

Feministische Interventionen zu einer vertrauenswürdigen KI

CORINNA BATH. TANJA KUBES. JANNIS STEINKE

Nach einer wechselvollen Geschichte voller Auf's und Abs wird dem Forschungsfeld Künstliche Intelligenz (KI) seit der Veröffentlichung des Chatprogramms ChatGPT im Herbst 2022 enormes mediales Interesse zuteil. Eine in den Beiträgen häufig formulierte Befürchtung lautet dabei, dass KI, die auf Deep Learning-Ansätzen beruht, soziale Ungleichheiten und Diskriminierungen mit Hilfe neuer Technologien in die Zukunft fortschreibt (AG DIG*IT*AL 2022, 1). Im Fokus der Kritik stehen vor allem die Datensätze, mit deren Hilfe KIs trainiert werden. Ohnehin bereits marginalisierte Gruppen sind darin signifikant unterrepräsentiert, so dass bestehende diskriminierende Strukturen und Muster nicht nur perpetuiert, sondern zugespitzt und verfestigt werden. Seitens der Politik wird seit einigen Jahren versucht, die Entwicklung, Anwendungen und Auswirkungen von KI zu regulieren und dabei auf Ethik und Vertrauenswürdigkeit als zentrale Leitkategorien zu rekurrieren (Europäische Kommission 2019, 2020). An Fahrt aufgenommen haben die Regulierungsbemühungen mit dem im April 2021 veröffentlichten „Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz und zur Änderung bestimmter Rechtsakte der Union“ (im Folgenden: EU-AIA; Europäische Kommission 2021). Der EU-AIA trägt der Tatsache nur unzureichend Rechnung, dass Entwicklung und Gestaltung von Technik eine lange Geschichte der Privilegierung einer *weißen*, cis-männlichen Perspektive ist (D'Ignazio/Klein 2020, o.S.), während neue Produkte vorrangig mit Blick auf die Bedürfnisse dieser spezifischen Anwender*innengruppe konzipiert wurden (Benjamin 2019, 36, 50ff.). Gerade vor dem Hintergrund solcher strukturellen Diskriminierungspraxen aber gilt es, alternativen Ansätzen der Technikgestaltung Gehör zu verschaffen (Costanza-Chock 2020, o.S.).

Die Arbeitsgruppe DIG*IT*AL¹ der deutschen Fachgesellschaft für Geschlechterstudien e.V., der die Autor*innen dieses Beitrags angehören, strebt deshalb eine aktive Rolle bei der Entwicklung entsprechender Ethikrichtlinien an. Mitglieder der AG arbeiteten u.a. bei der deutschen Normungsroadmap für KI mit. Im Juli 2022 veröffentlichte die Arbeitsgruppe zudem eine Stellungnahme zum EU-AIA, in der wir die Debatte um eine feministisch-intersektionale Intervention ergänzen (AG DIG*IT*AL 2022). Die wichtigsten Forderungen aus unserem Positionspapier werden im Folgenden summarisch dargestellt.

Feministische Ergänzungen zu den EU-Ethikleitlinien

In den für den EU-AIA richtungsweisenden „Ethikleitlinien für eine vertrauenswürdige KI“ (Europäische Kommission 2019) wird knapp umrissen, an welchen

Kriterien sich die Entwicklung künstlicher Intelligenzen orientieren sollte. Neben offensichtlichen Anforderungen wie der Einhaltung geltender gesetzlicher Bestimmungen, der Vermeidung von Diskriminierung und der Ausrichtung an den Erfordernissen des gesellschaftlichen Wohlergehens kommt dabei vor allem den Punkten Transparenz und Verantwortung eine zentrale Rolle zu. Die grundsätzliche Zielsetzung des EU-AIA ist mit unseren Forderungen kompatibel, sie geht jedoch in vielerlei Hinsicht nicht weit genug. Gefordert ist angesichts der rasanten Entwicklungen mehr als eine Reihe vager Absichtserklärungen. Es gilt, den heutigen KIs inhärenten Diskriminierungs- und Ausgrenzungsmechanismen *auf allen Ebenen* strategisch entgegenzuwirken. Unsere Forderungen nehmen hier vier thematische Teilbereiche in den Blick: 1. Soziale Ungleichheit und Diskriminierung, 2. Intersektionalität, 3. Partizipation, 4. Vertrauen.

Ungleichheit und Diskriminierung

Fragen der sozialen Ungleichheit und der sozialen Gerechtigkeit werden im EU-AIA nicht ausreichend berücksichtigt. Während Diskriminierung grundsätzlich als Problem erkannt und selektiv durch spezifische Antidiskriminierungsgesetze angegangen wird, bleiben ihre strukturellen Ursachen ausgeklammert. Hier ist dringend geboten, dass die Europäische Kommission explizit *alle* Faktoren, die in Entwicklung, Verbreitung und Nutzung von KI soziale Ungleichheit begünstigen, identifiziert und reguliert und dass bestehende Richtlinien zur Verhinderung von Diskriminierung konsequent umgesetzt werden. Konzepte wie Ungleichheit und Diskriminierung bedürfen dazu einer eindeutigen Klärung und Definition. Der Einbezug von feministischen Perspektiven ist dabei unabdingbar. Ebenfalls unverzichtbar ist, dass die Bekämpfung nicht einem reaktiven Modell folgt, sondern proaktiv und präventiv gegen strukturelle Diskriminierung vorgeht.

