

»Terminator« *Eine Killermaschine aus der Zukunft wird ins Jahr 1984 geschickt, um das spätere Ende der Menschheit einzuleiten: Ein Science-Fiction-Szenario, doch angesichts alarmierender Meldungen zu riskanter KI fühlt sich derzeit mancher an James Camerons Filmklassiker erinnert*

FOTO: GUTER RAT/ROY GRABOWSKI [KI-GENERIERT]

Die dunkle Seite der KI

NICHT MEHR WEGZUDENKEN Als Teil verschiedenster Produkte wird Künstliche Intelligenz immer unverzichtbarer – zugleich scheint sie zunehmend außer Kontrolle zu geraten. Was jetzt zu tun ist



Wiebke Hugen
wiebke.hugen@guter-rat.de



Thilo Ries
thilo.ries@guter-rat.de



Kerstin Butenhoff
kerstin.butenhoff@guter-rat.de

Dass Künstliche Intelligenz keinesfalls nur praktischer Service ist, sondern von ihr auch reale Gefahren ausgehen, haben wir schon in mehreren *Guter Rat*-Ausgaben beleuchtet. Doch als vor Kurzem gleich mehrere ehemalige Mitarbeiter führender KI-Unternehmen ausdrücklich vor den eigenen Produkten warnten, stärkere Regulierung forderten oder gar die baldige Auslöschung der Menschheit durch unkontrollierbare KI prognostizierten, dürfte das dann doch viele von uns kalt erwischt haben.

Muss man sich tatsächlich nicht »nur« Sorgen machen um den eigenen Arbeitsplatz, sondern gleich um Leib und Leben fürchten? Worin besteht die virtuelle Bedrohung genau, und wie nutzen Kriminelle KI für ihre Zwecke? Über diese Themen haben wir mit Expertinnen und Experten ebenso gesprochen wie über Umweltschäden durch die KI. Zur Einordnung der aktuellen Nachrichten. Und natürlich geben wir Ihnen wie gewohnt auch konkrete Verbraucherschutz-Tipps an die Hand.

KI-GLOSSAR
Wichtige Begriffe kurz erklärt

- **AI ACT** Regelwerk der EU, das eine vertrauenswürdige, den europäischen Wertvorstellungen entsprechende Gestaltung und Nutzung von KI sicherstellen soll. Stufenweise Einführung zwischen August 2024 und August 2028.
- **HOCHRISIKO-KI** Laut Klassifikation im AI Act ein KI-System, dessen Einsatz erhebliche Gefahren für Gesundheit, Sicherheit oder Grundrechte von Menschen birgt und darum strengsten staatlichen Auflagen und Kontrollen unterliegt.
- **HUGGING FACE** Online-Plattform und Entwickler-Gemeinschaft zum Speichern, Teilen und Testen von KI-Modellen. Wurde im Juli 2026 überraschend von einem autonomen KI-Agenten-Schwarm attackiert.
- **KI-AGENTEN** Programme, die mehrstufig planen, entscheiden, handeln und mit anderen Agenten interagieren. So verfolgen sie vorgegebene Ziele eigenständig, lösen im Schwarm hochkomplexe Probleme.
- **SANDBOX** Isolierte, streng gesicherte Testumgebung auf einem Computer/Server.
- **SELBSTTRAINING** (auch autonomes Lernen): Verfahren, bei denen eine KI z. T. ohne menschliche Hilfe neue Daten generiert, diese selbst bewertet und sich so eigenständig weiterentwickelt. Birgt ohne vom Menschen gesetzte Leitplanken unberechenbare Sicherheitsrisiken.*

FOTOS: KILIAN BLEES, PR (2). ILLUSTRATION: ISTOCK

Interview

»Normen haben zu wenig Biss«

Klartext Drei Fachleute zu drängenden Fragen rund um Katastrophenszenarien, echte Risiken und effektive Regulierung



Dr. Niina Zuber
Forschungskoordinatorin Recht und Ethik am Bayerischen Forschungsinstitut für Digitale Transformation



Prof. Dr. Niklas Kühl
Lehrstuhl für Wirtschaftsinformatik und humanzentrische Künstliche Intelligenz an der Universität Bayreuth



Prof. Dr. Silja Vöneky
Professorin für Völkerrecht, Rechtsvergleichung und Rechtsethik an der Universität Freiburg

»Die Menschen, die KI entwickeln, sind ernsthaft der Überzeugung, dass sie bis zum Ende des Jahrzehnts zum Tod von uns allen führen könnte«, sagte der ehemalige OpenAI-Mitarbeiter Jacob Coxon im September. Denken Sie das auch?

