

EU-KOMM

Stakeholders' Consultation on Draft AI Ethics Guidelines

ETHICS GUIDELINES FOR TRUSTWORTHY AI

High Level Expert Group on Ai (18.12.2018)

Einschätzung /Online-Konsultation (Frist: 31.01.2019)

DGB-Bundesvorstand/Abt. DAA/sy

30.01.2019

- Introduction: Rationale and Foresight of the Guidelines

Es ist zu begrüßen, dass der Entwurf der HLEG dem Ansatz folgt, KI mit europäischen Werten und Prinzipien zu verbinden. Dabei ist besonders hervorzuheben, dass ethische Grundsätze nicht nur benannt werden, sondern auch in konkrete Handlungsempfehlungen / Leitlinien bei der Entwicklung und der Umsetzung von KI münden.

Die formulierten Zielvorstellungen – „do good, do no harm, autonomy of humans, justice, and explicability“ – sind jedoch sehr allgemein, idealtypisch und im Hinblick auf die Arbeitswelt unzureichend. Bei der Zielsetzung im Sinne von „human centric“ wird dem Einsatz von KI ein ethischer Zweck („ethical purpose“) zugeordnet, der neben der technischen Zuverlässigkeit von KI-Systemen („technically robust and reliable“) als tragende Säule einer „vertrauenswürdigen KI made in Europe“ gelten soll.

Grundsätzlich ist zu bedenken, dass KI von Menschen gemacht wird – und damit immer unterschiedlichen Interessen folgt. Weitgehend ausgeblendet werden aber – auch im weiteren Verlauf des Entwurfs der Ethik-Richtlinien – ökonomische und strategische Fragen zum Einsatz von KI in Bezug auf die Veränderung von Wertschöpfung, Wirtschaftsstrukturen, Arbeitsmärkten und Beschäftigung) oder auch arbeitspolitischen Grundsätzen (z. B. individuelle Optimierung und Verantwortung, insbesondere im Arbeits- und Gesundheitsschutz).

KI kann zweifellos dazu beitragen, gesellschaftliche Herausforderungen (Gesundheit, Umwelt oder Klima) zu bewältigen. Jedoch ist auf europäischer Ebene nicht davon auszugehen, dass wirtschaftliche Interessen zur Anwendung und Verwertung von KI-Systemen generell deckungsgleich mit politischen Ansprüchen zur Lösung gesellschaftlicher Probleme sind.

Deshalb ist im Hinblick auf den Einsatz von KI-Systemen in der Arbeitswelt eine **ethische Komponente zur Förderung von Guter Arbeit und sozialem Fortschritt** zu ergänzen. Dieser Grundsatz gilt insbesondere im öffentlichen Bereich, der Vorbildfunktion für die kommerzielle Anwendung bzw. wirtschaftliche Nutzung von KI haben sollte. Dies gilt auch und nicht zuletzt für die notwendigen Prozesse zur Entwicklung und Umsetzung von „trustworthy AI“. **„Gute Arbeit by design“ sollte als Grundprinzip für den Einsatz von KI in der Arbeitswelt gelten.** Dies ist eine logische Konsequenz aus der Zielsetzung der HLEG, nach der KI kein Selbstzweck ist, sondern das menschliche Wohlbefinden verbessern soll. Dies sollte nicht zuletzt für das Arbeitsleben gelten.

Der „human centric“-Ansatz (HCD) setzt nicht nur Information, Transparenz, Nachvollziehbarkeit und Beteiligung voraus, sondern erfordert kontextspezifische Aushandlungsprozesse im Sinne des frühzeitigen Mit-Entscheidens von Stakeholdern wie der Beschäftigten und ihrer Interessenvertretungen über Ziele und Umsetzung von KI-Systemen.

- Chapter I: Respecting Fundamental Rights, Principles and Values - Ethical Purpose

Entscheidend für die Wirkungsweise von KI-Systemen ist die Frage der Entwicklung von Zielen, denen KI-Systeme folgen. Dies wird im Entwurf der HLEG nicht thematisiert und sollte dringend ergänzt werden.

