

Gemeinsam KI-Zukunft gestalten. Methoden partizipativer Technikgestaltung für den fairen Einsatz von KI in Österreich

White Paper des Projekts dAlalog.at

Beteiligte Institutionen

IFZ – Interdisziplinäres Forschungszentrum für Technik, Arbeit und Kultur

Center for Human-Computer Interaction, University of Salzburg

Research Group Cooperative Systems, Faculty of Computer Science,
University of Vienna

Robotics – Institut für Robotik und Mechatronik, Joanneum Research
Forschungsgesellschaft

Research & Data Competence



Gefördert im Rahmen des Ideen Lab 4.0 (2019) der FFG.



Inhalt

1	Auf einen Blick	3
2	Ein Brief aus der Zukunft	3
3	Grundsätzliches, oder: Was ist eigentlich "Faire KI"?	4
4	Formate partizipativer Gestaltung von KI im Projekt dAlalog.at.....	5
5	Partizipative Methoden als Begleitung von KI-Entwicklung und Einführung: eine Würdigung.....	6
6	Fazit und Ausblick.....	7
7	Weiterführende Literatur	7

1 Auf einen Blick

Künstliche Intelligenz (KI) ist auch in Österreich auf dem Vormarsch und beschäftigt die breite Öffentlichkeit genauso wie politische Entscheidungstragende in zunehmenden Ausmaß. Ein zentrales Thema ist hierbei die Frage danach, wie KI-Systeme möglichst fair gestaltet werden können und keine von KI-Algorithmen betroffenen Personen Nachteile davontragen müssen. Das Projekt dAlalog.at, gefördert im Rahmen des Ideen Labs 4.0 der FFG, hat sich dieser Frage angenommen und konzipierte Workshopformate für das Verstehen, die Entwicklung und den Einsatz von Künstlicher Intelligenz. Die Workshopformate sollen einerseits den Diskurs der interessierten Öffentlichkeit über faire Künstliche Intelligenz anregen und andererseits Betriebe bei der Entwicklung und Einbindung fairer Künstlicher Intelligenz unterstützen.

2 Ein Brief aus der Zukunft

Der großflächiger Einsatz von Künstlicher Intelligenz verspricht vielfältige Vorteile und wird unseren Privatbereich ebenso wie die Geschäftswelt und den öffentlichen Raum in Zukunft immer mehr beeinflussen. Die Einsatzgebiete reichen von verbesserter Gesundheitsvor- und -fürsorge über bessere Dienstleistungen, effizientere Landwirtschaft bis hin zu neuen öffentlichen Leistungen im Transportwesen¹. Auch wird Künstliche Intelligenz bei der Bekämpfung des Klimawandels zu einem wichtigen Werkzeug werden. Gleichzeitig ist die Diskussion über mögliche Gefahren von Künstlicher Intelligenz in aller Munde. Lassen wir zunächst exemplarisch die Demokratieaktivistin Glacier Kwong zu Wort kommen, die im Rahmen ihrer Kolumne „Brief aus Hongkong“ regelmäßig in der Zeitung „Die Welt“ darüber berichtet, was der Einsatz von KI durch die chinesische Regierung für sie bedeutet: „Als Aktivistin aus Hongkong lade ich mir keine chinesischen Apps aufs Handy, weil ich fürchte, dass die Daten für Überwachungszwecke nach China geschickt werden. Ich weiß, dass ... TikTok in Deutschland sehr beliebt ist. Aber wissen Sie, welche Daten [die App] sammelt? Wenn Sie TikTok erlauben, auf Ihr Fotoalbum zuzugreifen, bedeutet das, dass die App durch alle Fotos geht, die auf dem Handy gespeichert sind – und sogar Gesichtszüge von Ihnen und Ihren Freunden scannt. Wenn Sie der App erlauben, auf Ihren Standort zuzugreifen, bedeutet das, dass sie weiß, wo Sie sich aufgehalten haben.“²