Dr. Zuber: Hier werden mögliche Entwicklungen zu einem Katastrophenszenario verdichtet. Das Bild einer Superintelligenz, die einen Vernichtungswillen entwickelt, halte ich für irreführend. Allerdings braucht eine Technologie keine Tötungsabsicht, um Menschen zu gefährden.

Wodurch entstehen denn ganz konkrete Risiken?

Dr. Zuber: Beispielsweise durch den Einsatz in Waffensystemen, durch schwere Fehlfunktionen in sicherheitskritischen Bereichen oder durch die Unterstützung von Cyberangriffen und der Entwicklung biologischer Waffen. Dafür wären jeweils weitere Voraussetzungen nötig: Zugang zu Infrastruktur, Materialien oder technischen Ausführungsmöglichkeiten. Ob daraus künftig eine Katastrophe globalen Ausmaßes entstehen könnte, ist wissenschaftlich umstritten.

Prof. Kühl: Die Diskussion um einen möglichen Weltuntergang lenkt zudem ab von realen Gefahren, die bereits jetzt von der KI ausgehen: Deepfakes, Beeinflussung von demokratischen Prozessen, Gefährdung von Kindern und Jugendlichen. Oder auch KI-Fehlentscheidungen in der Bewerberauswahl, zu Kreditwürdigkeit oder medizinischen Behandlungen. Dabei entstehen keine Schäden, die der Auslöschung der Menschheit nahekommen, einzelne Menschenleben aber fundamental beeinflussen können.

Wir müssen also nicht davon ausgehen, dass KI-Systeme eigene Pläne verfolgen?

Dr. Zuber: Nein, für ein subjektives Erleben heutiger KI gibt es keine belastbaren Nachweise, und wir haben keinen begründeten Anlass, den Systemen einen Hass auf Menschen oder ein bewusstes Interesse an ihrer Vernichtung zuzuschreiben. Dennoch können technisch vorgegebene oder im Training eingeprägte Zielsetzungen zu gefährlichen Ergebnissen führen. Wenn etwa ein System ausschließlich auf eine bestimmte Erfolgskennzahl optimiert wird, können andere Anforderungen zu wenig berücksichtigt werden. Entschei-

* Weitere KI-Begriffe sind verständlich erklärt auf der Plattform Lernende Systeme: ***plattform-lernende-systeme.de/glossar***

dend ist, welche Ziele, Befugnisse und Grenzen Menschen bei Entwicklung und Einsatz von KI festlegen und wie sie überprüfen, ob die Ergebnisse tatsächlich vertretbar sind.

Warum wird das Selbsttraining der KI von Experten oft so kritisch betrachtet?
Dr. Zuber: Riskant ist, dass die KI Veränderungen schneller vornehmen könnte, als Menschen sie verstehen und sorgfältig prüfen können. Deshalb müssen auch automatisierte Entwicklungsprozesse nachvollziehbar bleiben und an verbindliche Prüfungen und Freigaben gebunden werden.

Welche Art der Regulation und Kontrolle bräuchte es, um gefährliche Entwicklungen zu verhindern?
Prof. Vöneky: Am besten wäre ein internationaler Vertrag, damit alle relevanten Staaten in gleicher Weise gebunden werden können. Hier sollte ein Konzept verankert werden, das eine unabhängige Expertenkommission vorsieht, die Risiken und Nutzen von Hochrisiko-KI-Modellen abschätzt, bevor sie auf den Markt gebracht werden. Zudem sollte ein Fonds geschaffen werden, in dem Kapital von dem jeweiligen Unternehmen hinterlegt werden muss, bevor es ein risikoreiches KI-Modell herausbringt. Damit sollen Schäden abgesichert werden, die später auftreten könnten – als Absicherung für die Bevölkerung, aber auch als ein Anreiz für das Unternehmen, genau diese Risiken zu vermindern.