Intersektionalität

Der EU-AIA geht an keiner Stelle ausdrücklich auf Intersektionalität ein. Zwar werden verschiedene Kategorien von Vielfalt und potenzieller Diskriminierung (u.a. Bildung, Behinderung, sexuelle Orientierung, Herkunft, Migration, Geschlecht) mit unterschiedlicher Häufigkeit erwähnt, ihre sich verstärkenden Auswirkungen durch Überschneidungen und Interferenzen bleiben jedoch unberücksichtigt. Die Arbeitsgruppe DIG*IT*AL fordert, dass die Europäische Kommission Intersektionalität als Analysekategorie anerkennt, um den exponentiell zunehmenden Auswirkungen von Mehrfachdiskriminierung entgegenzuwirken. Wünschenswert ist dabei auch der verstärkte Einsatz sprachlicher Formen, die eine kritische Reflexion von Fragen der Dekolonialität und Non-Binarität erkennen lassen.

Partizipation

Der EU-AIA stellt technische Lösungen ausschließlich als Top-down-Prozesse dar, ohne ihre Auswirkungen auf Menschen und Umwelt angemessen zu berücksichtigen. Partizipative Ansätze sind darin nicht vorgesehen. Hier fordert die Arbeitsgruppe DIG*IT*AL die Implementierung nachhaltig partizipativer Strukturen, in denen potenziell Betroffene auf allen Ebenen in Entwicklungsprozesse einbezogen werden. Dabei kann auf der vom kritischen Feminismus entwickelten inklusiven, intersektionalen Definition von Partizipation aufgebaut werden. Da der Ausbau von KI-Strukturen nicht nur Menschen betrifft, sondern – etwa durch den enormen Energiebedarf von Rechenzentren – auch massive Auswirkungen auf die natürliche Umwelt hat, sind bei künftigen Entscheidungsprozessen auch die Interessen nicht-menschlicher Akteure wie Tiere, Pilze, Pflanzen, Atmosphäre etc. in angemessener Weise zu berücksichtigen. In Planungs- und Entwicklungsprozesse sind Expert*innen und Repräsentant*innen für *alle* Betroffenen einzubeziehen.

Vertrauen

Auffällig häufig ist im EU-AIA die Rede von Vertrauen. Zweifellos ist das Vertrauen der Nutzenden sowohl für die Verwertung als auch für die Bereitstellung von Daten eine wesentliche Voraussetzung für eine funktionierende KI. Dieses Vertrauen kann allerdings naiver oder informierter Natur sein. Es kann Sorglosigkeit im Umgang mit Daten ebenso widerspiegeln wie das Ergebnis einer reflektierten Risikoabschätzung. Bislang stellt der Begriff der ‚informierten Nutzenden‘ im Text der EU-AIA eine auffällige Leerstelle dar. Um ein *informiertes* Vertrauen in künstliche Intelligenzen zu rechtfertigen, müssen die marktwirtschaftlichen und politischen Interessen hinter der Entwicklung und Gestaltung von KI transparent gemacht werden. Nur allzu oft in der Vergangenheit orientierten sich technologische Innovationen allein an den Interessen einer politischen und finanziellen Elite und ignorierten die Bedürfnisse anderer Gruppen. Sie trugen damit zugleich zur Zementierung bestehender Ungleichheitsverhältnisse bei. Mit unserem Forderungspaket schließen wir uns Appellen anderer kritisch-feministischer Initiativen² an. Ziel der vereinten Anstrengungen muss es sein, KI jenseits einer rein gewinnmaximierenden eurozentrischen Marktlogik neu zu denken, um die nachhaltige Ermöglichung eines lebenswerten Lebens für möglichst Viele zu gewährleisten.

Literatur

AG DIG*IT*AL, 2022: Stellungnahme der AG DIG*IT*AL zum Vorschlag einer EU-Verordnung zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung bestimmter Rechtsakte der Union vom 21. April 2021. Internet: https://www.fg-gender.de/wp-content/uploads/2022/07/Positionspapier-AG-DIG-IT-AL_EU-AIA_Juli2022_final.pdf (10.2.2023).

Benjamin, Ruha, 2019: *Race After Technology. Abolitionist Tools for the New Jim Code*. Cambridge.

Costanza-Chock, Sasha, 2020: Design Justice. Community-Led Practices to Build the Worlds We Need. Internet: <https://designjustice.mitpress.mit.edu/> (22.2.2023).

D'Ignazio, Catherine/**Klein**, Lauren, 2020: Data Feminism. Internet: <https://data-feminism.mitpress.mit.edu/> (22.2.2023).

Europäische Kommission, 2019: Ethik-Leitlinien für eine vertrauenswürdige KI. Internet: http://www.bc-witt.de/download/EU-Ethische-Leitlinien_KI_2019.pdf (10.2.23).

Europäische Kommission, 2020: Weißbuch zur Künstlichen Intelligenz: Ein europäisches Konzept für Exzellenz und Vertrauen. Internet: <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:52020DC0065> (10.2.2023).

Europäische Kommission, 2021: Vorschlag der Europäischen Kommission für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union. [EU-AIA] Internet: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0019.02/DOC_1&format=PDF (10.2.2023).

Anmerkungen

- 1 Diversitätskritische, Intersektionale, Gendertheoretische Interventionen in Technologien und Algorithmen/Kritische Interventionen in Digitalisierungsprozesse.
- 2 Feminist AI Research Network: <https://aplusalliance.org/about-fair/> (10.2.2023); European Digital Rights: <https://edri.org/about-us/> (10.2.2023); Algorithm Watch: <https://algorithmwatch.org/de/> (10.2.2023); netzforma: <https://netzforma.org/> (10.2.2023).