KI-Szenarien wie die von China betriebene mehr oder weniger totale Überwachung samt „Social Scoring“³, auf die hier angespielt wird, sind in Europa bislang noch nicht eingetreten. Trotzdem birgt Künstliche Intelligenz auch hierzulande das Risiko möglicher Eingriffe in persönliche Grundrechte⁴. In Österreich rückte die Diskussion um die Fairness von KI stark in den öffentlichen Fokus, als bekannt wurde, dass ein vom AMS geplanten

¹https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

² <https://www.welt.de/sonderthemen/welt-transformationsgipfel/article219652408/Riskante-Nebenwirkungen-der-kuenstlichen-Intelligenz.html>

³ Der Begriff „Social Scoring“ beschreibt ein auf Überwachungsdaten basierendes Punktevergabesystem der chinesischen Regierung, welches der Kontrolle der chinesischen Bevölkerung dient und Verhalten aus der Sicht der Regierung belohnt oder sanktioniert. Personen und Unternehmen, die zu wenig Punkte erhalten, müssen mit Einschränkungen wie Reisebeschränkungen oder dem Ausschluss von öffentlichen Ausschreibungen rechnen.

⁴ <https://digital-strategy.ec.europa.eu/en/library/communication-fostering-european-approach-artificial-intelligence>

Algorithmus⁵ zur Bewertung der Chancen für erfolgreiche Arbeitsvermittlung, vermutlich Frauen, ältere Personen und Menschen mit außer-österreichischer Herkunft diskriminieren würde.

Solche Gefahren der unfairen Behandlung von Individuen sollten bereits bei der Entwicklung von KI-Systemen bewertet und adressiert werden. Nicht nur würde man gewährleisten, dass niemand Schaden davontragen kann; man würde auch sicherstellen, dass das Vertrauen in KI-Systeme keinen zusätzlichen Schaden nimmt. Auf europäischer Ebene bietet hier der *Artificial Intelligence Act* der Europäischen Kommission⁶ einen Fingerzeig in die richtige Richtung und schlägt einen regulativen Rahmen für den Einsatz von Künstlicher Intelligenz in Europa vor. Das Ziel dieser europäischen Regularien ist es, die Chancen durch die Entwicklung und Anwendung von KI-Elementen zu wahren und gleichzeitig die Risiken bei deren Entwicklung und Einsatz zu minimieren. Hiermit will sich Europa an der Weltspitze der KI-Entwicklung für nachhaltigen, fairen und gerechten Einsatz von KI positionieren.

3 Grundsätzliches, oder: Was ist eigentlich “Faire KI”?

Fairness ist immer das Ergebnis eines gesellschaftlichen Aushandlungsprozesses auf Basis ethischer und moralischer Werte, die im Falle Künstlicher Intelligenz zwingend deren Gestaltung beeinflussen. Die Besonderheit KI-basierter Entscheidungs- und Optimierungssysteme besteht in der hohen Komplexität von Algorithmen, die einerseits oft schwer verständlich und andererseits im Falle von Biases (also Schätzungsfehler, die zu verzerrten Ergebnissen und letztendlich Diskriminierungen führen) oder anderer unerwünschter Konsequenzen im Nachhinein schwierig korrigierbar sind. Unfairness kann dabei zum einen bereits in der Modellierung eines Problems verankert sein, bei der bereits die grundlegenden ethischen Entscheidungen getroffen werden.⁷ Zum anderen sind die von einem Algorithmus gelernten Entscheidungsregeln auch von den verwendeten Trainingsdaten und den verwendeten Methoden abhängig. Die Qualität der Trainingsdaten trägt dabei nicht nur dazu bei, dass KI-basierte Systeme fehlerfrei arbeiten, sondern hat auch signifikanten Einfluss auf mögliche Biases und damit Fairness. Faire KI setzt Transparenz, Offenlegung und Nachvollziehbarkeit voraus, um die Überprüfbarkeit von ethischen Werten wie Diskriminierungsfreiheit, Inklusivität und Gerechtigkeit zu gewährleisten.

