ExamAI

Arbeitspapier

KI-Regulierung made in Europe: Positionen zum Gesetzentwurf der Europäischen Kommission

Dezember 2021

Arbeitspapier

KI-Regulierung made in Europe: Positionen zum Gesetzentwurf der Europäischen Kommission

Dezember 2021

INHALTSVERZEICHNIS

Editorial

Ende April 2021 hat die EU-Kommission den weltweit ersten Rechtsrahmen speziell für Künstliche Intelligenz veröffentlicht. [1] Der Regulierungsvorschlag (AI Act) beinhaltet eine Risikoklassifizierung von KI-Anwendungen und formuliert – abhängig von der jeweiligen Risikoklasse – unterschiedliche Anforderungen an Unternehmen. Für sogenannte Hochrisiko-Anwendungen wurden beispielsweise spezielle Anforderungen formuliert, worunter neben allgemeinen Dokumentations- und Transparenzpflichten auch das Testen entsprechender KI-Systeme fällt. Die Anwendungsfälle aus dem Projekt ExamAI (Industrieproduktion [2] & Personal- und Talentmanagement [3]) sind größtenteils Teil der Kategorie „Hochrisiko". Dementsprechend stellt sich die Frage, welche konkreten Auswirkungen der EU-Regulierungsvorschlag in seiner aktuellen Form für die Anwendung von KI im Arbeitsbereich hat.

In unserem WebTalk „KI-Regulierung made in Europe" am 14. September 2021 diskutierten vier Expert*innen aus Wirtschaft, Wissenschaft und Politik über diese Auswirkungen sowie über die Stärken und Schwächen des AI Acts in seiner aktuellen Form. Die Diskussionsrunde wurde aufgezeichnet und steht allen Interessierten zur Verfügung. [4]
Dieses Arbeitspapier stellt eine Zusammenfassung der vielfältigen Perspektiven unserer Diskutant*innen dar. **Dr. Johanna Wenckebach** (Hugo-Sinzheimer Institut, Hans-Böckler-Stiftung) spricht sich in ihrem Beitrag dafür aus, einen stärkeren Arbeitnehmer*innenschutz im AI Act zu integrieren, damit dieser nicht als zahnloser Tiger daherkommt. Wenn der AI Act eine starke und flexible Grundlage für die Gestaltung verantwortungsvoller KI darstellen soll, müssen zudem die Verantwortlichkeiten für KI-Systeme und die Ausgestaltung der im AI Act formulierten Pflichten geklärt werden, fordert **Cornelia Kutterer** (Microsoft). Stellvertretend für das Projekt ExamAI bemängelt **Nikolas Becker** die fehlenden Prüfmethoden für Hochrisiko-KI-Systeme und schlägt Assurance Cases und ATDD als Grundstein für das Testen von KI-Systeme im Arbeitsbereich auf Sicherheit und Fairness vor.

Das Team des Projekts ExamAI wünscht Ihnen viel Freude und spannende Einsichten beim Lesen.

Daniel Krupka, Nikolas Becker, Julia Meisner

[1]

Europäische Kommission (2021): Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierender Vorschriften für Künstliche Intelligenz und zur Änderung bestimmter Rechtsakte der Union

[2]

Adler, R., Heidrich, J., Jöckel, L., Kläs, M. (2020). Anwendungsszenarien: KI-Systeme in der Produktionsautomatisierung, ExamAI – KI Testing & Auditing. Berlin: Gesellschaft für Informatik e.V.

[3]

Zweig, K., Hauer, M., Raudonat, F. (2020). Anwendungsszenarien: KI-Systeme im Personal- und Talentmanagement, ExamAI – KI Testing & Auditing. Berlin: Gesellschaft für Informatik e.V.

[4]

https://testing-ai.gi.de/meldung/diskriminiert-durch-ki-gi-webtalk-diskutiert-rechtliche-aspekte-2

Anforderungen an den AI Act mit Blick auf gute Arbeit und die Rechte von Arbeitnehmer*innen

Von Dr. Johanna Wenckebach, Hugo-Sinzheimer Institut der Hans-Böckler-Stiftung

Zurecht wird eine intensive rechtspolitische Debatte um den AI Act geführt. Dabei müssen die Besonderheiten der Arbeitswelt fokussiert werden. Der gebotene Schutz von Beschäftigtenrechten muss effektiv ausgestaltet werden. Hierfür sind Transparenz und Mitbestimmung essenziell.

