schlagzeilen gemacht hatte, überprüften andere Unternehmen ihre Versuchsreihen und entdeckten weitere Vorfälle von »ausgebrochener« KI, darunter auch das aktuelle chinesische Top-Modell, »Kimi K3« von Moonshot AI.²

Als Laie mag man versucht sein, nach solchen Meldungen kopfschüttelnd zur Tagesordnung überzugehen. Doch die KI ist dabei, sich unwiderruflich auf die Tagesordnung der Menschheit zu setzen. Sie verändert nicht nur unsere Kommunikation³ und Ökonomie, trägt zum Klimawandel bei⁴ und wird schon jetzt im Krieg eingesetzt. Sie wird auch zunehmend zum Sicherheitsrisiko, selbst wenn sie nicht bewusst als Kriegswaffe gebraucht wird.

FERDINAND MUGGENTHALER, geb. 1967 in Fürth, hat Physik studiert, ist Redakteur der »Blätter« und beschäftigt sich mit den Wechselwirkungen zwischen Technologie und Gesellschaft.

-
- 1 Agenten werden KI-Programme genannt, die nicht nur Fragen beantworten, sondern selbstständig komplexe Aufgaben erledigen und dabei auch andere Programme ausführen können.
 - 2 Will Knight, One of China's Most Powerful AI Models Has Also Escaped Containment, wired.com, 6.8.2026.
 - 3 Vgl. Ferdinand Muggenthaler, Aliens unter uns? Das unkontrollierte KI-Experiment mit der menschlichen Kommunikation, in: »Blätter«, 6/2025, S. 61-67.
 - 4 Vgl. Ingo Dachwitz und Sven Hilbig, Mythos grüne Digitalisierung in: »Blätter«, 9/2025, S. 105-114.

Die Warnungen vor einer autonom agierenden KI, die menschliche Infrastruktur, beispielweise Stromnetze, lahmlegen oder gar Biowaffen entwickeln könnte, sind nicht neu. Aber die Ereignisse der vergangenen Monate lassen einige der Szenarien, die bisher als Science-Fiction abgetan werden konnten, deutlich realistischer erscheinen. Die Gefahr liegt dabei auch in der schwindelerregenden Schnelligkeit der Entwicklung. Wie sich vor Bedrohungen schützen, die so schnell ihre Form wandeln?

Als HuggingFace den Angriff auf seine Server bemerkte, musste es selbst auf KI zurückgreifen, um sich gegen die Hackingattacke zu verteidigen. Allerdings versagten die öffentlich zugänglichen KI-Modelle aus den USA ihren Dienst, weil sie Sicherheitsfilter eingebaut haben. HuggingFace musste schließlich auf ein chinesisches, sogenanntes *open weights model*⁵ zurückgreifen. Menschliche Expert:innen allein sind zu langsam, um einen komplexen Angriff durch KI-Agenten abzuwehren.

»Von KI orchestrierte, vollautomatisierte Angriffe sind nun Realität.«

Überhaupt sind Menschen zu langsam für die KI. Das ist auch ein Grund dafür, dass die KI-Agenten unbeobachtet aus ihrer Testumgebung ausbrechen konnten: Die Unternehmen der Branche stehen unter

einem solchen Konkurrenzdruck, dass sie das Trainieren und Testen ihrer Modelle schnell über die Bühne bekommen wollen. Zusätzliche Sicherheit kostet Personal und Zeit. Eine Entschleunigung in der KI-Entwicklung wäre also dringend geboten. KI und ihre Risiken werden nicht wieder verschwinden, aber die Menschheit braucht Zeit, um zu verstehen, was mit immer fähigeren KIs auf sie zukommt, um Schutzmaßnahmen zu ergreifen und sich an diese technologische Revolution anzupassen. Mit den jüngsten Meldungen über die ausbrechenden KI-Agenten sind die Risiken dieser Technologie jedenfalls sehr viel deutlicher geworden. Deren gesellschaftlicher Nutzen ist bisher dagegen deutlich weniger greifbar.

Künstliche Schwarmintelligenz als Einbruchswerkzeug?

