»Ansätze wie jene der Europäischen Kommission, KI in der Cybersicherheit generell zu vertrauen, sind irreführend und gefährlich«, warnt Mariarosaria Taddeo.

»Vertrauen in künstliche Intelligenz ist ein zweischneidiges Schwert«

Mariarosaria Taddeo ist Senior Research Fellow am Oxford Internet Institute, Oxford University, und stellvertretende Direktorin des Digital Ethics Lab. Report(+)PLUS hat anlässlich der IDSF-Konferenz in Wien mit ihr über Robustheit und Ethik beim möglichen Einsatz von künstlicher Intelligenz in der Cybersicherheit gesprochen. VON MARTIN SZELGRAD

(+) PLUS: Wie ist die Bedrohungslage bei Angriffen auf IT-Systeme? Warum könnte der Einsatz von KI dort notwendig sein?

Mariarosaria Taddeo: Das Cybersecurity-Unternehmen Norse stellte im Jahr 2014 mehr als 4.000 Cyberangriffe pro Minute auf IT-Systeme weltweit fest. Trotz ausgefeilter Cyberabwehr hat sich die Bedrohungslage bis heute nicht verbessert. Dem »Global Risks Report 2019« des Weltwirtschaftsforums zufolge gehören Cyberangriffe zu den fünf größten globalen Bedrohungen. Gemalto berichtet, dass im ersten Halbjahr 2018 durch Angriffe 4,5 Milliarden Datensätze kompromittiert wurden – fast doppelt so viel wie im gesamten Jahr 2017. Und eine Microsoft-Studie wies nach, dass 60 % der Angriffe im Jahr 2018 weniger als eine Stunde dauerten und neue Formen von Malware gebracht haben. Das zeigt uns, dass Cyberangriffe weiter zunehmen und auch effektiver werden. KI kann in der Unterstützung von Cybersecurity-Maßnahmen nun eine Schlüsselrolle spielen.

(+) PLUS: Wie definieren Sie KI in Cybersicherheitssystemen?

Taddeo: Abgesehen von einem allgemeinen Hype um dieses Thema wird der Einsatz von KI vielfältig sein und von unterschiedlichen Methoden abhängen. Ich definiere – dem klassischen Turing-Ansatz zufolge – KI als eine Ressource interaktiver, autonomer und selbstlernender Systeme, die zur Ausführung von Aufgaben eingesetzt werden können, die ansonsten menschliche Intelligenz erfordern würden.

Es gibt zwei wichtige Aspekte in dieser Definition. Der erste hat mit der Form von Intelligenz zu tun, die wir hier betrachten. Da ist kein Platz für Science-Fiction, wir sprechen nicht über Maschinen mit Bewusstsein. Die KI führt sehr wohl Aufgaben aus, die normalerweise eine Art von Intelligenz erfordern würden. Die Fähigkeit zu Intuition oder Kreativität hat eine KI aber nicht – das sind Attribute, die wir mit menschlicher Intelligenz verknüpfen.

Das andere Element ist die Art der Technologie. Es ist das erste Mal in der Geschichte der Menschheit, dass wir autonome Maschinen haben. Diese großartige Technologie kann für viele Zwecke eingesetzt werden, bringt aber auch neue Herausforderungen – ethische und technische.

(+) PLUS: Auf welche Weise könnte KI zu Sicherheitsprodukten und -dienstleistungen beitragen?

Fotos: Digital Ethics

Taddeo: KI kann in vielerlei Hinsicht helfen. In Hinblick auf die Robustheit von Technik kann KI eingesetzt werden, um Fehler und Lücken in Systemen zu identifizieren. Sie kann etwa Verifikations- oder Validierungsprozesse viel schneller erledigen. Diese sind im Allgemeinen oft zeitaufwendig und mühselig.

Weiters wird KI eingesetzt, um Systeme bei der Abwehr von Angriffen zu unterstützen. Wir haben 2016 bei der von der DARPA organisierten »Cyber Grand Challenge« gesehen, wie KI-Systeme gegeneinander ausgespielt werden können. Sie erkennen Schwachstellen in den eigenen und in den konkurrierenden Systemen und können Angriffsstrategien entwickeln, um den Gegner auszuschalten.

Schließlich sehen wir bereits viele Produkte auf dem Markt, die KI nutzen. Die Technologie ist auch von Vorteil, wenn man mit bislang unbekannten Angriffsvektoren konfrontiert ist. Ändert sich plötzlich etwas im Zustand eines Systems, kann die KI einen Angriff innerhalb von Stunden – und nicht Tagen – ausmachen. Das britische Unternehmen Darktrace nutzt bereits Methoden des maschinellen Lernens, um kompromittierte Teile von Systemen, die angegriffen wurden, zu identifizieren und unter Quarantäne zu stellen.

> UNTERNEHMEN UND DER ÖFFENTLICHE SEKTOR SOLLTEN DIE LÖSUNGEN UND KOMPONENTEN FÜR MASCHINELLES LERNEN BESSER SELBST ENTWICKELN.

Das sind alles gute Nachrichten und es ist der Grund, warum es weltweit großen Druck auf die Entwicklung von KI-basierten Produkten für die Cybersicherheit gibt. Die Rolle von KI in der Cybersicherheit wird in den Leitlinien der EU-Kommission und von anderen internationalen Initiativen betont. Normierungsgremien wie die IEEE arbeiten an Standards für KI in der Cybersicherheit. All diese Initiativen haben ein Element gemeinsam: die Idee, vertrauenswürdige KI voranzutreiben.

