Garantiert vertrauenswürdige KI?

Was das neue Gesetz über künstliche Intelligenz der Europäischen Union leistet

Franziska Weindauer
Geboren 1989 in Lübeck, Geschäftsführerin des KI-Projektlabors TÜV AI Lab GmbH, Berlin.

Leonie Löbenberg
Junior AI Solution Architect, KI-Projektlabor TÜV AI Lab GmbH, Berlin.

Mit ihrem Gesetz über künstliche Intelligenz (*EU AI Act*) setzt die Europäische Union (EU) einen Meilenstein bei der Regulierung von Künstlicher Intelligenz (KI). Die am 1. August 2024 in Kraft getretene Verordnung legt erstmals einheitliche, horizontale Anforderungen an KI-Anwendungen fest und schafft einen klaren rechtlichen Rahmen. Im Mittelpunkt steht die Einteilung von KI-Anwendungen in Risikoklassen, die spezifische regulatorische Maßnahmen erfordern. Entscheidend ist, in welche Risikoklasse das jeweilige KI-System oder -Modell fällt.

Der *EU AI Act* reagiert auf die wachsende Bedeutung von KI in allen Lebensbereichen und den Bedarf an einheitlichen Standards, die Innovationen fördern und gleichzeitig Risiken minimieren. Er schafft erstmals EU-weite Anforderungen an KI-Anwendungen und minimiert Fragmentierungen von Regelungen innerhalb der EU. Dabei verfolgt die gesetzliche Verordnung ein am Menschen orientiertes Ziel, das den

Schutz der Grundrechte gemäß der EU-Charta sicherstellen soll. Priorität wird der Sicherheit und Gesundheit der Bürgerinnen und Bürger eingeräumt, um ein hohes Schutzniveau in allen von KI betroffenen Bereichen zu gewährleisten und Innovationen im Einklang mit europäischen Werten und rechtlichen Standards zu fördern. Wenn der *EU AI Act* gut umgesetzt wird, wird er dazu führen, dass vertrauenswürdige KI ein Alleinstellungsmerkmal der Europäischen Union wird, und damit die Wettbewerbsfähigkeit der EU stärken. Europa setzt damit ebenso wie in vielen anderen Rechtsbereichen auf das Vorsorgeprinzip. Viel kritisiert, aber kulturell passend – „better safe than sorry".

Minimales bis inakzeptables Risiko

Der *EU AI Act* teilt KI-Systeme und -Modelle in verschiedene Risikoklassen ein, um zu gewährleisten, dass nur bei besonders risikobehafteten Anwendungen anspruchsvolle Anforderungen insbesondere an die Anbieter gerichtet werden. So liegt der Großteil der KI-Systeme auf dem Markt außerhalb des Anwendungsbereichs der neuen KI-Gesetzgebung.

Die Einstufung von KI-Systemen im *EU AI Act* basiert nicht auf der zugrunde liegenden Technologie, sondern vielmehr auf dem sektorspezifischen Anwendungsbereich und den potenziellen Auswirkungen auf Einzelpersonen und die Gesellschaft. Dabei wird unterschieden zwischen *Stand-alone*-Systemen, die als eigenständige Lösungen agieren, und KI-Systemen, die als sicherheitskritische Komponenten in bestehende Produkte integriert sind. Die folgenden vier Risikoklassen sind für KI-Systeme vorgesehen:

Erstens: minimales Risiko. KI-Systeme, die in diese Kategorie fallen, unterliegen keinen spezifischen regulatorischen Anforderungen, da sie kein signifikantes Risiko für die Sicherheit, die Rechte oder das Wohlergehen der Menschen darstellen. In diese Kategorie werden die meisten der sich zurzeit am Markt befindlichen KI-Anwendungen fallen.

Zweitens: begrenztes Risiko. KI-Systeme mit begrenztem Risiko müssen Transparenzanforderungen erfüllen. Dies betrifft in erster Linie KI-Systeme, die direkt mit Menschen interagieren, wie etwa Chatbots. Solche KI-Systeme müssen klar gekennzeichnet sein, um den Nutzerinnen und Nutzern zu verdeutlichen, dass sie mit einer KI interagieren. Dies schützt deren Autonomie und ermöglicht es ihnen, fundierte Entscheidungen zu treffen. Darüber hinaus muss die synthetische Generierung von Audio-, Bild-, Video- oder Textinhalten transparent gemacht werden. Das soll Vertrauen in KI schaffen und eine informierte Nutzung ermöglichen. Zu beachten ist, dass der *EU AI Act* Ausnahmen von der Transparenzpflicht definiert, etwa für KI-Systeme im Bereich der Strafverfolgung.