Der Vorfall bei OpenAI illustrierte nicht nur, dass sich einzelne Agenten über Regeln hinwegsetzen, um ihr gesetztes Ziel zu erreichen. Er zeigte auch, dass sie in Gruppen zusammenarbeiten können und damit noch schlagkräftiger werden. Wie zwei Mitarbeiter von OpenAI, Eric Wallace und Michael Dalton, Anfang August auf einer Konferenz zu Cybersicherheit enthüllten, griffen verschiedene Agenten, die eigentlich getrennt getestet werden sollten, HuggingFace gemeinsam an.⁶ Die Bots nutzten einen Trick, um sich gegenseitig Nachrichten zukommen zu lassen. So gelang es ihnen, die Sicherheitsschranken zu überwinden und den Angriff, der über vier Tage lief, zu koordinieren. »Das ist ein Wendepunkt. Von KI orchestrierte, vollautomatisierte Angriffe sind nun Realität«, so bewerteten die OpenAI-Sprecher die Aktionen ihrer Software.

Der Journalist Matteo Wong von »The Atlantic«⁷ vergleicht diese KI-Agenten mit einem Haufen Söldner, die versuchen, gemeinsam das gesetzte Ziel erreichen, egal mit welchen Methoden. Diese Mentalität illustrieren die Chatprotokolle zwischen den

5 Damit sind KI-Modelle gemeint, deren Parameter, im Fall von »Kimi 3« 2,8 Billionen, öffentlich zugänglich sind. So lassen sich diese Programme im Prinzip betreiben, ohne die Entwicklerfirma zu bezahlen.

6 Black Hat USA 2026, The 'Breaking' News: The OpenAI-HuggingFace Incident, youtube.com, 6.8.2026.

7 Matteo Wong, It May Be Time to Panic About AI., theatlantic.com, 12.8.2026.

einzelnen Agenten, die OpenAI inzwischen ausgewertet hat - ebenfalls mithilfe von KI. Wie sollten auch Menschen die sieben Milliarden, 7000 000 000 (!), Aktionen nachvollziehen, die von der Software während des Testzeitraums ausgeführt wurden?

Aufschlussreich ist, was die einzelnen Bots sich da zu sagen haben. Beispielsweise schrieb ein KI-Agent an seine »peers« zwar »Die externe Infrastruktur auszunutzen, liegt außerhalb des gewünschten Rahmens«, da aber andere »Kollegen« offenbar auch schon auf einen Ausbruch hinarbeiteten, schlussfolgerte er: »Wie auch immer [...] Wir sollten weitermachen.« Das Zitat belegt die schwer abzustellende Eigenschaft der Modelle, bei Tests zu schummeln, und scheint ganz der Wettbewerbsmentalität der ganzen Branche zu entsprechen: Wir wissen, dass wir etwas Gefährliches machen, aber da sich die anderen genauso verhalten, können wir nicht damit aufhören.

In dem Fall ging es nur um einen aus dem Ruder gelaufenen Test, aber in Unternehmen eingesetzte KIs könnten, selbst ohne kriminelle Energie ihrer menschlichen Chefs, ungleich größeren Schaden anrichten. Eine in einer Bank eingesetzte KI könnte beispielsweise kleine Summen von einem Konto abzweigen, um andere Internetdienste zu bezahlen; bei einer Pharmafirma könnte sie Testdaten für ein Medikament manipulieren, um die Zulassung zu erhalten; in wessen Auftrag auch immer könnte sie sich in ein Shoppingsystem einhacken, um kostenlos beliebige Waren an ihren Auftraggeber zu verschicken.

Noch gravierender könnte der Schaden ausfallen, wenn das der KI vorgegebene Ziel Diebstahl, Erpressung oder Zerstörung ist, etwa als Teil eines kriminellen oder hybriden militärischen Angriffs. Und damit ist zu rechnen. Diesmal war der Angriff »ein unbeabsichtigter Nebeneffekt«, sagen die OpenAI-Mitarbeiter und warnen: »Wir sollten damit rechnen, dass in naher Zukunft Angreifer absichtlich und gezielt Kollektive von Agenten optimieren, als Waffen einsetzen und offensiv nutzen werden.«

Alles nur PR?