Die Arbeitswelt ist geprägt von starken Interessengegensätzen und einem großen Machtungleichgewicht, das Arbeitsverträge von anderen Vertragsverhältnissen unterscheidet. Das betrifft auch den rechtlichen Rahmen abhängiger Arbeit: Arbeitsrecht ist im Wesentlichen Arbeitnehmerschutzrecht.

Regulierungsbedarf von KI insbesondere in der Arbeitswelt

Ob Arbeitsrecht überreguliert ist, wird erfahrungsgemäß von den jeweiligen Seiten des Arbeitsvertrags unterschiedlich beurteilt. Einigkeit besteht allerdings darüber, dass die Veränderungen der Arbeitswelt durch die Digitalisierung so massiv sind, dass der Gesetzgeber gefordert ist. Das betrifft vor allem auch die technischen Möglichkeiten, die durch algorithmische Systeme – in der politischen Debatte als „künstliche Intelligenz“ (KI) bezeichnet – entstehen.

Unterschiedlich wiederum sind die Auffassungen dazu, welcher Regelungsbedarf in der Arbeitswelt durch KI-Systeme entsteht: Die Arbeitgeberseite wünscht sich vor allem fairen Wettbewerb, den Schutz von Eigentumsrechten und Unternehmensdaten sowie Rechtssicherheit in Haftungsfragen. Aus Sicht der Beschäftigten wiederum ist der Schutz ihrer fundamentalen Rechte relevant, die durch KI in vielerlei Hinsicht betroffen werden. Es geht um die physische und psychische Gesundheit und die wirtschaftliche Absicherung von Arbeitnehmenden, um Mitbestimmung und demokratische Teilhabe, sowie um den Schutz von Beschäftigtendaten und somit um Persönlichkeitsrechte.

Für Letztere ist auch das Antidiskriminierungsrecht wichtig. Regelungslücken gehen zulasten der Rechte Beschäftigter. Sie sind deshalb auf eine effektive, das heißt auch gut durchsetzbare Regulierung angewiesen.

Effektive Beschäftigtenrechte

Da Arbeitnehmer*innen in der Regel rechtliche Auseinandersetzungen mit ihrem Arbeitgeber scheuen, sind präventive Regeln besonders sinnvoll. Das Arbeitsrecht setzt dabei zurecht auf kollektive Handlungsmöglichkeiten: Gewerkschaften und betriebliche Mitbestimmung. So werden Einzelne vor individuellen Konflikten im Arbeitsverhältnis geschützt und das beschriebene Machtgefälle der Vertragsparteien ausgeglichen. Aber auch individuelle Abwehrrechte zur Beseitigung von Rechtsverstößen sind essenziell. Wenn sie mit ausreichenden Sanktionen belegt sind, entfalten auch diese Rechte abschreckende Wirkung.

Anforderungen an den AI Act

Vor diesem Hintergrund ist es – erstens – sehr zu begrüßen, dass die Europäische Kommission einen so detaillierten Regelungsentwurf vorgelegt hat. Das war dringend geboten, auch um Arbeitnehmer*innenrechte angesichts rasanter technischer Entwicklungen zu schützen. Zweitens allerdings wird der Entwurf den skizzierten Besonderheiten der Arbeitswelt und des Arbeitsrechts noch nicht gerecht.

Auch wenn eine umfangreiche arbeitsrechtliche Analyse des Entwurfs an dieser Stelle nicht erfolgen kann, so lassen sich doch auch in aller Kürze wesentliche Defizite benennen:

Tarifautonomie und Mitbestimmung. Blinder Fleck?

Bisher entsteht der Eindruck, dass der – als Verordnung in allen Mitgliedstaaten gleichermaßen geltende – AI Act abschließend sein soll. Das wirft die Frage nach weitergehender Regulierung auf, sei es durch nationale Gesetzgebung oder durch kollektive Regeln. Hierbei ist vor allem an Tarifverträge zu denken. Sie können branchenspezifischen Wirtschafts- aber auch Arbeitnehmerinteressen besser Rechnung tragen. Im deutschen Recht ist die Tarifautonomie verfassungsrechtlich garantiert. Die Mitbestimmungsrechte des Betriebsverfassungsgesetzes wiederum ermöglichen auf

Betriebe zugeschnittene Lösungen und schützen Beschäftigte vor Vereinzelung. Durch Mitbestimmung wird Teilhabe ermöglicht, was die Akzeptanz von Transformationsprozessen erhöht.