Es wird zwar darauf hingewiesen, grundlegende Prinzipien zu beachten (Accountability, Data Governance, Design for all, Governance of AI Autonomy (Human oversight), Non-Discrimination, Respect for Human Autonomy, Respect for Privacy, Robustness, Safety, Transparency).

Es wird auch darauf hingewiesen, Stakeholder bei der Konzeption und Entwicklung von KI zu beteiligen, wobei aber anzumerken ist, dass alle Stakeholder inkl. insbesondere der Beschäftigten beteiligt werden müssen („Ensure participation and inclusion of stakeholders in the design and development of the AI system“) – analog zu Kap. II bei der Rolle von Bildung und Sensibilisierung bei der Sicherstellung von vertrauenswürdiger KI (wo es heißt: „Trustworthy AI requires informed participation of all stakeholders“).

Es ist zu begrüßen ist, dass der soziale Dialog als wichtiges Element benannt wird, denn der wirtschaftliche Erfolg hängt sehr eng mit der gesellschaftlichen Akzeptanz, nicht zuletzt bei den Beschäftigten zusammen. Hier ist jedoch eine Beschreibung der Verhandlungsprozesse erforderlich, wie zum Beispiel zentrale Kontrollstrukturen für Branchenlösungen (vgl. ‚AI-Now‘-Bericht 2018): „Governments need to regulate AI by expanding the powers of sector-specific agencies to oversee, audit, and monitor these technologies by domain.“ (https://ainowinstitute.org/AI_Now_2018_Report.pdf).

Problematisch ist jedoch, dass sich der Beteiligungsansatz nicht explizit auf die Entwicklung der Ziele des KI-Systems bezieht, obwohl konstatiert wird, dass Zielkonflikte („fundamental tensions between different objectives“) bestehen können.

Es ist unzureichend, diese Trade-offs nur zu kommunizieren und zu dokumentieren, wie es im Entwurf vorgeschlagen wird. Vielmehr ist es für die Umsetzung und Akzeptanz von entscheidender Bedeutung, dass Optimierungsziele und mögliche Zielkonflikte bereits vor der Implementation von KI-Systemen verhandelt und gelöst werden. Dies gilt insbesondere für den Einsatz von KI in der Arbeitswelt.

Die HLEG beschreibt im Entwurf zwar die Notwendigkeit, Machtverhältnissen besondere Beachtung zu schenken („asymmetries of power or information“) und nennt hier exemplarisch das Verhältnis von Arbeitgebern und Beschäftigten.

Gleichwohl wird (in Chapter II) kein Prozess empfohlen, der auf gleichberechtigten Aushandlungsprozessen der Stakeholder beruht. Die Verhandlung von Optimierungszielen für ein KI-System im betrieblichen Kontext ist jedoch eine zwingende Voraussetzung, um (a) das Erfahrungswissen der Beschäftigten für die Prozessoptimierung zu nutzen, (b) die nutzer- bzw. arbeitnehmerfreundliche Umsetzung zu erleichtern, mögliche Zielkonflikte zu lösen, (d) die Akzeptanz bei den Beschäftigten zu erhöhen und damit (e) Synergien für eine erfolgreiche Umsetzung

im Sinne von Guter Arbeit (Verbesserung der Arbeitsbedingungen) und wirtschaftlichem Erfolg (Erhöhung der Produktivität) – oder auch gesellschaftlichem Nutzen – zu erreichen.

Die Aushandlungsprozesse sind von besonderer Bedeutung, wenn KI-Systeme auf persönlichen Daten von Beschäftigten basieren oder diese tangieren. Information und Beteiligung sind hier ebenso unzureichend wie das Prinzip der informierten Einwilligung („informed consent“), das angesichts der Machtasymmetrie von abhängig Beschäftigten ohnehin kein Maßstab im Arbeitsleben sein kann.

Um den Machtasymmetrien in Beschäftigungs- und Auftragsverhältnissen insgesamt Rechnung zu tragen, sollte daher unter den Fundamental Rights of Human Beings ein eigener Punkt 3.6 zu „Workers rights“ aufgenommen werden. Darunter sind insbesondere „Gute Arbeit by design“, gleichberechtigte Aushandlungsprozesse im Sinne von Mitbestimmungsrechten, informationelle Selbstbestimmung von Beschäftigten (bzw. Erwerbstätigen in Abhängigkeitsverhältnissen), Diskriminierungsverbot, sowie insbesondere das Recht auf Vereinigungsfreiheit inkl. des Rechts auf Streik aufzuführen, um effektive Machtressourcen Beschäftigter beim Mit-Entscheiden über Ziele und Einsatzweisen von KI rechtlich-institutionell abzusichern.