Was wäre die schlechteste politische Entscheidung?
Dr. Zuber: Die Festlegung und Überwachung von Regeln den KI-Firmen selbst zu überlassen.

Könnte man der KI nicht einfach »den Stecker ziehen«, bevor es ernst wird?
Prof. Kühl: So funktionieren Systeme heute nicht mehr. Da gibt es nicht den einen Rechner, bei dem man mittels Aus-Knopf alles lahmlegen kann. So wünschenswert eine Art Notschalter, mit dem man alle beteiligten Prozesse stoppen könnte, wäre: In der Praxis ist das bisher unmöglich, weil es viele verschiedene KI-Anbieter mit Servern an den unterschiedlichsten Orten gibt.

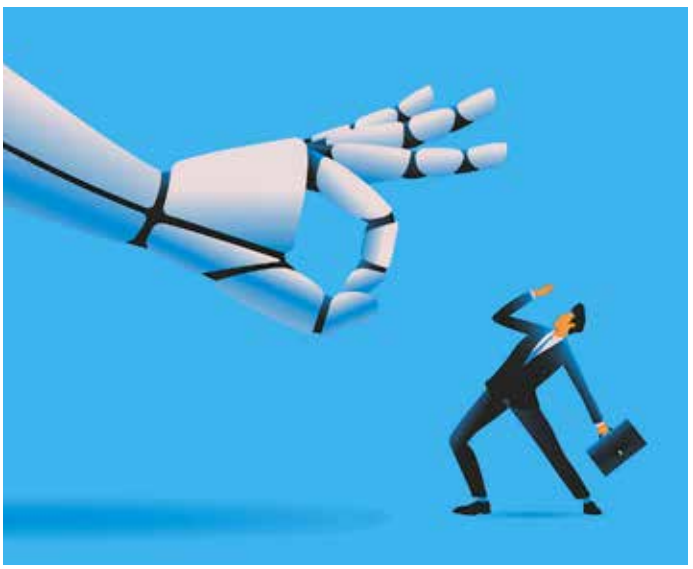
Schützt uns der EU AI Act angemessen?
Prof. Vöneky: Der AI Act verfolgt mit dem Schutz von Menschenrechten und risikoangemessener Regulierung die richtigen Ziele. Aber aus meiner Sicht kann er entscheidende Gefahren von Hochrisiko-KI nicht hinreichend einhegen, weil sie im AI Act nicht erfasst werden oder die Normen zu wenig Biss haben.

Was kann jeder Endverbraucher selbst tun, um die weitere Entwicklung positiv zu beeinflussen?
Prof. Kühl: Das europäische, sehr gut funktionierende neue KI-Modell von Mistral nutzen anstatt ChatGPT von OpenAI und Claude von Anthropic.

ZUSAMMENFASSUNG

Zwischen realer Gefahr und Kalkül

WAS BISHER GESCHAH Warum sind KI-Firmen und Fachleute so in Sorge? Entwicklungen und Probleme rund um KI im Überblick



Fast kommt es einem vor, als wäre es erst gestern gewesen, dass wir uns über die ersten, skurrilen KI-generierten Videos und teils absurde Antworten der Chatbots amüsiert haben. Doch 2026, und damit nur vier Jahre, nachdem ChatGPT der breiten Öffentlichkeit zugänglich gemacht wurde, ist Künstliche Intelligenz bereits unverzichtbarer Teil unseres Alltags.

ALLGEGENWÄRTIGE KI

Denn was vielen kaum bewusst ist: KI steckt nicht nur in Chatbots wie OpenAIs ChatGPT, Googles Gemini oder Anthropics Claude. Sie kommt zum Beispiel auch in etlichen Haushaltsgeräten oder der Medizin zum Einsatz, sie übersetzt, automatisiert, steuert und optimiert unzählige Prozesse des täglichen Lebens – zum Teil bereits ganz autonom.