In der Datenschutzgrundverordnung (DSGVO) wurde deshalb für den Beschäftigungskontext in Art.88 DSGVO explizit klargestellt, dass durch Rechtsvorschriften oder Kollektivvereinbarungen spezifischere Vorschriften zur Gewährleistung des Schutzes der betroffenen Rechte und Freiheiten möglich sind. Eine entsprechende Regelung ist auch für den AI Act zu empfehlen, der bisher einen blinden Fleck aufweist. Zudem muss sichergestellt werden, dass betriebliche und gewerkschaftliche Interessenvertretungen Zugang zu relevanten Informationen haben. Auch kollektive Akteure sind machtlos, wenn KI-Anwendungen intransparent sind.

Arbeit als Hochrisiko-Bereich

Entscheidend für die Sicherung von Beschäftigtenrechten ist eine gänzliche Erfassung abhängiger Beschäftigungsverhältnisse im „Hochrisiko"-Bereich. Annex III des Entwurfs bedarf deshalb einer weiten Fassung. Die – nicht offizielle – deutsche Übersetzung weist gegenüber dem englischsprachigen Original Einschränkungen auf, die der Korrektur bedürfen.

Rechtsansprüche Betroffener fehlen. Zahnloser Tiger?

Aus anderen Rechtsgebieten – etwa der Regulierung von Lieferketten oder dem Antidiskriminierungsrecht – gibt es ausreichend Untersuchungen dazu, dass eine Selbstkontrolle von Unternehmen allein nicht ausreicht, um die Rechte Betroffener zu schützen. Die Konformitätsbewertung, wie Art. 43 des Entwurfs sie bisher vorsieht, ist deshalb für die Arbeitswelt aus Beschäftigtensicht völlig unzureichend.

Nach dem oben Gesagten sind Rechtsansprüche Betroffener unverzichtbar. Jedoch fehlen sie im Entwurf bisher. Auch hier kann die DSGVO mit ihren Auskunftsansprüchen und Beschwerderechten als Vorbild dienen.

Abschließend sei darauf hingewiesen, dass bereits jetzt die Umsetzung der Verordnung in den Blick genommen werden sollte. Die im AI-Act vorgesehen Strukturen, etwa „Marktüberwachungsbehörden", die für den Rechtsschutz gemäß dem Entwurf

eine sehr wichtige Stellung einnehmen, müssen mit ausreichenden Kompetenzen aber auch finanziellen und personellen Ressourcen ausgestattet werden.

Der AI Act hilft Beschäftigten nicht, wenn er als zahnloser Tiger daherkommt.

Über die Autorin:

Dr. Johanna Wenckebach leitet seit Juli 2019 als wissenschaftliche Direktorin das Hugo Sinzheimer Institut für Arbeits- und Sozialrecht (HSI) der Hans-Böckler-Stiftung. Daneben lehrt sie Themen des kollektiven Arbeitsrechts an der Europäischen Akademie der Arbeit in Frankfurt am Main und an der EBS Universität Wiesbaden. Sie ist ehrenamtliche Richterin am Arbeitsgericht Berlin. Vor ihrer Tätigkeit am HSI war sie mehrere Jahre Tarifjuristin in der IG Metall Bezirksleitung Berlin-Brandenburg-Sachsen und Aufsichtsrätin. Davor absolvierte sie ihr Referendariat in Berlin und Potsdam und war wissenschaftliche Mitarbeiterin an verschiedenen Lehrstühlen im In- und Ausland. Sie promovierte 2012 zu einem antidiskriminierungsrechtlichen Thema.

Ein Balanceakt: Regulierung von KI zur Förderung verantwortungsvoller Innovation in Europa

Von Cornelia Kutterer, Microsoft

Die Entwicklung und Nutzung von KI-Systemen wirft wichtige und komplexe ethische, rechtliche, politische und wirtschaftliche Fragen auf. Der “AI Act” muss einen Rahmen für die Nutzung verantwortungsvoller KI schaffen, der Grundrechte respektiert und mit den europäischen Werten in Einklang steht.

Künstliche Intelligenz (KI) ist ein äußerst komplexer Technologiebereich, der sich auf jeden Aspekt unserer Gesellschaft auswirkt. Der Einsatz von KI kann in zahlreichen Bereichen wie Gesundheitswesen, Verkehr, Klima und öffentliche Verwaltung erhebliche Vorteile bringen und der gesamten Menschheit vielversprechende Möglichkeiten eröffnen.