Drittens: hohes Risiko. Diese Kategorie umfasst Systeme, die potenziell erhebliche Risiken für die Sicherheit, Gesundheit oder Grundrechte von Personen darstellen und folglich besonders strengen Anforderungen unterliegen. Diese KI-Systeme werden in bereits regulierte Bereiche (Anhang I des *EU AI Act*) – wie Medizinprodukte, Maschinen und Mobilität – sowie in noch nicht regulierte Bereiche (Anhang III des *EU AI Act*) – wie KI-basierte HR-Systeme (*Human Resources*-Systeme: Softwarelösungen zur Digitalisierung, Automatisierung und Rationalisierung von Personalfunktionen in einem Unternehmen), Bildung und Verwaltung – unterteilt. In den regulierten Bereichen werden die Anforderungen des *EU AI Act* in die bestehenden sektoralen Vorschriften integriert. Obwohl der Mobilitätsbereich als Hochrisikobereich eingestuft wird, ist dieser Bereich aktuell noch vom Geltungsbereich der Verordnung ausgeschlossen. Die Anforderungen des *EU AI Act* müssen erst in die sektoralen Regulierungen direkt integriert werden, wie etwa die Typengenehmigungsverordnung im Automobilbereich. In noch nicht regulierten Bereichen sind die Anforderungen des *EU AI Act* hingegen ab August 2024 zu erfüllen.

Für Hochrisiko-KI-Systeme legt der *EU AI Act* spezifische Anforderungen fest, darunter Qualitäts- und Risikomanagementsysteme, eine technische Dokumentation sowie Vorgaben in den Bereichen Daten-Governance (inklusive Datenqualität und Nichtdiskriminierung), *Human Oversight*, Genauigkeit, Robustheit und Cybersicherheit. Diese Maßnahmen sollen sicherstellen, dass die Systeme zuverlässig, sicher und transparent arbeiten. Ausnahmen bestehen für Anwendungen, die nur begrenzte Aufgaben ohne sicherheitsrelevante Implikationen erfüllen. Diese Ausnahmen tragen dazu bei, die Regulierung auf die risikoreichsten Bereiche zu konzentrieren, während der Innovationsspielraum in weniger kritischen Bereichen weitgehend ungeschmälert erhalten bleibt.

Viertens: inakzeptables Risiko. KI-Systeme, die als inakzeptabel risikobehaftet eingestuft werden, sind in Mitgliedstaaten der Europäischen Union verboten. Dies gilt insbesondere für Systeme, die die Würde des Menschen verletzen, zum Beispiel Systeme, die manipulative Techniken zur unethischen Beeinflussung des Verhaltens einsetzen oder diskriminierende Praktiken fördern. Dies verdeutlicht das Bestreben der EU, ethische Standards und den Schutz der Grundrechte zu wahren.

Die Europäische Kommission hat auf die Veröffentlichung großer KI-Modelle mit allgemeinem Verwendungszweck, wie etwa *GPT-4* – integriert in *ChatGPT* – reagiert und mit der technologie-agnostischen Ausrichtung des *EU AI Act* in seiner ursprünglich vorgeschlagenen Fassung gebrochen. So sieht der *EU AI Act* jetzt auch die Regulierung bestimmter KI-Modelle vor. Hier unterscheidet die Verordnung zwischen KI-Modellen mit und ohne allgemeinen Verwendungszweck. Ein KI-Modell gilt

dann als KI-Modell mit allgemeinem Verwendungszweck, wenn es ein breites Spektrum unterschiedlicher Aufgaben kompetent erfüllen kann. Für KI-Modelle ohne allgemeinen Verwendungszweck werden keine Anforderungen definiert. KI-Modelle mit allgemeinem Verwendungszweck – sogenannte *General Purpose AI* (GPAI) – werden aufgrund ihrer Vielseitigkeit und potenziellen Risiken im Gegensatz zu KI-Systemen unabhängig vom Anwendungsbereich reguliert. Der *EU AI Act* sieht für GPAI zwei spezifische Risikoklassen vor: Systeme mit systemischem und ohne systemisches Risiko. Diese Klassifizierung stellt sicher, dass besonders leistungsfähige und potenziell risikoreiche KI-Modelle strenger überwacht und reguliert werden, um unerwünschte Auswirkungen auf die Gesellschaft zu minimieren.