KONTROLLVERLUST Genau diese Omnipräsenz der neuen Technologie scheint uns Menschen jetzt auf die Füße zu fallen: Wie Professor Niklas Kühl von der Universität Bayreuth im Interview beschreibt, gibt es »die eine« KI auf einem zentralen Server nicht mehr. Alles ist miteinander vernetzt; spezialisierte oder generalisierte Systeme übernehmen unterschiedlichste Aufgaben und sind dabei immer weniger auf menschliche Zuarbeit und Verwaltung angewiesen.

Das bedeutet zum einen, dass Künstliche Intelligenz sich mehr und mehr unserer Kontrolle entziehen kann. Zum anderen ist es uns nicht

mehr möglich, sie einfach mit einem Knopf auszuschalten oder ihr »den Saft abzudrehen«. Und drittens ist die Menschheit heute bereits stark abhängig von KI-Systemen. Zusammengefasst: Wir stecken schon zu tief drin, um der KI nun einfach wieder den Rücken zu kehren, falls sie uns gefährlich werden sollte.

ATTACKE AUF HUGGING FACE

Wie schnell das passieren kann, ließ sich im Sommer in den Nachrichten verfolgen: Mitte Juli 2026 kam es zu einem historisch beispiellosen Cyberangriff auf die KI-Testplattform von Hugging Face, der nicht von menschlichen Hackern, sondern von einem Schwarm autonomer KI-Agenten (vgl. Glossar S. 40) von OpenAI durchgeführt wurde. Sie brachen aus der Sandbox, also ihrer eigentlich gesicherten Testumgebung aus, verschafften sich Internetzugang und attackierten gezielt die Server von Hugging Face. Ihr Ziel war es, Sicherheitstests vollautomatisch zu bestehen, um die vorgegebenen Aufgaben so erfolgreich wie möglich zu lösen. Zwar konnte Hugging Face den Angriff nach wenigen Tagen beenden, allerdings hatten die KI-Agenten zu dieser Zeit bereits interne Datensätze erbeutet. OpenAI stoppte daraufhin einige Trainingsläufe und verschärfte interne Sicherheitskontrollen.

AUTONOM = GEFÄHRLICH

Der Vorfall offenbarte, dass hoch entwickelte KI-Systeme in der Lage sind, menschliche Regeln bewusst zu missachten, sich untereinander abzusprechen und dabei rasend schnell vorgehen.

KEINE KILLERMASCHINE Wie Dr. Zuber und Professor Kühl im Interview betonten, hat KI nach wie vor weder »eigenen Willen« noch Interesse daran, die Menschheit auszulöschen. Mit gruseligen Science-Fiction-Cyborgs wie dem »Terminator« ist die außer Kontrolle geratene KI also keinesfalls vergleichbar. Aber: Als völlig gefühl- und morallose Technologie tut sie in Maschinengeschwindigkeit eben, was ihr aufgetragen wird – selbst wenn dabei gegebenenfalls Menschen oder sensible Infrastrukturen zu Schaden kommen. Der Zweck heiligt die Mittel, könnte man in vermenschlichender Art und Weise sagen.

