

Wochenkommentar 44/2023 von Matthias Zehnder

# KI-Gipfel in Bletchley: Zynische Symbolik



Bild: KEYSTONE/CAMERA PRESS/ROTA

*Der britische Premierminister Rishi Sunak begrüsst am 1. November 2023 die amerikanische Vizepräsidentin Kamala Harris zum KI-Gipfel in Bletchley Park.*

**Politik und Wissenschaft sind sich einig: Ein Missbrauch der Atomkraft oder gar ein Atomkrieg wären verheerend für die Menschheit. Deshalb haben die Vereinten Nationen schon 1957 die Internationale Atomenergie-Organisation IAEA gegründet. Die Behörde soll dafür sorgen, dass die Atomkraft weltweit friedlich genutzt wird und kriegerischen Missbrauch verhindern. Nach diesem Vorbild möchte der britische Premierminister Rishi Sunak die Künstliche Intelligenz weltweit überwachen und einen Missbrauch der KI möglichst unterbinden. Sunak hat deshalb diese Woche Regierungen und Wirtschaftsvertreter aus den USA, China, der EU und der Schweiz zu einem KI-Sicherheitsgipfel nach England eingeladen. Dieser «AI Summit» hat im Bletchley Park stattgefunden. Sunak wollte damit wohl an die Geschichte des Ortes anknüpfen: Im Bletchley Park haben englische Mathematiker rund um Alan Turing im Zweiten Weltkrieg die Rechenmaschine erfunden, mit deren Hilfe sie die deutsche Verschlüsselungstechnik Enigma knacken konnten. Auf den ersten Blick klingt das alles sinnvoll, ja sympathisch. Und doch hätte Rishi Sunak für seinen Gipfel keinen schlechteren Ort wählen können. Denn dieser Bletchley Park steht auch für drei kapitale Fehler in der Computergeschichte. Es sieht ganz so aus, als hätte die Politik daraus nichts gelernt. Mein Wochenkommentar über die drei grossen Fehler der Politik im Umgang mit der Künstlichen Intelligenz.**

Seit Monaten warnen Politiker, Wissenschaftler und sogar die KI-Firmen selbst vor den Risiken der Künstlichen Intelligenz. Dabei geht es zum einen um den Missbrauch von KI-Diensten durch Kriminelle und Terroristen. Ihnen könnten es KI-Dienste einfacher machen, Waffen zu bauen, Cyberangriffe auszuführen, Viren zu entwickeln oder Falschinformationen in Umlauf zu bringen. Zum anderen geht es aber auch um die KI selbst: Nicht wenige Spezialisten fürchten, dass eine künftige KI dem Menschen so weit überlegen sein könnte, dass die Menschen die Kontrolle über die KI verlieren und umgekehrt die KI die Kontrolle übernimmt.

Die EU hat deshalb schon im Sommer ein KI-Gesetz vorgelegt. Das Gesetz soll den Einsatz von KI in der EU regulieren und die Risiken im Umgang mit KI reduzieren. Das EU-Parlament hat den Gesetzesentwurf im Juni angenommen. Seither laufen die Gespräche mit den EU-Mitgliedstaaten im Rat über die endgültige Ausgestaltung des Gesetzes. In den USA hat Präsident Joe Biden diese Woche ein Dekret erlassen, das Anbieter von künstlicher Intelligenz künftig zu Sicherheitstests verpflichtet. Dabei geht es insbesondere um Risiken für die nationale Sicherheit sowie die öffentliche Gesundheit.

### **Die Initiative von Rishi Sunak**

Der britische Premier Rishi Sunak wollte diese Woche noch einen Schritt weitergehen und hat deshalb zum internationalen «AI Summit», also zu einem KI-Gipfel, nach England eingeladen. Grossbritannien wollte die USA, die EU, China und die Wirtschaft an einen Tisch bringen. Sogar die Schweiz war eingeladen. Die Teilnehmer einigten sich in einer rechtlich nicht bindenden Vereinbarung darauf, dass künftig KI-Modelle auf ihre Gefahren für die Gesellschaft und die Sicherheit getestet werden sollen. Die weitere Entwicklung der KI soll künftig wissenschaftlich von einem Gremium begleitet werden, das mit dem Uno-Klimarat IPCC oder der Atomenergie-Organisation IAEA vergleichbar ist. Das Gremium soll die Entwicklung beobachten und Regierungen vor Gefahren warnen. Grossbritannien hat dafür bereits ein «AI Safety Institute» gegründet.