(+) PLUS: Was sind die Herausforderungen für vertrauenswürdige KI?

Taddeo: Für die Sicherheit unserer Gesellschaft, für Infrastrukturen und für den Einzelnen brauchen wir Vertrauen in Technologie. Nun ist aber echte KI eine Blackbox. Wir wissen nicht im Detail, wie ein auf maschinellem Lernen basierender Prozess ein bestimmtes Ergebnis hervorbringt.

Vertrauen ist eine vielseitige Angelegenheit, die auf unterschiedliche Weise definiert wird. Wir können es vereinfachen: Vertrauen ist eine Form des Delegierens von Aufgaben, ohne die weitere Kontrolle darüber zu haben. Doch wir müssen das Risiko abschätzen können. Einer KI-Lösung bei einer einfachen Bilderkennung in einer Büroumgebung zu vertrauen, ist einfach, da die negativen Folgen bei einem Fehler eher gering sind. Dem gleichen System würde ich aber weniger vertrauen, wenn davon die Fahrsicherheit in einem autonomen Fahrzeug abhängig ist. Vertrauen ist also immer von der Umgebung und einem Kontext abhängig.

(+) PLUS: Welchen Bedrohungen sind KI-Systeme selbst ausgesetzt?

Taddeo: KI ist eigentlich eine sehr fragile und fehleranfällige Technologie. Das betrifft zum Beispiel die Manipulation von Daten. Studien zeigen, dass mit einem minimalen Aufwand das Ergebnis eines KI-Systems in großem Umfang verändert werden kann. In einem Fall in einem Krankenhaus wurde gezeigt, dass bei 8 % verfälschten Daten in der automatisierten Medikamentenverteilung 75 % der Patienten die falschen Dosierungen bekommen. In einer anderen Studie haben Forscher das Muster eines Schildkrötenpanzers so verändert, dass eine KI ihn fälschlicherweise als Gewehr identifiziert hat.

Schließlich besteht die Gefahr von Hintertüren in neuronalen Netzwerken: Da diese Technologien keinen einsehbaren Quellcode im klassischen Sinn haben, werden Backdoors kaum erkannt. Es gibt Fälle, in denen eine Erkennungssoftware ein Stoppschild falsch interpretiert, wenn jemand ein Post-It auf das Verkehrsschild klebt. Autonome Fahrzeuge halten dann an dieser Kreuzung nicht an. Was also, wenn jemand eine solche Hintertür einbaut, um irgendwann die Ergebnisse dieses Systems in eine gewünschte Richtung zu drehen?

KI gegenüber Manipulationen robust zu machen, ist eine Bemühung, die wir überall auf der Welt sehen. Es ist ein Teil der Standardisierungsverfahren. Die Robustheit von KI ist freilich rechnerisch ein unlösbares Problem, da die Anzahl der möglichen Zustände eines Systems astronomisch groß ist. Deshalb sind Ansätze wie jene der Europäischen Kommission, KI in der Cybersicherheit generell zu vertrauen, konzeptionell irreführend und operativ gefährlich.

Da KI eine lernende Technologie ist, führt uns die Idee des Vertrauens – Delegieren ohne Kontrolle – in die falsche Richtung. Der OECD-Grundsatz lautet: KI-Systeme müssen während ihres gesamten Lebenszyklus robust und sicher funktionieren, und potenzielle Risiken sollten kontinuierlich bewertet und behandelt werden.

(+) PLUS: Sollten wir uns besser auf herkömmliche Software und hart geschriebenen Code verlassen?

Taddeo: KI kann relativ effektiv bei der Unterstützung von Cybersicherheitsaufgaben sein. Aber wir sollten vom Terminus der vertrauensbasierten KI zu jenem zuverlässiger KI-Systeme übergehen. Das bedeutet für die Betreiber kritischer Infrastrukturen, dass bei der Beschaffung »KI als Dienstleistung« keine Option ist: Unternehmen und der öffentliche Sektor sollten Lösungen und Komponenten für maschinelles Lernen selbst entwickeln. Und wir werden Standards für das Training von KI-Systemen brauchen. Zudem benötigen wir ein dynamisches, paralleles Monitoring. Wenn es ein unterschiedliches Verhalten der beiden Systeme – das eine draußen in der Praxis und das andere im Labor – gibt, können wir entsprechend eingreifen.

(+) PLUS: Brauchen wir Zertifizierungen für die Sicherheit und Qualität von KI-Lösungen?

Taddeo: Durchaus. Aber bevor Zertifizierungen möglich sind, brauchen wir Standards. Letztere bilden eigentlich die größere Herausforderung. Standards treiben Märkte an und haben auch politischen Einfluss. KI-Lösungen, die sich an gewisse Standards halten, werden Zugang zu diesen Märkten bekommen. Wir müssen uns also entscheiden, mit welchem Standard wir arbeiten wollen und welche Werte darin verankert sind. Menschen nehmen Technologien an, wenn diese mit den Werten einer Gesellschaft in Einklang sind. Wenn das nicht der Fall ist, werden wir eine große Chance verpassen. ■

ZUR PERSON

Mariarosaria Taddeo ist Associate Professor und Senior Research Fellow am Oxford Internet Institute, University of Oxford, und stellvertretende Direktorin des Digital Ethics Lab. Sie ist zudem Faculty Fellow und Defence Science and Technology Fellow am Alan Turing Institute.