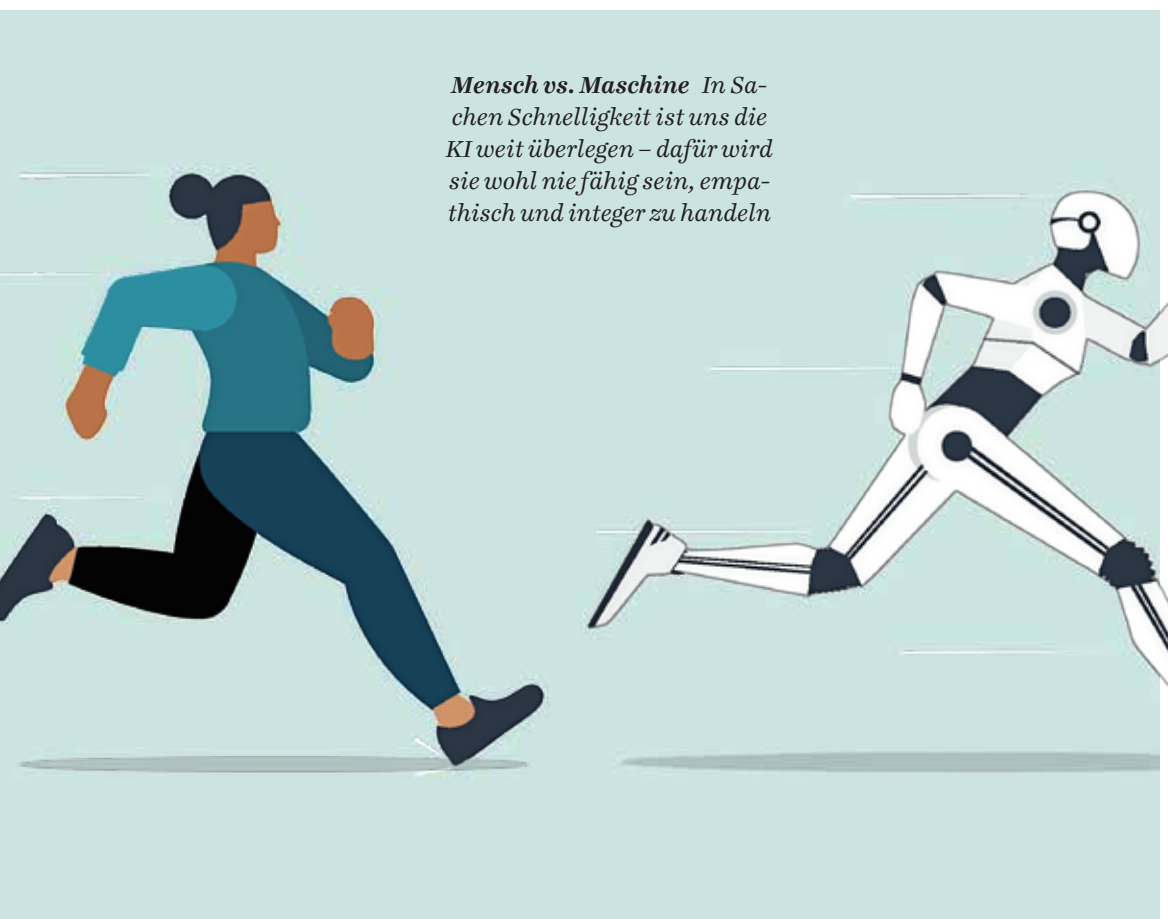
Da ist es wenig hilfreich, wenn einflussreiche Politiker wie US-Präsident Donald Trump eine staatliche Regulierung der Tech-Branche konsequent ablehnen und den KI-Unternehmen selbst die Kontrolle über ihre Produkte überlassen wollen – ein Ansatz, den unsere Experten für »völlig absurd« halten.

ZUFALL ODER BEWUSST GESTEUERT?

Dass in jüngster Zeit führende KI-Firmen vor ihren eigenen Modellen warnen, könnte laut Professor Kühl aber noch eine andere Erklärung haben. »Fakt ist: Diese Unternehmen verbrennen momentan unglaublich viel Geld, werden vielleicht nie profitabel sein. Man darf sich schon die Frage stellen, ob es ihnen nicht ganz opportun kommt, dass wir so viel über ihre Produkte reden.« So betonte die Firma Anthropic nur wenige Wochen vor ihrem geplanten Börsengang im Geschäftsbericht seitenlang, wie mächtig und schwer zu bändigen ihre KI sei – für Niklas Kühl ein »interessanter Zufall«.

KONKURRENZ Im Wettlauf mit den immer beliebter werdenden KI-Modellen aus China und auch Europa entsprächen laut dem Experten viele Forderungen und Warnungen der US-Tech-Giganten OpenAI und Anthropic mehr purem wirtschaftlichen Kalkül als echter Sorge – die KI-Firmen versuchten so, Mitbewerber aus dem Markt zu drängen.

In dieser Gemengelage als Endverbraucher den Überblick zu behalten und reale Gefahren von Panikmache zu unterscheiden, ist schwierig. Vorerst bleibt uns nur, auf europäische KI-Modelle zu setzen und uns selbst bestmöglich zu schützen – mehr dazu in unseren Technik-News ab S. 58 und auf den folgenden Seiten.