Insbesondere im Kontext von sozialer Gerechtigkeit und Diversity ist etwa zu befürchten, dass bestehende Ungerechtigkeiten, Diskriminierungen und vorhandene Stereotypen durch Algorithmen weiter verfestigt werden. Dabei könnte gerade KI bestehende Verzerrungen und ungerechtfertigte Diskriminierung auffinden und somit als höchst effizientes Diagnoseinstrument zu mehr Gerechtigkeit beitragen (vgl. Lopez 2021).⁸ Darüber hinaus könnten faire KI-Systeme auch dazu beitragen, die wirtschaftliche Effizienz als derzeit

⁵ <https://www.derstandard.at/story/2000125884443/regeln-fuer-kuenstliche-intelligenz-ams-algorithmus-auf-dem-eu-pruefstand>

⁶ <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>

⁷ Vgl. O’Neill, Cathy (2016): *Weapons of Math Destruction*. UK: Penguin Books.

⁸ Vgl. Lopez Paola (2021). Artificial Intelligence und die normative Kraft des Faktischen. In *Mercur* #863. <https://www.mercur-zeitschrift.de/2021/03/25/artificial-intelligence-und-die-normative-kraft-des-faktischen/>

dominierendes Optimierungskriterium in einem breiteren gesamtgesellschaftlichen Kontext (bspw. sozialen, ökologischen und nutzer:innenzentrierten Kontext) zu stellen.

Zum Erreichen größerer Fairness von KI werden Interventions- und Verhandlungsräume für einen breiten Diskurs sowie unterschiedliche Formate für das Kennenlernen und Verstehen von KI für die Normalbevölkerung benötigt. Entsprechende Räume und Formate können dazu beitragen, Handlungsspielräume auszuloten und zu realisieren. KI-basierte Systeme werfen viele Fragen und Risiken auf, die in einem transparenten, für alle Personengruppen öffentlich zugänglichen Diskurs einzugrenzen sind. Big Data, maschinelles Lernen und selbstlernende Systeme eröffnen ungeahnte Möglichkeiten im positiven und negativen Sinn und benötigen das Verständnis und die Möglichkeit der kritischen Auseinandersetzung der Bürger:innen. In einem demokratischen System liegt es an den gewählten Vertreter:innen diesen Rahmen zu schaffen sowie Transparenz von Entwickler:innen und Unternehmer:innen einzufordern und dadurch dazu beizutragen die Rechte der Bürger:innen zu schützen. Zur Einbindung von Unternehmen und anderer Stakeholdergruppen sowie der Zivilgesellschaft können partizipative Prozesse unterstützen.

4 Formate partizipativer Gestaltung von KI im Projekt dAlalog.at

Das Projekt dAlalog.at setzt bei der Frage von fairer Entwicklung und fairem Einsatz künstlicher Intelligenz auf das Prinzip partizipativer Gestaltung. Hierzu wurden, basierend auf existierender Fachliteratur und Ergebnissen aus Expert:inneninterviews,⁹ zwei konkrete partizipative Formate entwickelt, um gemeinsam mit unterschiedlichsten Zielgruppen (von konkreten Unternehmen und Organisationen bis hin zur interessierten Öffentlichkeit) faire KI reflektieren, gestalten und einsetzen zu können. Diese Formate (kostenfrei verfügbar unter www.dAlalog.at) bestehen jeweils aus

- (1) Factsheets mit Rahmenbedingungen und notwendigen Überlegungen zur Vorbereitung,
- (2) Hinweisen zum didaktischen Ablauf einschließlich detaillierter Erläuterungen zu Methoden, Inhalten und Themen, und
- (3) Feedbackinstrumenten, die Daten und Rückmeldungen zur Wirksamkeit der Veranstaltung liefern.