- Chapter II: Realising Trustworthy AI

Wie bereits in den Anmerkungen zu Chapter I beschrieben, mangelt es an der Beschreibung von Verhandlungsprozessen, die zur Erreichung der Zielsetzung eines „human centered design“ bedürfen. KI wird von Menschen gemacht und folgt somit auch unterschiedlichen Interessen.

Der Ansatz, die o. g. Leitlinien bereits am Beginn der Entwicklung von KI-Systemen zu berücksichtigen und den Einsatz von KI als kontinuierlichen Prozess ist zu begrüßen. Die Beschränkung auf „Information und Beteiligung der Stakeholder“ ist jedoch unzureichend. Erforderlich sind **partnerschaftliche Aushandlungsprozesse** sowie eine Dokumentation der Optimierungsziele und deren kontinuierliche Evaluation. Dafür sind die Mitbestimmungsrechte zu verbessern, denn es geht nicht nur um das Einholen von Meinungen, sondern es geht darum, bei dem „was, wozu und wie“ des Einsatzes von KI mitzuentcheiden.

Die Sicherstellung der Nachvollziehbarkeit (Traceability) von der Wirkungsweise von KI-Anwendungen ist zu begrüßen, erfordert jedoch auch Interventionsmöglichkeiten. Der Vorschlag für ein Opt-Out (nach dem Entwurf im Bereich Scoring) ist für Beschäftigte im betrieblichen Kontext keine realistische Option. Vielmehr sind die Überwachungs-, Optimierungs- und Vorhersagemöglichkeiten menschlicher Arbeit bzw. Leistungsfähigkeit durch KI zu beachten und negative Konsequenzen für Beschäftigte gemäß Art. 22 DSGVO und Erwägungsgrund 71 DSGVO auszuschließen. Das Letztentscheidungsrecht muss immer beim Menschen liegen. Klärungsbedürftig ist auch die Frage der Verantwortung im Umgang mit Entscheidungsvorschlägen von KI-Systemen für Beschäftigte.

Es ist zu begrüßen, dass die HLEG die Bedeutung von Aus- und Weiterbildung anerkennt. Insbesondere die Ausbildung von KI-Entwicklern sollte die ethischen Prinzipien für KI beinhalten. Zu kritisieren ist allerdings, dass Beschäftigte (außer im Kontext der Anwendung (users) nicht berücksichtigt sind. Darüber hinaus sollte es insbesondere im Arbeitskontext nicht nur um die Ausbildung für KI gehen, sondern um die Qualifizierung für den Einsatz und Umgang mit KI-Systemen (Veränderung von Tätigkeitsprofilen etc.).

Dafür ist eine **technische und soziale Folgenabschätzung von KI-Systemen** am Beginn der Entwicklung und in der Umsetzung erforderlich, die sich neben Qualifizierungserfordernissen auch auf Belastungsveränderungen und die Frage der Datennutzung bezieht. Menschenzentrierte KI bedeutet auch, dass eine Richtlinie Stellung dazu bezieht, unter welchen Bedingungen z. B. Trainingsdatensätze generiert werden.

- Chapter III: Assessing Trustworthy AI

Es ist zu begrüßen, dass konkrete Ansätze und Fragestellungen aufgelistet werden, um KI-Anwendungen kontinuierlich zu monitoren und zu bewerten.

An der – offenen – Frage zur Verantwortung und zum Umgang mit potenziellen Problemen oder Risiken zeigt sich exemplarisch die Notwendigkeit, gemeinsam verbindliche Prozesse der Überprüfung zu vereinbaren (s. Anmerkungen zu Chapter II).