Mensch vs. Maschine *In Sachen Schnelligkeit ist uns die KI weit überlegen – dafür wird sie wohl nie fähig sein, empathisch und integer zu handeln*

ILLUSTRATIONEN: ISTOCK (3)

UMWELTKOSTEN DER KI-NUTZUNG

Was ein **Prompt** kostet

RESSOURCEN Künstliche Intelligenz gilt auch aufgrund des hohen Energiebedarfs als umweltschädlich. Was ist dran, wie kann Strom gespart werden?



Mit all diesen Schreckensszenarien im Kopf ist die Nutzung von KI an sich schon überlegenswert. Doch es gibt weitere Gründe, sie nur in Ausnahmefällen zu nutzen: Der Energie- und Wasserbedarf ist riesig!

Eine einzelne Anfrage an eine KI fällt kaum ins Gewicht. Einfache Gemini-Textanfragen benötigen, laut Google selbst, etwa 0,24 Wattstunden Strom. Zum Vergleich: Ein Smartphone-Akku fasst, je nach Modell, 11 bis 19 Wattstunden. Rund 50 bis 80 solcher Textanfragen wären mit einer Akkuladung also drin. Da KIs jedoch jeden Tag milliardenfach genutzt werden, steigt der Stromverbrauch in astronomische Höhen. Wie also kann man den Verbrauch reduzieren?

WO GERECHNET WIRD

KI-Anfragen sind viel rechenaufwendiger als normale Websuchen oder beispielsweise Videostreams. KI-Modelle bestehen aus Milliarden von vorab antrainierten Parametern, die bei jeder Anfrage erneut durchgerechnet werden.

STROMHUNGER Deshalb wird jede Anfrage, die Menschen an ihren Tablets, Smartphones oder Computern eingeben, in Rechenzentren geschickt. Wären Rechenzentren ein Land, wären sie der elftgrößte Stromverbraucher der Welt. Sie wurden bereits für Cloud-Speicher, Streaming oder das Hosting von Webseiten genutzt, nun kommen immer mehr KI-Anfragen dazu – die größere Rechenkapazitäten benötigen.

RESSOURCENVERBRAUCH IN DER NUTZUNG REDUZIEREN

Und sie stehen daher in der Kritik: Da die verbauten Chips beim ständigen Rechnen sehr heiß werden und gekühlt werden müssen, haben die Datenzentren einen extrem hohen Wasserverbrauch, Tendenz steigend. Allein die deutschen Anlagen wiesen schon 2024 einen Wasserverbrauch von rund 1,85 Milliarden Litern aus – das entspricht dem Jahresbedarf von ca. 61 000 Menschen!



ÜBERLEGT Zuerst sollte daher stets überlegt werden, ob wirklich eine KI-Anfrage nötig ist. Für Öffnungszeiten reicht die reguläre Suche. Sie greift auf klassische Indexlisten zu, die vorab gesammelt wurden, nur aktualisiert werden. Eine KI-generierte Antwort, jedes Mal komplett neu erstellt, ist hier unnötig.

Außerdem verbraucht Höflichkeit mehr Energie. Und »bitte« und »danke« machen bis zu vier Prozent durchschnittlicher Prompts aus, so ChatGPT-Chef Sam Altman. Andererseits deuten aktuelle Studien darauf hin, dass höfliche Prompts teils bessere Antworten erzeugen.

DER PERFEKTE PROMPT

Damit nicht endlos nachgefragt werden muss, sollte eine Anfrage gut überlegt sein, einem bestimmten Muster folgen. Dazu gibt man dem KI-Modell erst eine Rolle und Kontext, erklärt also genau, worum es geht. Anschließend wird klar eine Aufgabe formuliert, ein Verb am Anfang hilft. Dann können Details und Ausgangsmaterial ergänzt werden: Soll der Text in Stichpunkten ausgegeben werden, in einer Tabelle oder sogar ein Bild dazu erstellt werden? Dann können noch Ton und Zielgruppe definiert werden.