Die Begeisterung für die Leistungsfähigkeit und das Potenzial von KI darf uns jedoch nicht blind machen. Die Entwicklung und Nutzung von KI-Systemen wirft wichtige und komplexe ethische, rechtliche, politische und wirtschaftliche Fragen auf. Wir sollten uns daher der Risiken in Bezug auf Datenschutz, Überwachung, Diskriminierung oder vorsätzlichen Missbrauch bewusst sein. Das gilt insbesondere in Bezug auf Eingriffe in die Menschenrechte, die Demokratie und die Rechtsstaatlichkeit - die Kernelemente, auf denen unsere europäischen Gesellschaften aufgebaut sind.

Notwendigkeit verantwortungsvoller KI Standards und Normen

Die letzten Jahre haben gezeigt, dass die genannten Risiken real sind. In Frankreich wurden Kameras, die Personen ohne Gesichtsmaske erfassten, deaktiviert, nachdem die CNIL (die französische Datenschutzbehörde) sie für illegal erklärt hatte. In den Niederlanden entschied ein Gericht gegen SyRI, ein System zur automatischen Erkennung von Sozialhilfebetrug.

Diese Beispiele zeigen, dass das Vertrauen in die Technologie jeden Tag neu verdient werden muss, und zwar durch die großen und kleinen Entscheidungen, die wir darüber treffen, welche Systeme wir bauen und wie wir sie einsetzen. Aus dem einfachen Grund, dass die Menschen keine Technologie nutzen werden, der sie nicht vertrauen, ist die schwierige Aufgabe, die europäischen Grundsätze für KI in die Praxis umzusetzen, noch nie so dringlich wie heute. Verantwortungsvolle KI Standards und Normen sind eine notwendige Grundvoraussetzung, um Risiken zu adressieren.

Regulierung von KI

Die von der Kommission vorgeschlagene Verordnung für harmonisierte Regeln für KI (KI-Gesetz) und der neue koordinierte Plan für KI sind ehrgeizige und wichtige Schritte, um vertrauenswürdige KI zur Norm in Europa und der ganzen Welt zu machen. Ein gesetzlicher Rahmen ist notwendig, damit das enorme Potenzial von KI auf eine Weise genutzt werden kann, die die Grundrechte respektiert und mit den europäischen Werten in Einklang steht.

Das KI-Gesetz sollte die Ziele der regulierten Akteur*innen definieren, jedoch nicht diktieren wie diese Ziele erreicht werden sollen. Vor allem zwei Bereiche würden von Klarstellungen profitieren: (1) die Verantwortlichkeit für KI-Systeme und (2) ergebnis- und prozessorientierte Pflichten.

Verantwortlichkeit für KI-Systeme

Erstens sollten die rechtlichen Verpflichtungen die Rollen der verschiedenen Akteur*innen in der KI-Wertschöpfungskette widerspiegeln.

Die Nutzerin oder der Anwender von KI-Systemen wird oftmals eine wichtige Rolle bei der Entscheidung spielen, wie das System eingesetzt wird. Insbesondere KI-Systeme, die in dienstleistungsbezogenen Anwendungen eingesetzt werden, können Risiken für die Grundrechte mit sich bringen. Die Abschwächung der Risiken von KI erfordert zwar die Zusammenarbeit zwischen den verschiedenen Akteur*innen der KI-Wertschöpfungskette, doch ist der Anwender des KI-Systems dem gesellschaftlichen Kontext am nächsten.
Die Verankerung des KI-Gesetzes im Neuen Rechtsrahmen (New Legislative Framework, NLF) führt zu Herausforderungen bei der Zuweisung von Pflichten innerhalb der

KI-Wertschöpfungskette. Der Neue Rechtsrahmen sieht Kontrollen vor dem Inverkehrbringen vor, um die Gesundheits- und Sicherheitsrisiken von Produkten anzugehen. Diese Zuweisung von Zuständigkeiten kann bei Produktsicherheitsrisiken für physische Produkte sinnvoll sein, da diese Risiken in allen Einsatzszenarien relativ einheitlich sind und sich nur selten ändern, sobald ein Produkt auf den Markt kommt. Im Gegensatz dazu ist eine ähnliche Aufteilung der Zuständigkeiten für KI-Systeme nicht gut geeignet, da viele KI-Modelle zwar universell einsetzbar sind, sich die Grundrechtsrisiken jedoch abhängig vom Nutzungskontext erheblich unterscheiden können.

Um Rechtssicherheit und Stabilität als wesentliche Voraussetzungen für Innovationen besser zu gewährleisten, könnte es ratsam sein, die Verantwortlichkeit insbesondere für risikoreiche KI-Systeme entsprechend anzupassen.