So ist es zwar zu begrüßen, dass Prozesse darauf geprüft werden sollen „to allow a human control, if needed“ (Assessment List Punkt 4. „Governing AI autonomy“). Dabei darf es aber nicht darum gehen „to keep a human in the loop“. Es bedarf klarer Maßnahmen, die den Menschen sowohl hinsichtlich der Ressourcen (technische Vorrichtungen u.ä.) und organisatorisch (Zeit, Haftung, etc.) als auch qualifikatorisch in allen Prozessen befähigen, diese Kontrolle auszuüben.

Es wird vorgeschlagen, die geplante Liste zur Bewertung der use cases (p. 28) um die Frage nach den Prozessen zu erweitern, um KI für Gute Arbeit zu nutzen (Entwicklung und Folgenabschätzung).

- General Comments

Es ist zu begrüßen, dass Regulierungsfragen in einem weiteren Schritt diskutiert werden sollen. Um das Ziel zu erreichen, „ein Klima für Innovation und Akzeptanz von KI zu fördern“, sollte der rechtliche Rahmen für Aushandlungsprozesse und Gestaltungsoptionen von KI in der Arbeitswelt auf die Agenda gesetzt werden. Die Akzeptanz bei den Beschäftigten ist eine entscheidende Sollbruchstelle für die Umsetzungsaussichten von KI in Unternehmen.

Zu begrüßen ist auch, dass kritische Anmerkungen formuliert worden sind, wenngleich dazu kein Konsens innerhalb der HLEG gefunden werden konnte. Um Risiken zu minimieren und damit die Chancen für gesellschaftlichen Fortschritt zu erhöhen, ist eine offene Diskussion erforderlich. Dies betrifft das Verhältnis von Mensch und KI – und letztlich immer die Frage, wofür KI eingesetzt wird und wer über die Optimierungsziele entscheidet.

In dem Zusammenhang stellt sich jedoch die Frage, weshalb bei der Entwicklung von Leitlinien, die ausdrücklich zu einer „vertrauenswürdigen KI Made in Europe“ – also auch entsprechende Leitlinien Made in Europe – führen sollen, auch außereuropäische Unternehmen wie z.B. Google nicht lediglich als Sachverständige angehört wurden, sondern eine direkte Mitgliedschaft mit entsprechendem Mitspracherecht in der HLEG erhielten.

- **Grundsätzliche Anmerkung/Fundamentals**

Grundsätzlich wirkt sich der Einfluss von KI sowohl auf Arbeitswelt als auch auf Demokratie und Gesellschaft aus. Es ist zu begrüßen, dass in Kap. I Punkt 5.3 Bezug darauf genommen wird, dass der Einsatz Künstlicher Intelligenz keinem großangelegtem staatlichen „citizen scoring“ dienen darf. Allerdings sollte dies auch in Bezug auf private Unternehmen gelten. Sowohl Staaten als auch Unternehmen sollte es weder erlaubt *noch möglich* sein, menschliche Profilbildungen wie bspw. „moral personality“ oder „ethical integrity“ zu erstellen. Die hier vorgeschlagenen „Opt-out“-Funktionen sind klar abzulehnen und selbst mögliche „Opt-In“-Funktionen dürfen nicht so gestaltet sein, dass sie Grundrechten widersprechen und mit einem Verzicht auf für den Menschen nützliche Dienstleistungen einhergehen. Für Arbeit und Leben wichtige und nützliche KI-basierte Dienstleistungen müssen so gestaltet werden, dass sie keine Datenerhebungen voraussetzen, die für menschliche Profilbildungen nutzbar wären.

Die Erstellung großer Datensätze ist immer mit dem Risiko der Hackbarkeit (hackability) und beabsichtigter wie unbeabsichtigter Leaks verbunden. Insofern ist auch die Leitidee von „Datensouveränität“ nur tragfähig, wenn dieser Sicherheitsaspekt mitgedacht wird. Das bedeutet, strukturelle Maßnahmen der Dezentralisierung bzw. in manchen – für den Menschen politisch, privat oder arbeitstechnisch brisanten – Bereichen auch die ausdrückliche Nichterhebung von Daten festzulegen. Das Grundrecht auf informationelle Selbstbestimmung, auf Meinungs- und Koalitionsfreiheit darf nicht durch die Schaffung solcher Datenbasen gefährdet werden.