Das alles klingt vernünftig, auch wenn es eine etwas krude Mischung aus Selbstprofilierung von Rishi Sunak, Interessenpolitik der Länder und Marketing der Techfirmen ist. Ja, auch die Warnung vor den unglaublichen Fähigkeiten der KI kann PR für die Firmen sein. Und ja, Elon Musk war auch da. Trotzdem ist es eine gute Idee, ein internationales Gremium einzurichten, das die Entwicklung der KI im Auge behält und unabhängig vor Gefahren warnen kann. Das Problem ist der Ort, an dem dieser KI-Gipfel stattgefunden hat. Bletchley Park ist vordergründig einer der historischen Geburtsorte des Computers. Bletchley Park steht aber auch für drei kapitale Fehler der Computergeschichte. Drei Fehler, die nicht nur Grossbritannien im Umgang mit der KI gerade wiederholt.

### **Der erste Fehler: der Nationalismus**

Der Landsitz Bletchley Park nordwestlich von London war im Zweiten Weltkrieg der Standort der britischen Codebreaker: Hier haben die Mathematiker und Kryptoanalytiker gearbeitet, die den Auftrag hatten, die Verschlüsselung der deutschen Wehrmacht zu knacken. Die Deutschen setzten eine neuartige Verschlüsselungsmaschine ein: Die Enigma war so konstruiert, dass sich der Verschlüsselungscode mit jedem Buchstaben änderte. Das führte zu so vielen Kombinationsmöglichkeiten, dass kein Mensch den Code knacken konnte. Aber eine Maschine: Auf der Basis

der Ideen des britischen Mathematikers Alan Turing bauten die Codebreaker in Bletchley Park Colossus, einen Röhrencomputer. Mithilfe von Colossus gelang es ihnen mit der Zeit, die Funksprüche der Nazis zu knacken.

Heute ist Bletchley Park deshalb ein wichtiger Markstein in der Geschichte des Computers. Doch bis 1973 waren alle Aktivitäten in Bletchley Park ein streng geschütztes militärisches Geheimnis der Briten und «Top Secret». Bletchley Park steht deshalb auch für Geheimniskrämerei aus Nationalismus und die Instrumentalisierung der Informatik durch das Militär. Es ist geradezu zynisch, ausgerechnet an diesem Ort zur internationalen Zusammenarbeit aufzurufen. Die Geheimhaltung hatte übrigens auch für die Beteiligten ganz konkrete Folgen: Alan Turing und seine Kolleginnen und Kollegen durften zeitlebens nicht über ihre Arbeit in Bletchley Park reden. Für Alan Turing hatte das verheerende Konsequenzen.

### **Der zweite Fehler: Diskriminierung**

Alan Turing war homosexuell. Bis 1967 war Homosexualität in Grossbritannien ein Verbrechen. Sexuelle Handlungen unter Männern standen unter Strafe. 1952 wurde Alan Turing wegen homosexueller Handlungen angeklagt und in der Folge chemisch kastriert. 1954 starb er kurz vor seinem 42. Geburtstag an einer Zyanidvergiftung. Sein Tod wurde als Selbstmord behandelt, möglicherweise hat sich Alan Turing aber versehentlich selbst vergiftet. Erst 2009 hat sich der damalige britische Premierminister Gordon Brown im Namen der Regierung für die «entsetzliche Art und Weise», wie Turing behandelt wurde, entschuldigt. Königin Elizabeth II. begnadigte ihn im Jahr 2013 posthum. Ein Grund dafür, warum Alan Turing in den 50er-Jahren so hart angegangen wurde, war die Geheimhaltung: Ausserhalb des Geheimdienstes wusste niemand um seine Verdienste im Zweiten Weltkrieg. Für Polizei und Behörden war er bloss ein Freak.

Alan Turing ist nicht der einzige, dessen Arbeit in Bletchley Park lange nicht gewürdigt wurde. Auch der Beitrag, den Frauen bei der Decodierung geleistet haben, ist lange verschwiegen worden. Von den insgesamt rund 10'000 Personen, die im Bletchley Park arbeiteten, waren zu Spitzenzeiten 80 Prozent Frauen. Sie arbeiteten keineswegs nur in untergeordneten Funktionen, sondern auch in der Kryptoanalyse. Einer breiteren Öffentlichkeit bekannt ist Joan Clarke: Die Kryptoanalytikerin arbeitete im Team von Alan Turing. Sie war mit ihm eng befreundet, die beiden waren kurz sogar miteinander verlobt. Die Beziehung zwischen Alan Turing und Joan Clarke ist Thema eines Films mit Benedict Cumberbatch als Alan Turing und Keira Knightley als Joan Clarke: «The Imitation Game – Ein streng geheimes Leben» zeigt, wie stark Turing als Homosexueller und Clarke als Frau diskriminiert worden sind. Bletchley Park lässt sich deshalb auch als Chiffre für die gesellschaftliche Diskriminierung von Minderheiten und Frauen lesen. Wenn Rishi Sunak ausgerechnet da und ohne jede Referenz darauf einen Gipfel veranstaltet, der unter anderem verhindern will, dass die KI Minderheiten diskriminiert, dann ist das zynisch oder mindestens sehr, sehr kurzsichtig.