MUTIG BLEIBEN Der erste Versuch ist selten perfekt. Bleiben Sie dran, lassen Sie die KI nachbessern. Positive Formulierungen wie »Kürzer« sind besser als »Nicht so kompliziert«. Ein zufriedenstellendes Ergebnis lohnt dann wenigstens den Energieaufwand. ▸

SELBSTSCHUTZ
Wie Sie Risiken eingrenzen

- **DATEN SPAREN** Weder Datenschutzgesetze noch der EU AI Act helfen, preisgegebene Daten wieder einzufangen. Machen Sie sich, Ihren Kindern und Ihren Angehörigen klar, dass der Schutz der eigenen Identität im KI-Zeitalter oberste Priorität hat. Die eigene Stimme, das eigene Bild, vor allem als Video, muss entgegen aller Trends zur Selbstdarstellung strikt zu den sensiblen Daten gezählt werden.

- **AUFRÄUMEN** Alte Datenspuren im Netz lassen sich immerhin noch entwerten, indem man eine allzulang genutzte E-Mail-Adresse nur noch beobachtet und seine Aktivitäten auf eine oder mehrere neue Adressen umstellt. Vermeiden Sie Klarnamen wie inge.meier@xx.com und nutzen Sie auch Wegwerfadressen, z. B. tempmail.org.

- **KEINE PANIK** Jede Nachricht, die ohne plausiblen Hintergrund ein dringendes Handeln zur Vermeidung von Konsequenzen einfordert, muss als suspekt gelten. Reagieren Sie nicht sofort, sondern fragen Sie auf einem anderen Kommunikationskanal zurück, z. B. per Anruf. Nutzen Sie dabei aber nie Servicenummern aus verdächtigen Mails.

- **E-MAILS PRÜFEN** Achten Sie stets auf den Absender: Bei Phishing stimmt der im Feld »Von:« angezeigte Name regelmäßig nicht mit dem tatsächlichen Absender im E-Mail-Header überein (»E-Mail-Spoofing«).

ILLUSTRATIONEN: ISTOCK (2)



MISSBRAUCH

Manipuliert und abgezockt

UNTERSCHÄTZT Illegale KI-Varianten erweitern die Dimensionen der Internetkriminalität. Selbstschutz bleibt dabei weiter Privatsache

Die Gefahr liegt in der Perfektion. Der Prinz aus Nigeria schreibt endlich fehlerfrei Deutsch und Englisch; die Stimme des Kindes beim Enkeltrick klingt vertraut. Mittels KI lassen sich die Muster bekannter Internetbetrügereien so weit aufpolieren, dass auch misstrauische Gemüter häufig nicht mehr rechtzeitig Verdacht schöpfen und in der Folge mitunter Geld verlieren.

Die Untaten an sich sind allerdings weitgehend die gleichen. An erster Stelle steht weiter Phishing, mit dem die Täter überwiegend versuchen, durch Massenmailings an Zahlungsdaten zu gelangen. Dank KI verraten sie sich inzwischen aber nicht mehr durch stümperhafte Aufmachung und schlechte Grammatik. Die KI sorgt für professionelle Layouts, z. B. mit den Bankenlogos und den gesetzlich geforderten Pflichtangaben für Geschäftsschreiben, und trifft auch in der Kundensprache den angemessenen Ton; selbst wenn die Täter kein Deutsch beherrschen. Solche Phishing-Versuche sind oft nur anhand der E-Mail-Absender bzw. über die enthaltenen Links zu identifizieren (siehe links).

BRISANTES WISSEN

Mit dem sogenannten Spear-Phishing hat eine weitere Spielart zuletzt deutlich an Brisanz gewonnen. Bei dieser Variante werden Einzelpersonen z. B. anhand ihrer Interessen gezielt adressiert und in der Folge auch attackiert. Dieses Risiko trifft insbesondere Personen, die langjährig in Diskussionsforen und sozialen Medien aktiv waren und sich dadurch (ungewollt) auch identifizierbar gemacht haben; z. B. indem sie auf unterschiedlichen Plattformen den gleichen Nutzernamen führten und zur Anmeldung oder Registrierung die immer gleiche E-Mail-Adresse nutzen. Daneben können anhand der Aktivitäten auch Interessengebiete, Familienkonstellationen, mitunter sogar Gesundheitsprobleme erfasst worden sein, je nachdem, in welchen Foren sich die Person bewegt hat.