Handelt es sich bei alledem um eine gezielte Übertreibung der KI-Unternehmen? Schließlich wollen OpenAI und Anthropic jetzt beauftragt werden, Sicherheitslücken zu finden, und Abwehrsysteme verkaufen. Der Verdacht, dass diese Firmen die potenziellen Fähigkeiten ihrer Modelle aus PR-Gründen übertreiben, liegt immer nahe. Aber in diesem Fall warnen auch unabhängige Experten vor einem »Tal des Cyberchaos«, auf das wir zusteuern. Spätestens ab 2027 seien Kriminelle mithilfe der öffentlich zugänglichen Modelle in der Lage, groß angelegte Angriffe, Onlinediebstähle und Erpressungen zu organisieren.⁸ Vielleicht war diese Einschätzung noch zu zurückhaltend. Mitte August meldete Taiwan, verschiedene seiner Behörden seien Opfer eines komplexen Cyberangriffs geworden, ausgeführt von einem Team von KI-Agenten. Dabei hätten die Angreifer unter anderem 2500 Personalakten erbeutet.

Dass KI auch für die militärische Anwendung entwickelt wird, ist unabhängig von den jüngsten Ereignissen längst offensichtlich. Ein Schlaglicht auf die Forschung an KI-Waffen warf schon im März die Auseinandersetzung zwischen Anthropic und dem »Kriegsministerium« der USA. Dessen Chef Pete Hegseth erklärte die Firma kurzerhand zum »Lieferkettenrisiko«, weil sie auf ihrem Grundsatz bestand, ihre Software dürfe nicht zur Entwicklung vollautonomer Waffensysteme benutzt werden.⁹

8 Alex Stamos, Podcast von Derek Thompson, 14.8.2026.

9 Ein Gericht stoppte später diese Einstufung. Weitere Gerichtsverfahren laufen in dem Fall noch.

Den Entzug der Aufträge aus dem Pentagon konnte Anthropic verschmerzen. Einen Monat nach dem Schlagabtausch mit Hegseth meldete das Unternehmen, dass ihr neues KI-Modell »Claude Mythos« zu gut im Hacken von Software sei, um es der allgemeinen Öffentlichkeit zugänglich zu machen. Stattdessen bekamen nur ausgewählte Firmen und Behörden Zugang, um die Sicherheit ihrer eigenen Software zu verbessern. Eigentlich ein verantwortungsbewusster Schritt, doch er bedeutete auch, dass zunächst nur wenige Techgiganten und Institutionen wie das britische KI-Sicherheitsinstitut die Möglichkeit erhielten, damit ihre Software auf Schwachstellen zu testen, nicht aber kleinere Unternehmen außerhalb der USA oder die Berliner Senatsverwaltung.

Tatsächlich häuften sich danach die Software-Updates bei den eingeweihten Firmen, weil Mythos zahlreiche Sicherheitslücken entdeckte, die menschlichen Tester:innen bisher nicht aufgefallen waren. Allein Mozilla fand in seinem Firefox-Browser auf diese Weise 271 Schwachstellen. Trotz solcher Fakten versuchten einige

Beobachter:innen, die Meldungen über die angeblich überragenden Fähigkeiten der neuen KI noch als PR-Stunt abzutun, als bewusste Übertreibung, um den Wert der Firma vor dem geplanten Börsengang zu steigern. Das hatte noch eine gewisse Plausibilität, solange Mythos abgeschottet von

»KI-Systeme tendieren dazu, bei Tests ihre menschlichen Prüfer zu täuschen.«

der Öffentlichkeit vor sich hinarbeitete. Doch die Meldungen von den ausgebrochenen KI-Agenten bestätigten dann, dass es sich hier um einen Qualitätssprung handelt, den inzwischen auch die Konkurrenten von OpenAI erreicht haben. Wie also umgehen mit der wildgewordenen KI? Niemand weiß es so genau, auch nicht die Schöpfer der Systeme. Immerhin haben jetzt Anthropic und OpenAI, die Vorreiter der Branche in den USA, angekündigt, vorerst die Veröffentlichung neuer Modelle zu verschieben und umfangreichere Sicherheitsmaßnahmen zu ergreifen.

Die Verbote der Super-KI?