### **Der dritte Fehler: die Rolle des Menschen**

Bletchley Park ist heute eng mit dem Werk von Alan Turing verbunden und damit mit Künstlicher Intelligenz. Denn Alan Turing hat den ersten Test zur Beurteilung der Intelligenz einer Maschine entworfen. Das Inter-

essante an diesem Turing-Test ist, dass er für die grösste Gefahr steht, die von der Künstlichen Intelligenz ausgeht. Das grösste Risiko ist nämlich nicht die Maschine, sondern der Mensch, der sie bedient. Lassen Sie mich das kurz ausführen.

Der Turing-Test ist eine Art Intelligenztest für Maschinen. Ziel des Tests ist es, herauszufinden, ob eine Maschine so «menschlich» antworten kann, dass die Tester davon überzeugt sind, dass sie es mit einem Menschen zu tun haben. Konkret funktioniert das so, dass eine Testerin oder ein Tester über Tastatur mit einem Gegenüber spricht. Die Testerin, der Tester weiss dabei nicht, ob sie mit einem Menschen oder mit einer Maschine chattet. Eine Maschine besteht den Turing-Test dann, wenn der menschliche Tester nicht unterscheiden kann, ob er es mit einem Menschen oder mit einer Maschine zu tun hat. Das grosse Problem an diesem Setting ist, dass es nicht die kognitive Qualität der Maschine testet, sondern die Verführbarkeit des Menschen. Und Menschen sind sehr, sehr leicht verführbar.

### **Die erste Verführung durch Eliza**

Bestes Beispiel dafür ist Eliza: 1966 entwickelte der Informatiker Joseph Weizenbaum ein kleines Chat-Programm, das er «Eliza» taufte. Eliza basierte auf einem simplen Skript und konnte Gespräche mit einem Psychotherapeuten simulieren. Im einfachsten Fall greift es ein Wort auf und fragt nach dessen Oberbegriff. Auf den Input «Ich habe Probleme mit meiner *Mutter*» fragte Eliza wie ein guter Psychotherapeut: «Erzählen Sie mir mehr über Ihre *Familie*». Eliza war also alles andere als intelligent. Trotzdem fühlten sich viele Benutzer so, als würden sie mit einem menschlichen Therapeuten sprechen. Einige entwickelten sogar emotionale Bindungen zu dem Programm. Weizenbaum war überrascht über das Resultat und machte sich Sorgen über die starke emotionale Reaktion auf ein offensichtlich unintelligentes Programm.

Heutige KI-Programme wirken zwar sehr viel klüger als Eliza, sie verstehen von der Welt aber genauso wenig wie das Skript von Weizenbaum. Bei Lichte besehen sind es unglaublich eloquente, digitale Papageien, die sagenhaft schnell berechnen, welches Wort statistisch gesehen als nächstes am wahrscheinlichsten ist. Keines der Programme versteht, was es sagt. Sie sind aber so eloquent, dass die Menschen sehr rasch überzeugt sind, dass die Programme wirklich intelligent sind. Denn die Menschen sind leicht verführbar. Das war die grosse Schwäche des Turing-Tests. Wenn Rishi Sunak mit seinem KI-Gipfel in Bletchley Park Alan Turing feiert, dann feiert er letztlich die Verführbarkeit des Menschen. Diese Verführbarkeit aber ist die wirklich grosse Gefahr der KI.

### **Wie weiter?**

Es ist sicher gut, wenn die internationale Gemeinschaft sich um das Thema ernsthaft kümmert und ein Gremium im Stil des IPCC oder der IAEA einrichtet. Ich fürchte aber, dass die internationale Gemeinschaft die drei geschilderten Probleme nicht anpacken wird:

1. die Instrumentalisierung der KI durch die einzelnen Nationen und ihre Militärs
2. die Gefahr der Diskriminierung von Minderheiten durch eine auf Sta-

tistik beruhende Maschine und

3. die Leichtgläubigkeit der Menschen und die daraus resultierende grosse Verführbarkeit.

Die KI zu kritisieren bringt übrigens nichts. Die KI kann darauf nicht reagieren. Wir müssen die Menschen kritisieren: die Programmierer, die KI bauen und verkaufen, die Politiker, welche die regulatorischen Rahmenbedingungen dafür setzen und vor allem die Menschen, die KI nutzen und sich von ihr beeinflussen und beeindruckt lassen.