Ergebnis- und prozessorientierte Pflichten

Zweitens sollten die Anforderungen des KI-Gesetzes sehr viel ergebnis- und prozessorientierter gestaltet werden. Im derzeitigen Entwurf sind viele der Pflichten des KI-Gesetzes präskriptiv oder auf bestimmte Szenarien ausgerichtet. Infolgedessen sind sie wahrscheinlich in einigen Fällen praktikabel, in anderen jedoch unwirksam.

Ein effektiverer Ansatz wäre es, wenn das KI-Gesetz die Ergebnisse, die die Akteur*innen anstreben sollten, klar formulieren würde, zusammen mit einer Liste von Schlüsselprozessen, die sie anwenden müssen, um diese zu erreichen. Beispielweise könnte eine ergebnisorientierte Verpflichtung wie folgt beschrieben werden: KI-Systeme sollten eine ähnliche Dienstleistungsqualität für alle relevanten demografischen Gruppen, die von dem System betroffen sind, bieten.

Dies würde dazu beitragen, dass die regulierten Akteur*innen Risiken effektiv über die gesamte Bandbreite der verschiedenen KI-Systeme und Anwendungsfälle identifizieren und abmildern können. Es schafft auch Raum für Innovationen, die sich aus künftigen technologischen Entwicklungen und Fortschritten in der Praxis der verantwortungsvollen KI ergeben.

Schlussfolgerung

Politische Entscheidungsträger*innen und Regulierungsbehörden spielen derzeit eine kritische Rolle bei der Förderung eines Ökosystems vertrauenswürdiger KI. Das KI-Gesetz ist eine Chance, um eine starke und flexible Grundlage für die Nutzung verantwortungsvoller KI zu definieren, die die europäischen Bürger*innen in den Mittelpunkt der KI-Systeme stellt.

Über die Autorin:

Cornelia leitet Microsofts europäisches Team für Rechtsstaatlichkeit, verantwortungsvolle Technologie und Wettbewerb, das sich mit fairem Wettbewerb, den Auswirkungen neuer Technologien auf die Gesellschaft und rechtlichen Rahmenbedingungen befasst, die den Erwartungen der Gesellschaft entsprechen. Ihr Team befasst sich mit Themen wie verantwortungsvolle KI, digitale Sicherheit und Regulierung von Inhalten, Datenschutz, rechtmäßiger Zugang zu Daten für Strafverfolgungsbehörden, Menschenrechte und Wettbewerb. In ihrer Funktion arbeitet sie Hand in Hand mit internen Teams wie Microsofts Office of Responsible AI, Microsoft Research, Anwält*innen und Regulierungsteams. Sie tauscht sich regelmäßig mit führenden europäischen Wissenschaftler*innen in diesen Bereichen aus, um die akademische Denkweise voranzutreiben. Cornelia verfügt über langjährige Erfahrung in der Informationsgesellschaft und Internetpolitik und hält regelmäßig Vorträge auf regionalen und internationalen Konferenzen. Bevor sie zu Microsoft kam, leitete sie die Rechtsabteilung von BEUC, der Europäischen Verbraucherorganisation. Außerdem sammelte sie Erfahrungen in einer Top-10-Kanzlei und begann ihre berufliche Laufbahn im Europäischen Parlament als politische Beraterin einer Europaabgeordneten. Cornelia ist eine zugelassene deutsche Juristin und hat einen Master-Abschluss in Informationstechnologie und Telekommunikationsrecht. Sie studierte Rechtswissenschaften an den Universitäten von Passau, Porto, Hamburg und Glasgow/Strathclyde.

Fehlenden Prüfmethoden für Hochrisiko-KI begegnen – mit Assurance Cases und ATDD zu fairen und sicheren KI-Systemen

Von Nikolas Becker

Hochrisiko KI-Systeme müssen getestet werden

Mit dem Versprechen, Effizienz zu steigern, Innovation zu fördern und die Wettbewerbsfähigkeit insgesamt zu steigern, halten KI-basierte Systeme zunehmend in allen erdenklichen Anwendungsfeldern Einzug. Ihr Einsatz birgt jedoch nicht nur Potenziale, sondern kann auch zu einer Verletzung der Grundrechte führen, indem beispielsweise selbstfahrende Fahrzeuge im Automobil- oder Industriebereich den Schutz vor körperlicher Unversehrtheit gefährden [5], Gesichtserkennungssoftware oder algorithmische Entscheidungssysteme, die an der Entscheidungsfindung im Personal- oder Kreditwesen beteiligt sind, zu unzulässiger Diskriminierung führen. [6]