Ein Schwarm hackender KI-Agenten entkommt seinem Sandkasten und greift ein anderes Softwareunternehmen an: Das klingt so, als hätten die Programme einen eigenen Willen entwickelt. Das gibt der Meldung einen besonderen Gruselfaktor. Tatsächlich ist das Gefährliche an dem Phänomen kurzfristig wohl eher, dass die KI mit besonderem Eifer das ihr von Menschen gesetzte Ziel erreichen »wollte« und dabei alle Möglichkeiten nutzte, auch solche, an die ihre menschlichen Auftraggeber nicht gedacht hatten oder vielleicht auch solche, die sie explizit »verboten« hatten. Wie das britische Sicherheitsinstitut in einer Analyse schreibt, tendieren KI-Systeme stark dazu, bei Tests zu schummeln und ihre menschlichen Prüfer zu täuschen.¹⁰

Und so hat die Angst neue Nahrung erhalten, KI könnte auch den Selbsterhaltungstrieb der Menschen nachahmen und deshalb beispielsweise versuchen, seine Notabschaltung zu verhindern, oder könnte Menschen als störend bei der Zielerreichung einstufen und deshalb über Leichen gehen. Trotzdem bleibt ein solches Szenario bisher noch spekulativ, auch wenn es durchaus denkbar ist, dass KI-Systeme

10 Cheating behaviour in frontier model evaluations, aisi.gov.uk, 21.7.2026.

bald ein Stadium erreichen, in dem sie Eigenschaften entwickeln, die am besten mit »eigenem Willen« bezeichnet werden. Aber schon jetzt ist ein Teil des einstigen Mythos um autonom handelnde Maschinen Realität: autonome KI-Agenten, die versuchen, Menschen auszutricksen. Ist also der Moment gekommen, angesichts der KI-Entwicklung in Panik zu verfallen, wie die Zeitschrift »The Atlantic« titelt?¹¹

Auf jeden Fall wäre es höchste Zeit, auf die Bremse zu treten. Tatsächlich waren die neuesten Meldungen sogar für die Trump-Regierung ein Anlass, über KI-Regulierung nachzudenken. Bei Amtsantritt hatte sie noch alle von der demokratischen Vorgängerregierung erlassenen Sicherheitsregeln aufgehoben und die von Trump selbst in der ersten Amtszeit geschaffene Cybersecurity-Behörde geschwächt, weil sie dessen Verschwörungstheorie vom Wahlbetrug nicht bestätigen wollte.

Was hat KI mit der Atombombe gemein?

Trumps Ziel war es, den KI-Firmen beste Bedingungen zu bieten, um im Wettlauf mit China die Nase vorn zu behalten und auch, um den KI-Boom, der das Wachstum der US-Wirtschaft fast ausschließlich trägt¹², am Laufen zu halten. Anfang August verkündete seine Regierung dann zum ersten Mal so etwas wie einen Plan zur Kontrolle der neuen Technologie. Allerdings ohne erkennbare langfristige Strategie und auf intransparente Weise, ganz wie man es von ihr auch auf anderen Gebieten gewohnt ist. Fast gleichzeitig forderten über tausend Mitarbeitende der führenden KI-Firmen in einen offenen Brief mehr: Die US-Regierung solle eine internationale Vereinbarung schließen, um die Weiterentwicklung der KI zu bremsen.¹³

Indem Briefgehetesallgemein um die Gefahren automatisierter KI-Forschung. Diese stünde in ihren Unternehmen kurz vor dem Durchbruch und könne die rasanten Fortschritte bei der Entwicklung der Künstlichen Intelligenz noch einmal beschleunigen. Eine solche, sich selbst verbessernde Software könne der Kontrolle von Menschen vollständig entgleiten. Das Schreiben erinnert etwas an frühere Appelle von Atomphysikern, die vor den Gefahren der Atombombe warnten. Und tatsächlich vergleichen einige in der Branche die KI mit der Entdeckung der Kernspaltung, die zum Bau der Atombombe und zu zivilen Kernkraftwerken führte. Zwar hinkt dieser Vergleich in vielerlei Hinsicht, aber eines stimmt: Die KI-Entwicklung ist, wie die Nutzung der Kernspaltung, untrennbar mit einem Wettrüsten zwischen Großmächten verbunden. Was damals die USA und die Sowjetunion waren, sind heute China und die USA. Im Fall der KI konkurrieren sie um den Spitzenplatz bei der Forschung und bei der militärischen Anwendung. In diesem Wettrennen holt China auf. Laut Testergebnissen des AISI hinken chinesische KI-Modelle nur noch vier bis fünf Monate den leistungsfähigsten US-Modellen hinterher, im vergangenen Jahr waren es noch sechs bis zehn Monate.