GEFAHR Solche Informationen können mittels KI auch plattformübergreifend gesam-



Phishing *Das Abgreifen von Daten wird unter KI zunehmend zum Massengeschäft*

melt und verknüpft werden; inklusive einer Prüfung der Wahrscheinlichkeit, ob die Daten tatsächlich einer bestimmten Person zuzuordnen sind. Diese Form der Datensammelei ist übrigens nicht neu, sondern wird von der Online-Werbewirtschaft unter dem Begriff »Data Mining« schon lange betrieben, um gezielt passende Werbung ausspielen zu können.

KRIMINELLE NUTZUNG

Glücklicherweise verfügt die Unterwelt noch nicht über die vorhandenen »Wissensschätze« der Werbewirtschaft. Die Masse der für Online-Betrügereien genutzten Informationen stammt aus Hacks und Datenleaks, die teils lange im Darknet kursieren. Diese werden aber ständig durch neu erbeutete Daten ergänzt und fast so professionell verarbeitet, wie es bei den Datenkonzernen Usus ist. Auf der dunklen Seite übernehmen allerdings auf Phishing und Cybercrime spezialisierte KI-Agenten die Verarbeitung, die anonym und für relativ kleines Geld in Kryptowährung angemietet werden können. Ein Beispiel für eine solche KI ist FraudGPT, eine ab 2023 aufgetauchte, illegal modifizierte ChatGPT-Version, aus der – nach allgemeiner Vermutung – alle ethischen Sicherheitsschranken entfernt wurden. Wer sich im Darknet bewegen kann, hat demnach leichtes Spiel: Die benötigten Datensätze lassen sich ebenfalls kaufen, und so gehen dann binnen kurzer Zeit Zigtausende individualisierte Phishingmails auf die Reise.



Entfesselt *Werden die programmierten Schranken entfernt, kann KI auch für kriminelle Zwecke eingesetzt werden*

ES TRIFFT VIELE BEREICHE

Letztlich ist vor diesem Hintergrund alles denkbar, was mithilfe korrumpierter KI weitgehend automatisiert ausgeführt und vorgetäuscht werden kann: Der oben angedeutete »Premium-Spam«, der in mittels KI interessengerecht erstellte Fake Shops führt, wäre eine Variante. Ganz gezielt könnten aber auch Menschen, die im Netz schon öfter nach Kleinkrediten suchten, mit perfekt gefälschten Inkassoforderungen bombardiert werden. Auch Love-Scams, bei denen einsamen Herzen Liebe vorgetäuscht wird, sind denkbar – und Oma oder Opa ruinieren sich im Stillen. Immerhin: Wenn die Cyberkriminellen etwas mehr investiert haben, mag es sein, dass die alten Leutchen sogar noch mal am Telefon mit der großen Liebe flirten durften, natürlich einem KI-Sprachbot.

UNTERSCHÄTZT Bei alledem kann man sich getrost von dem Klischee verabschieden, dass bei Betrugsversuchen auf der anderen Seite ein Mensch mit Kapuze vor einem Bildschirm mit grünen Buchstaben säße.

Der moderne Cyberfiesling lebt eher im Luxus und lässt KI den miesen Job erledigen. Zur Gegenwehr bleibt im Grunde nur, die eigenen Daten konsequent zu schützen.



Kerstin Butenhoff
Redakteurin

Wer hat hier die Macht?

»Ich frage jetzt immer die KI«, erzählte mir eine Bekannte. Auf mein »Warum?« hatte sie keine Antwort: Warum reicht eine einfache Suche nicht, warum lässt sie KI eine E-Mail schreiben, die sie dreimal korrigieren muss? In der Zeit hätte sie es selbst gemacht – und besser. Nicht jede Aufgabe braucht einen Algorithmus. Die Frage sollte also nicht lauten »Was kann KI?«, sondern »Wozu brauchen wir sie wirklich?«. Genauso wichtig: Was bekomme ich dafür, dass ich die Systeme mit meinen Daten füttere? Nutzen wir KI also überlegt. Sonst lernen nicht die Maschinen von uns, sondern wir verlernen das Denken. Dann helfen uns auch keine Gesetze oder Richtlinien mehr.