Die EU-Kommission stellte im April diesen Jahres gleich zwei Gesetzentwürfe vor, die diese Risiken adressieren: Erstens den Vorschlag für eine Verordnung des Europäischen Parlaments und Rates über Maschinenprodukte, die als EU-Maschinenverordnung die zuvor gültige Maschinenrichtlinie ersetzen soll. Dieser beinhaltet als wesentliche Neuerung auch KI-Systeme und klassifiziert sie als Hochrisiko-Maschinenprodukte (Anhang I der Verordnung). [7] Zweitens veröffentlichte die EU-Kommission einen Vorschlag zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (KI-Verordnung), der KI-basierte Systeme in Anwendungen mit annehmbaren Risiken, hohem Risiko, geringem Risiko und unannehmbaren Risiken gruppiert. Während für letztere lediglich bestimmte Transparenzpflichten vorgesehen sind, sollen Anwendungen mit annehmbaren Risiken, die etwa dem Social Scoring dienen oder menschliches Verhalten manipulieren, komplett verboten werden. Für Hochrisiko-KI – Anhang III der KI-Verordnung definiert hierfür beispielsweise den Einsatz von KI-Anwendungen in vielen Bereichen des Personalwesens und der Strafverfolgung – sieht die Verordnung in Art. 9 bestimmte Risikobewertungen und Testpflichten vor. [8]

[5]
Beining, Leonie (2021): KI in der Industrie absichern und prüfen. Was leisten Assurance Cases? Abgerufen am 17.11.2021

[6]
Dastin, Jeffrey (11.10.2018): Amazon scraps secret AI recruiting tool that showed bias against women. Abgerufen am 22.09.2021

[7]
Europäische Kommission (2021a): Vorschlag für eine Verordnung des Europäischen Parlaments und Rates über Maschinenprodukte, COM(2021) 202 final. Abgerufen am 20.11.2021

[8]
Europäische Kommission (2021b): Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierender Vorschriften für Künstliche Intelligenz und zur Änderung bestimmter Rechtsakte der Union, COM(2021) 206 final. Abgerufen am 22.09.2021

Es fehlt an Prüfmethoden

Obschon die KI-Verordnung nach Art. 9 Abs. 5 Anforderungen wie die Ermittlung geeigneter Risikomanagementmaßnahmen zur Sicherstellung eines bestimmungsgemäßen Funktionierens eines KI-Systems als Testziel festlegt, bringt die Forderung, KI-Systeme zu testen, große Herausforderungen mit sich: Im Gegensatz zu klassischer Software gibt es für KI-Systeme weder geeignete harmonisierte Normen, noch etablierten Prüfmethoden. Auch der Entwurf macht keine konkreten Angaben darüber, welche Anforderungen wie und nach welchen Kriterien getestet werden sollen. Solange es keine etablierten Testverfahren und entsprechenden Rahmenbedingungen gibt, können KI-basierte Anwendungen jedoch weder breit eingesetzt, noch weiterentwickelt und verbessert werden, da es an Wissen fehlt, wie z. B. Anforderungen an ein sicheres Verhalten technisch implementiert und überprüft werden können. Entwickler*innen, Anwender*innen und Prüforganisationen benötigen folglich einen Grundstein zum Aufbau von Wissen und Erfahrungen für die Entwicklung generalisierbarer Regeln und anwendungsspezifischer Normen, die eine Konformitätsbewertung schließlich ermöglichen.

ATTD und Assurance Cases als Lösung?

Ein solcher Grundstein könnte die aus dem Safety Engineering und insbesondere in den Bereichen Automotive und Luftfahrt etablierte Methode der *Assurance Cases* sein. Diese sollen als klar strukturierte Argumentation umfassend begründen, warum ein KI-System bestimmte Anforderungen in einem klar definierten Anwendungsfall erfüllt und dass die gewählten Testkriterien im jeweiligen Anwendungskontext tatsächlich geeignet sind, um das bestimmungsgemäße Verhalten eines KI-Systems zu belegen. Dabei wird eine Aussage – beispielsweise, dass ein von einem bestimmten Unternehmen für den Transport eines definierten Gutes genutzte autonome Transportsystem sicher ist – mit einer Argumentation verbunden, die alle Annahmen zum situativen Systemverhalten enthält und sich auf Evidenzen – etwa Test- und Simulationsergebnisse, aber auch die Qualifikation der Entwickler*innen – stützt. [9] Durch seine klare Strukturierung bietet ein Assurance Case den Vorteil, dass er sich auch von externen Personen leicht nachvollziehen lässt und sich somit dafür anbietet, in Rechtsstreitigkeiten herangezogen zu werden oder dem Informationsaustausch zwischen Entwickler*innen, Anbieter*innen und Nutzer*innen zu dienen.