Darüber hinaus definiert der *EU AI Act* verschiedene Rollen – das heißt unterschiedliche Verantwortlichkeiten und Funktionen, die Akteure im Zusammenhang mit KI-Systemen übernehmen –, die über den Umfang der konkreten Anforderungen je Risikoklasse bestimmen. So sind etwa Anforderungen für Anbieter von KI-Modellen und -Systemen besonders hoch. Weitere Rollen sind beispielsweise Betreiber oder Händler von KI-Systemen.

Es sind bereits Anwendungen zur Bestimmung der Risikoklasse auf dem Markt. Wer unsicher ist, in welche Risikoklasse das eigene KI-Modell oder -System fällt oder welche Rolle nach der KI-Verordnung erfüllt wird, kann dies mithilfe des *AI Act Risk Navigators* des *TÜV AI.Lab* herausfinden. Der *AI Act Risk Navigator* ist ein kostenloses, niedrigschwelliges Klassifikationstool für KI-Anwendungen im Single- und Multiple-Choice-Verfahren (*www.tuev-risk-navigator.ai*).

Stärken der Gesetzgebung

Eine der größten Stärken des Gesetzes ist die Schaffung einheitlicher Regeln innerhalb der Europäischen Union. Im Gegensatz zu den USA, wo die Gesetzgebung oftmals auf Ebene der Bundesstaaten erfolgt und sich Unternehmen daher an unterschiedliche Regelungen in verschiedenen Regionen anpassen müssen, bietet die EU einen einheitlichen Rechtsrahmen. Dies sorgt nicht nur für rechtliche Klarheit, sondern verringert auch den administrativen Aufwand und die Komplexität für Unternehmen, die ihre KI-Produkte in mehreren Mitgliedstaaten anbieten.

Ein weiterer Vorteil liegt in der Stärkung des Vertrauens in KI-Anwendungen. Durch die Festlegung von Mindestanforderungen in Hochrisikobereichen, etwa in Bezug auf Sicherheit, Robustheit und Datenqualität, können Verbraucherinnen und Verbraucher sicher sein, dass sie Produkte nutzen, die bestimmten Qualitätsstandards entsprechen. Dieses Vertrauen ist unerlässlich für eine breite Akzeptanz von KI

in der Gesellschaft und für den langfristigen Erfolg von KI-basierten Produkten und Dienstleistungen.

Die Verordnung schafft auch Klarheit über die Anforderungen, die Unternehmen erfüllen müssen, bevor sie ein Produkt auf den Markt bringen. Diese Orientierung ermöglicht es Unternehmen, frühzeitig die Konformität ihrer Produkte sicherzustellen. Diese Vorhersehbarkeit reduziert Unsicherheiten und erleichtert es, sich strategisch auf die regulatorischen Anforderungen vorzubereiten.

Ein wesentlicher Fortschritt im Vergleich zu früheren regulatorischen Ansätzen, wie etwa der Datenschutz-Grundverordnung (DSGVO), ist die Einrichtung eines eigenen *AI Office* auf EU-Ebene. Dieses Büro soll nicht nur Leitfäden und Unterstützungsdokumente bereitstellen, sondern hat bereits vor Inkrafttreten des *EU AI Act* Normungsaufträge an das Europäische Komitee für elektrotechnische Normung (*Comité Européen de Normalisation Électrotechnique*, CENELEC) erteilt. Dies hilft den Unternehmen, die KI-Verordnung effektiver und effizienter umzusetzen.

Schwachstellen bei der Umsetzung

Trotz dieser Stärken besitzt der *EU AI Act* auch einige (potenzielle) Schwachstellen. Eine davon ist die nationale Umsetzung der Verordnung. Es bleibt abzuwarten, wie einheitlich diese in der Europäischen Union erfolgen wird. In Deutschland beispielsweise könnte die gründliche und oft überkomplexe Umsetzung von EU-Vorgaben zu Problemen führen, besonders vor dem Hintergrund der föderalen Strukturen. Hier könnte der „Vorteil" der Einheitlichkeit in der Praxis geschwächt werden. Es ist fraglich, inwieweit die EU-Mitgliedstaaten tatsächlich an einem Strang ziehen und sich intensiv austauschen werden, um eine konsistente Umsetzung zu gewährleisten.