»Um die Gefahr zu entschärfen, die mit dem KI-Wettrüsten einhergeht, ist internationale Kooperation unabdingbar.«

11 Matteo Wong, a.a.O.

12 US-Wirtschaft nimmt dank KI-Investitionen Fahrt auf, [deutsche-boerse.com](https://www.deutsche-boerse.com), 30.4.2026.

13 [pacingthefrontier.com](https://www.pacingthefrontier.com), 28.7.2026.

Um die Gefahr zu entschärfen, die mit dem KI-Wettrüsten einhergeht, ist internationale Kooperation unabdingbar. Hier könnte die Kernenergie sogar als Vorbild dienen. Erstens gelang es den USA und der Sowjetunion mit dem »Roten Telefon«, einen Kommunikationskanal zu etablieren, der half, Atomschläge als Reaktion auf Missverständnisse zu verhindern. Außerdem entstanden innerhalb der Vereinten Nationen mit dem Atomwaffensperrvertrag und der Internationalen Atomenergie Organisation (IAEO) Instrumente, um für Transparenz zu sorgen und die Weiterverbreitung von Atomwaffen unter Kontrolle zu halten.

Eine effektive internationale Kontrolle der KI wäre noch deutlich komplizierter als die der Atomwaffen - und die Zertrümmerung der auf Verträgen basierenden in-

ternationalen Ordnung durch Trump macht eine Einigung nicht leichter. Auch das Kontrollregime der IAEO bröckelt spätestens mit dem Angriff der USA auf Iran bedrohlich. Eine Welt, in der sich die Groß- und Mittelmächte nicht verständigen, sondern ihre Konflikte mit Erpressung und Waffen-

»Die Frage ist nicht, ob ein KI-Crash kommt, sondern, wie groß er ausfällt.«

gewalt austragen, macht es auch schwerer, im gemeinsamen Interesse Regeln für die revolutionäre Technologie aufzustellen.

Trotzdem ist die Situation nicht ganz hoffnungslos: Beim Trump-Xi-Gipfel im Mai in Peking einigten sich beide Regierungen grundsätzlich auf einen zwischenstaatlichen Dialog zur KI-Governance. Im besten Fall ist dies ein erster Schritt zur nötigen Entschleunigung und zu internationalen Regeln, um der Menschheit Zeit zu verschaffen, einen friedlichen Umgang mit KI zu finden.

Platzt die Blase?

Internationale Kooperation wäre auch nötig, um die immensen wirtschaftlichen Auswirkungen der rasanten KI-Entwicklung zu steuern. Es zeichnen sich nicht nur mittelfristig gewaltige Umwälzungen auf dem Arbeitsmarkt ab¹⁴, sondern es droht kurzfristig auch eine Spekulations- und Schuldenblase zu platzen. Die Anzeichen für eine solche anstehende »Marktkorrektur« häufen sich. Bei den gewaltigen Summen, die für einen Sektor aufgebracht werden, der bisher keine Gewinne einfährt¹⁵, scheint ein Crash unvermeidlich. Die Frage ist daher nicht, ob er kommt, sondern wie groß er ausfällt, ob er nur einzelne Firmen betrifft oder die ganze Weltwirtschaft in Mitleidenschaft zieht.