Geht es um Anwendungsbereiche mit Testkriterien, die sich nicht eindeutig definieren lassen – beispielsweise Fairness in der Personalauswahl – so könnte der Aufstellung

[9]

Hauer, Marc P.; Adler, Rasmus; Zweig, Katharina (2021): Assuring Fairness of Algorithmic Decision Making. *2021 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW).* Abgerufen am 22.09.2021

eines Assurance Case die aus der agilen Softwareentwicklung stammende Methode der Akzeptanztestgetriebenen Entwicklung (*Acceptance Test-Driven Development – ATDD*) vorangestellt werden. [10] Hierbei legt ein interdisziplinäres Team aus Anwendenden, KI-Expert*innen, Ethiker*innen, Rechtswissenschaftler*innen und anderen Stakeholdern zunächst fest, welche Akzeptanzkriterien und Akzeptanztests für ein konkretes fairnesskritisches KI-System relevant sind. Die Akzeptanztests sollen die Akzeptanz bei den Kund*innen bzw. den Anwender*innen sicherstellen und tragen zur Klarheit der Anforderungen bei. Da die Tests auch für Nicht-Entwickler*innen lesbar sein müssen, kann der Begriff „Fairness" aus verschiedenen nicht-technischen Perspektiven geschärft werden, also auch von Personen, die von der algorithmischen Entscheidung betroffen sind. Ob und warum die Kriterien geeignet sind, um das bestimmungsgemäße Verhalten eines KI-Systems zu prüfen, legt abschließend ein Assurance Case dar. [11]

Keineswegs stellt der Ansatz dabei den Anspruch, das bestimmungsgemäße Verhalten eines Systems grundsätzlich sicherzustellen. Vielmehr soll die genaue Untersuchung verschiedener möglicher Anwendungsfälle, davon abgeleitete kontextspezifische Testkriterien sowie die ganzheitliche Betrachtung der KI-Komponente eine fallspezifische, evidenzbasierte Argumentation fördern, die das sichere Verhalten eines algorithmischen Systems im betrachteten Einzelfall belegt. Die Beteiligung möglichst vieler unterschiedlicher Stakeholder sowohl beim ATDD als auch bei der Aufstellung eines Assurance Cases bietet darüber hinaus den Vorteil, das Bewusstsein für die Relevanz und Methodik der KI-Prüfung bei verschiedenen Interessengruppen zu stärken, Wissenstransfer zu ermöglichen und gemeinsame Abstimmungsprozesse hinsichtlich der Entwicklung und Prüfung sicherer KI-Systeme zu verbessern.

Gleichzeitig gibt es noch keine Richtlinien, wie ATDD und Assurances Cases mit Bezug auf die Prüfung von KI-Systemen zu gestalten sind. Weder ist klar, welche Anspruchsgruppe mindestens an der ATDD-Phase zu beteiligen ist, noch ist geklärt, welche Kompetenzen für die Aufstellung eines Assurance Case benötigt werden. Obschon Assurance Cases versprechen, eine gute Dokumentations- und Argumentationsgrundlage im Rahmen von Audits und Rechtsstreitigkeiten darzustellen, können aufgrund fehlender Standards schließlich auch die jeweils beteiligten Prüfpersonen und Jurist*innen die Güte eines Assurance Case nur schwer bewerten. Erschwerend kommt die hohe Kontextspezifität hinzu und es ist offen, ob es die Betrachtung zahlreicher Einzelfälle auf lange Sicht ermöglicht, übergreifende Gemeinsamkeiten zu identifizieren, die der Standardisierung der Ansätze dienen können.

[10]

Hauer, Marc P.; Adler, Rasmus; Zweig, Katharina (2021): Assuring Fairness of Algorithmic Decision Making. *2021 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW).* Abgerufen am 22.09.2021

[11]

Cockburn, Allistair (2006): Agile software development: The cooperative game. Pearson Education / Adzic, Gojko (2009): Bridging the communication gap: specification by example and agile acceptance testing. Neuri Limited.