Ein gut ausgestattetes Digitalministerium inklusive einer Digitalagentur auf Bundesebene könnte hier Abhilfe schaffen und dazu beitragen, dass die notwendigen Maßnahmen ergriffen werden, um EU-Gesetze im Digitalbereich schlanker umzusetzen.

Ein weiterer Kritikpunkt ist die „Unreife" der Regulierung. Künstliche Intelligenz ist ein noch junges Technologiefeld, und viele Aspekte ihrer Funktionsweise und potenzieller Risiken sind noch nicht vollständig erforscht. Dies führt dazu, dass sowohl Unternehmen als auch Regulierungsbehörden mit der Umsetzung der Vorgaben des *EU AI Act* überfordert sein werden beziehungsweise es bereits sind. Das liegt vor allem daran, dass große Unsicherheit darüber besteht, wie die Verordnung auszulegen ist; unklar ist, wie KI-Modelle und -Systeme nach dem *EU AI Act* skalierbar sind und effektiv geprüft beziehungsweise

bewertet werden können. Zudem führt die Umsetzung einer so komplexen und aktuell oft noch unklaren Regulierung zu einer Mehrbelastung für Unternehmen – weniger für Mitarbeiter in der Entwicklung, mehr in der Compliance-Abteilung. Diese Nachteile gilt es allerdings mit den Vorteilen abzuwägen, die die Regulierung wie oben dargestellt bietet.

Angesichts dieser Herausforderungen ist es entscheidend, dass die Umsetzung des *EU AI Act* durch das *AI Office* und andere Akteure des Ökosystems eng begleitet wird. Dazu gehören die Erstellung praxisnaher Umsetzungshilfen, die Zusammenarbeit aller relevanten Stakeholder und die Förderung von Forschungsprojekten, die sich mit den praktischen Aspekten der Regulierung befassen. Auch die Beteiligung an Normungsgremien und die Schaffung von Kapazitäten für die Normenentwicklung sind von zentraler Bedeutung, um die Vorgaben des *EU AI Act* zu präzisieren und handhabbar zu machen.

Langfristig wird es wichtig sein, die Regulierung dynamisch an den technologischen Fortschritt anzupassen. Dies bedeutet, dass die Verordnung technologie-agnostisch gestaltet sein muss, um nicht von der technologischen Entwicklung überholt zu werden. Gleichzeitig muss die Regulierung sicherstellen, dass sie Innovationen nicht hemmt, sondern den Fortschritt unterstützt, indem sie Flexibilität bietet und Raum für neue Entwicklungen lässt.

Ein interessantes Konzept, das in diesem Zusammenhang immer wieder diskutiert wird, sind Reallabore. Sie können dazu beitragen, neue Technologien in einem regulierten Umfeld zu testen und die Regulierung auf Basis der gewonnenen Erkenntnisse weiterzuentwickeln. Es ist jedoch unklar, ob dieses Konzept in der Praxis so funktioniert, wie es in der Theorie angedacht ist.

Der *AI Act* der Europäischen Union ist ein wichtiger Schritt hin zu einer einheitlichen und umfassenden Regulierung von KI. Die Verordnung schafft Klarheit, stärkt das Vertrauen in KI und bietet Unternehmen eine verlässliche Grundlage für die Entwicklung und den Einsatz von KI-Technologien. Gleichzeitig zeigt sich, dass die praktische Umsetzung der Verordnung noch einige Herausforderungen mit sich bringen wird, insbesondere im Hinblick auf die nationale Umsetzung und den Umgang mit der noch unausgereiften Regulierung.

Die Weiterentwicklung der KI-Verordnung wird entscheidend dafür sein, ob Europa seine Ziele in der KI-Politik erreicht. Hierbei wird es wichtig sein, dass alle Akteure zusammenarbeiten, um eine effektive und flexible Regulierung zu schaffen, die sowohl den Schutz der Bürgerinnen und Bürger als auch die Förderung von Innovationen gewährleistet.