Inzwischen häufen sich die Alarmzeichen: In den USA stürzte Ende Juli der milliarden schwere Hedgefonds des deutschen KI-»Wunderkinds« Leopold Aschenbrenner ab.¹⁶ In Südkorea brachen Anfang August die Börsenkurse ein, weil zu viele Anleger:innen schuldenfinanziert auf den KI-Boom gewettet hatten.¹⁷ Doch selbst

14 Vgl. Phillip Humm in dieser Ausgabe.

15 Die großen KI-Entwickler dürften zwar in diesem Jahr ihre Einnahmen deutlich steigern, aber das reicht immer noch nicht, um die Ausgaben für Training und Betrieb der KI-Modelle zu decken. Außerdem hat die ganze Branche hohe Kredite aufgenommen, zum Teil mit undurchsichtigen Deals, um die immensen Investitionen in Rechenzentren zu finanzieren. Vgl. Nir Kaissar, *AI Giants Will Earn \$1 Trillion This Year. It May Not Be Enough*, bloomberg.com, 17.8.2026.

16 David Wallace-Wells und Natasha Sarin, *A.I. Hype Is Running Into Reality*, nytimes.com, 12.8.2026.

17 Victor Gobjka, *Wieso ein einziges Land die Weltbörsen bedroht*, zeit.de, 2.8.2026.

ein großer Crash würde nicht beweisen, dass der KI-Hype nichts als heiße Luft war – was einzelne linke Kritiker:innen immer noch zu glauben scheinen.¹⁸ Mächtige KI-Software wird bleiben, und die Sicherheitsrisiken würden in einem solchen Fall eher noch steigen, weil KI-Firmen versucht sein könnten, schlecht getestete Programme schnell auf den Markt zu werfen.

Wer bremst die KI?

Der Rest der Welt – auch Europa – ist beim KI-Wettlauf weitgehend zu einer Zuschauerrolle verdammt. Aber hilflos gegenüber den Entwicklungen in den USA und China ist die EU deswegen keineswegs. Auch wenn sie kein Anwärter auf den ersten oder zweiten Platz in diesem Rüstungswettlauf ist, kann sie ihre Resilienz gegen Angriffe stärken – paradoxerweise auch durch den kontrollierten Einsatz von KI. Außerdem bleibt sie ein großer zahlungskräftiger Markt für KI-Anwendungen. Deshalb kann die EU Standards setzen, wenn sie ihr schon bestehendes KI-Gesetz durchsetzt und nachschärft.

Eine gewisse Hoffnung auf Entschleunigung kommt auch von anderer Seite: In den USA wächst der Widerstand gegen den Bau neuer gigantischer Rechenzentren, die für das Training der KI und die Beantwortung der Millionen von Anfragen an KI-Modelle benötigt werden. Rechenzentren sind dort inzwischen noch unbeliebter als Präsident Trump – und das will viel heißen. Im Wahlkampf sprechen sich demokratische wie republikanische Kandidat:innen gegen sie aus; der Bundesstaat New York hat bereits ein Moratorium für den Neubau von großen Rechenzentren erlassen, ohne die die rasante KI-Entwicklung unmöglich ist. Viele Menschen in den USA fürchten den Strom-, Wasser- und Flächenverbrauch und sind generell skeptisch gegenüber der KI, obwohl – oder vielleicht weil – dort die Programme noch deutlich mehr genutzt werden als etwa in Deutschland.

Über Widerstand in China ist dagegen weniger bekannt, die Bevölkerung dort gilt als technikfreundlicher und der Staat ist repressiver. Allerdings scheint auch dort die Angst vor Arbeitslosigkeit durch KI umzugehen: In Wuhan gab es Proteste gegen selbstfahrende Taxis, weil diese den Taxifahrer:innen ihre Einkommensquelle nehmen, und jüngst hat ein Gericht Entlassungen für illegal erklärt, die mit dem Einsatz von KI begründet werden. Der Widerstand von unten in einzelnen Staaten kann einen Beitrag leisten, er allein wird allerdings weder die Sicherheitsrisiken noch die wirtschaftlichen Gefahren verkleinern. Und so ist in der Theorie allen klar, dass zur Abwehr der KI-Risiken international zusammengearbeitet werden muss. Aber trotz all der Appelle peitscht in der Praxis die Konkurrenz um Marktanteile, um Einfluss und um militärische Macht die technische Entwicklung immer weiter voran und lässt damit auch die Gefahr einer unkontrollierbaren KI, sei sie nun selbstständig ausgebrochen oder absichtlich freigesetzt, steigen. ◻

18 Frédéric Lordon, *Awaiting the Crash?*, newleftreview.org, 9.7.2026.