Format 1: Workshop „Erster Einblick in faire KI“

Welchen Einfluss hat KI bereits auf unsere Leben, und welche Rollen wollen wir von KI übernehmen lassen? Dieser Online-Workshop erlaubt es den Teilnehmenden ohne speziellem technischen Vorwissen, sich mit den Besonderheiten von und Ansprüchen an eine faire künstliche Intelligenz auseinanderzusetzen. Hierdurch wird eine Mitsprache am Diskurs über erwünschte Rollen von KI in ermöglicht. Konkret werden einer ansprechend designten virtuellen Umgebung mit Hilfe von Fiktions-, Ethik-, Wertekarten sowie

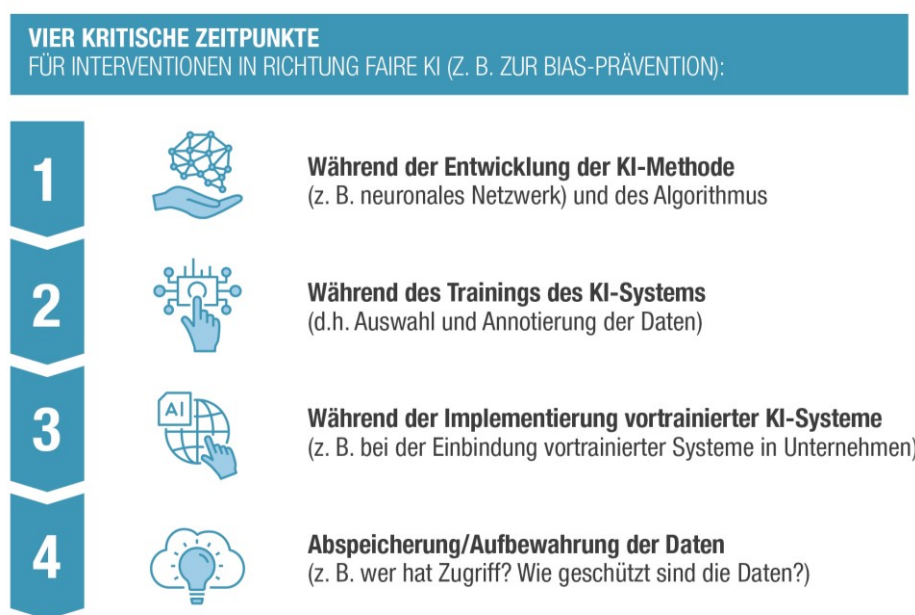
⁹ Thaler, Anita; Vogel, Susanna; Anslinger, Julian; Frauenberger, Christopher; Bentegeac, Philippe; Fiser, Scarlet; Krischkowsky, Alina & Lindlbauer, Andreas (2021). **How to make it fAIr – Methoden partizipativer Technikgestaltung für das Anwendungsfeld der Künstlichen Intelligenz.** Download: https://daialog.at/wp-content/uploads/2021/09/dAlalog.at_-_Methoden-partizipativer-Technikgestaltung-f%C3%BCr-KI.pdf [26.6.2021]

Konsequenzen-Karten verschiedene KI-Szenarien entwickelt und in der Gruppe diskutiert.

Format 2: Workshop-Serie zur Partizipativen KI-Entwicklung

Wie können KI-Systeme für die betriebliche Anwendung fair entwickelt werden? Im Laufe einer Serie von Veranstaltungen wird hier eine „Community of Practice“ aufgebaut, die gemeinsam Wissen schöpft und darauf basierend miteinander agiert. Dabei beschäftigt sich das Workshopformat nicht nur mit Fragen der technischen Entwicklung von KI-Anwendungen, sondern mit dem gesamten sozio-technischen System, d.h. auch der Implementierung oder Anpassung von KI-Systemen an soziale Gegebenheiten im jeweiligen Unternehmen.

Hierzu gibt es vier kritische Zeitpunkte, die für die Entwicklung und Implementierung fairer KI entscheidend sind:



Kritische Interventionszeitpunkte (Grafik aus Thaler et al. 2021, S. 9)

Abhängig vom Zeitpunkt der Intervention ergibt sich schließlich der Spielraum, um gemeinsam an fairer KI zu arbeiten, wobei Technik-Entwickler:innen und Interessensvertreter:innen aus Betrieben in der Umsetzung der partizipativen Prozesse durch moderationserfahrene Personen mit Genderexpertise (z.B. RRI-Expert:innen) begleitet werden.