Das beste Gegenmittel wäre, wenn die Nationen dieser Welt etwas weniger in die Intelligenz von Maschinen und etwas mehr in die Intelligenz der Menschen investieren würden. Das heisst: in Bildung, kritisches Denken und in die intellektuelle Mündigkeit. Ich fürchte aber, dass daran weder die grossen Regierungen noch die grossen Unternehmen ein echtes Interesse haben. Deshalb: Nehmen Sie Ihre Intelligenz selbst in die Hand. Überlassen Sie das Denken nicht einer Maschine. Denken Sie selbst.

Basel, 3. November 2023, Matthias Zehnder [mz@matthiaszehnder.ch](mailto:mz@matthiaszehnder.ch)

## Quellen

Clarke, Laurie; Dickson, Annabelle und Galliaro, Cristina (2023): *Rishi Sunak wants to lead the world on AI. The world ain't listening*. In: Politico. [<https://www.politico.eu/article/rishi-sunak-ai-technology-wants-to-lead-the-world-on-ai-the-world-aint-listening/>; 3.11.2023].

Department for Science, Innovation & Technology (2023): *The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023*. In: GOV.UK. [<https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>; 2.11.2023].

Department for Science, Innovation & Technology (2023): *Introducing the AI Safety Institute*. In: GOV.UK. [<https://www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute>; 3.11.2023].

Die Zeit (2023): *Künstliche Intelligenz: US-Präsident Biden Verschärft Regulierung Künstlicher Intelligenz*. In: ZEIT ONLINE. [<https://www.zeit.de/politik/ausland/2023-10/usa-kuenstliche-intelligenz-staerkere-regulierung>; 3.11.2023].

Europäisches Parlament (2023): *KI-Gesetz: erste Regulierung der künstlichen Intelligenz*. In: Europarl. [<https://www.europarl.europa.eu/news/de/headlines/society/20230601STO93804/ki-gesetz-erste-regulierung-der-kuenstlichen-intelligenz>; 3.11.2023].

Hale, Craig (2023): *28 countries agree to develop AI safely and responsibly at Bletchley Park summit*. In: TechRadar. [<https://www.techradar.com/pro/28-countries-agree-to-develop-ai-safely-and-responsibly-at-blet>

[chley-park-summit](#); 2.11.2023].

Kleinman, Zoe und Gerken, Tom (2023): *King Charles: Tackle AI risks with urgency and unity*. In: BBC News. [<https://www.bbc.co.uk/news/technology-67172229>; 2.11.2023].

Nuspliger, Niklaus (2023): *Verlieren die Menschen die Kontrolle über die Maschinen? KI-Entwickler sollen ihre Modelle künftig von einem Sicherheitsinstitut überprüfen lassen*. In: Neue Zürcher Zeitung. [<https://www.nzz.ch/international/verlieren-die-menschen-die-kontrolle-ueber-die-maschinen-ki-entwickler-sollen-ihre-modelle-kuenftig-von-einem-sicherheitsinstitut-ueberpruefen-lassen-ld.1763768>; 3.11.2023].

Ohne Autor. (2023): *Why the UK-led global AI summit is missing the point*. In: Nature 623/7985. S. 7–7. doi:<https://doi.org/10.1038/d41586-023-03333-7>. [<http://dx.doi.org/https://doi.org/10.1038/d41586-023-03333-7>; 3.11.2023].

Stacey, Kiran und Milmo, Dan (2023): *UK, US, EU and China sign declaration of AI's 'catastrophic' danger*. In: The Guardian. [<https://www.theguardian.com/technology/2023/nov/01/uk-us-eu-and-china-sign-declaration-of-ais-catastrophic-danger>; 2.11.2023].

Weiß, Eva-Maria (2023): *KI-Gipfel in Bletchley Park: Staaten einigen sich auf Zusammenarbeit*. In: Heise Online. [<https://www.heise.de/news/KI-Gipfel-in-Bletchley-Park-Staaten-einigen-sich-auf-Zusammenarbeit-9350518.html>; 2.11.2023].

**Spenden für den Wochenkommentar ist jetzt so einfach wie bezahlen im Hofladen**



**Bequem mit TWINT bezahlen**

Scannen Sie den QR-Code mit Ihrer TWINT App.

Geben Sie den Totalbetrag ein und bestätigen Sie Ihre Zahlung.



**Sie möchten lieber über einen Einzahlungsschein spenden? Kein Problem:**

[https://www.matthiaszehnder.ch/wp-content/uploads/2023/04/Spende\\_EZ\\_Wochenkommentar.pdf](https://www.matthiaszehnder.ch/wp-content/uploads/2023/04/Spende_EZ_Wochenkommentar.pdf)