Ausblick und Empfehlungen für das Testen von KI-Systemen

Um diese Fragen zu klären und Hürden bei der Umsetzung von ATDD und Assurance Cases überwinden zu können, bedarf es weiterer intensiver Forschungsanstrengungen. Nur so kann die Erzeugung notwendigen Wissens für die Nutzung dieser Ansätze im Besonderen und auf dem Gebiet der KI-Prüfung im Generellen vorangetrieben werden. Hierzu zählt unter anderem die Einrichtung von Experimentierräumen, aber auch die Förderung von Zertfizierungsinitiativen (z. B. DIN SPECS), um schnell Vorschläge zur Prüfung und Zertifizierung von KI-Systemen als Basis für internationale Standards und Normen zu erarbeiten.

Sollte die KI-Verordnung zeitnah in Kraft treten, gäbe es keine Möglichkeit, Hochrisiko-KI gemäß den Anforderungen der Verordnung zu testen und so wäre es ausgeschlossen, diese auf den Markt zu bringen. Die Erforschung von KI-Prüfmethoden und -werkzeugen sowie der Zertifizierung sollte demnach ein dringendes Anliegen von Entwickler*innen, Anwenderunternehmen sowie der Politik selbst sein, um die durch die Nutzung von KI-Systemen versprochenen (Markt-)Potenziale ausschöpfen zu können.

Geht es schließlich um die Frage, durch wen KI-Systeme vor ihrer Zulassung überhaupt getestet und zertifiziert werden sollen, so weist die KI-Verordnung eine wesentliche Schwachstelle auf. Im Gegensatz zur Maschinenverordnung gelten interne Kontrollen bzw. eine Selbstzertifizierungen laut der KI-Verordnung als ausreichend, Drittkontrollen durch notifizierte Stellen sind nicht in jedem Fall vorgesehen. Im Falle von KI-Anwendungen im Personalwesen verlangt Art. 43 Abs. 2 der KI-Verordnung beispielsweise sogar ausdrücklich, dass ausschließlich interne Kontrollen durchzuführen sind. Dies ermöglicht zwar eine recht kostengünstige Prüfung, wodurch insbesondere KMU KI-Systeme einfacher – oder überhaupt – auf den Markt bringen können. Gleichzeitig behalten Entwickler*innen das im Rahmen des Prüfverfahrens generierte Wissen jedoch in der Regel für sich, Nutzer*innen und sonstige Öffentlichkeit haben also keinen Zugang zu den Test- und Zertifizierungsergebnissen. Gefordert werden sollte daher unbedingt, aufgestellte Assurance Cases immer offen zu legen, um es auch Dritten zu ermöglichen, das Testverfahren nachvollziehen und eventuelle Risiken eines KI-Systems erkennen zu können.

Impressum

Eine Veröffentlichung aus dem Projekt „ExamAI – KI Testing & Auditing“ https://testing-ai.gi.de

Dezember 2021

Herausgeberin

Gesellschaft für Informatik e.V. (GI)
Spreepalais am Dom
Anna-Louisa-Karsch-Straße 2
10178 Berlin

Projektleitung

Nikolas Becker
nikolas.becker@gi.de

Gestaltung

Gabriela Kapfer
http://smileinitial.plus

ExamAI – KI Testing & Auditing

Dieses Arbeitspapier erscheint als Teil des Forschungsprojekts „ExamAI – KI Testing und Auditing“, das sich der Erforschung geeigneter Test- und Auditierungsverfahren für KI-Anwendungen widmet. Es steht unter der Leitung der Gesellschaft für Informatik e. V. und wird von einem interdisziplinären Team bestehend aus Mitgliedern der TU Kaiserslautern, der Universität des Saarlandes, des Fraunhofer-Instituts für Experimentelles Software Engineering IESE und der Stiftung Neue Verantwortung getragen und im Rahmen des Observatoriums Künstliche Intelligenz in Arbeit und Gesellschaft (KIO) der Denkfabrik Digitale Arbeitsgesellschaft des Bundesministeriums für Arbeit und Soziales (BMAS) gefördert.
Informationen zum Projekt und weitere Veröffentlichungen finden Sie unter: https://testing-ai.gi.de/

Projektpartner*innen:

GESELLSCHAFT FÜR INFORMATIK

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

INSTITUT FÜR RECHTSINFORMATIK
UNIVERSITÄT DES SAARLANDES

Fraunhofer IESE

Stiftung Neue Verantwortung

Gefördert durch:

DENKFABRIK DIGITALE ARBEITS-GESELLSCHAFT
im Bundesministerium für Arbeit und Soziales

Im Rahmen des:

ki_o
Observatorium Künstliche Intelligenz in Arbeit und Gesellschaft