5 Partizipative Methoden als Begleitung von KI-Entwicklung und Einführung: eine Würdigung

Die Einbindung von Stakeholdern in die Entwicklung von KI, sowie die Schaffung eines gesamtgesellschaftlichen Diskurses der diese Entwicklung richtungsgebend begleitet, sind die notwendigen Voraussetzungen dafür, um zu fairer, gerechter und nützlicher KI zu gelangen. Wie schon der Technologiephilosoph Günther Anders festgestellt hat, prägen die Werkzeuge, die sich die Menschen geben, ganz entscheidend das eigene Selbstverständnis und Selbstbild. Entsprechend ist die größte Aufgabe der KI-Entwicklung nicht die Frage nach der besten Lösung eines Problems, sondern die Frage wie wir zu den KI-Algorithmen kommen, die uns zu den Menschen machen, die wir sein wollen. Um in

einer fairen, gerechten und egalitären Gesellschaft leben zu können, müssen wir technologische Systeme entwickeln, die das Bestehen so einer Gesellschaft durch ihre strukturelle Ordnungsmacht erlauben, und im besten Falle sogar fördern.

Wie sich diese Ordnungsmacht manifestiert, also wie KI aussieht, die diese Qualitäten besitzt, welche Ziele und Funktionen sie hat, sollte so wie bei allen Dingen die uns als Gesellschaft formen einer demokratischen Aushandlung unterliegen. Und genau dafür, so argumentieren wir mit den Arbeiten in diesem Projekt, braucht es vielfältige Formen der Teilhabe an der technologischen Gestaltung von KI. Das reicht von partizipativen Designprozessen, die potentielle User:innen in die industrielle Entwicklung einbindet, bis hin zur Aushandlung im öffentlichen Raum, aus der eine politische Schaffung von geeigneten Rahmenbedingungen für Innovation abgeleitet wird. In jedem Fall, so unser Argument, wäre es fahrlässig diese Aufgabe privaten Akteur:innen außerhalb unserer demokratischen Gestaltungshoheit zu überlassen, oder gar Innovation und Technologie als apolitische oder neutrale Bereiche unsers Lebens zu verstehen.

„Technology is neither good, nor bad – nor is it neutral” (Malvin Kreuzberg)

6 Fazit und Ausblick

Die Frage nach der Fairness von KI ist von der Frage nach Gerechtigkeit nicht zu trennen, und muss von daher höchste Priorität innerhalb der Diskussion um die weitere technologische Ausgestaltung unserer Lebenswelt besitzen. Dem Projekt dAlalog.at kommt dabei eine zweifache Rolle zu: zum einen zeigt es, dass ein interdisziplinär gelebtes Projekt unter Einbindung verschiedenster Stakeholder zu einem wirklichen Fortschritt im Verständnis der vor uns liegenden Herausforderungen beitragen kann, zum anderen lassen sich die im Rahmen des Projekts entwickelten Formate direkt dazu verwenden, den dringend notwendigen öffentlichen Diskurs weiter voranzutreiben und für alle Seiten gewinnbringend zu gestalten. Denn insbesondere für die Zukunft der KI-Entwicklung gilt: dAlalog tut not.

7 Weiterführende Literatur

Thaler, Anita; Vogel, Susanna; Anslinger, Julian; Frauenberger, Christopher; Bentegeac, Philippe; Fiser, Scarlet; Krischkowsky, Alina & Lindlbauer, Andreas (2021). **How to make it fAIr – Methoden partizipativer Technikgestaltung für das Anwendungsfeld der Künstlichen Intelligenz.** Download: https://daialog.at/wp-content/uploads/2021/09/dAlalog.at_-_Methoden-partizipativer-Technikgestaltung-f%C3%BCr-KI.pdf [26.6.2021]

Wolf-Eberl, Susanne; Wolf, Paul; Reichl, Peter; Funk, Michael & Löw, Christian. **Projektbericht zu fairer KI.** Download: https://daialog.at/wp-content/uploads/2021/09/dAlalog.at_Bericht_-_Faire-KI.pdf [26.6.2021]