

**Ethische Leitlinien für die Entwicklung und den Einsatz von Künstlicher Intelligenz (KI):
Gemeinwohl und Gute Arbeit by Design**
(Diskussionspapier: Berlin, März 2020)

Die ethischen Leitlinien richten sich in erster Linie an Entwickler*innen, Programmierer*innen, Entscheider*innen, aber auch an Beschäftigte, die an der Konzeptionierung, Planung, Entwicklung und dem Einkauf sowie dem Einsatz von KI-Systemen in Unternehmen beteiligt sind und damit Verantwortung tragen.

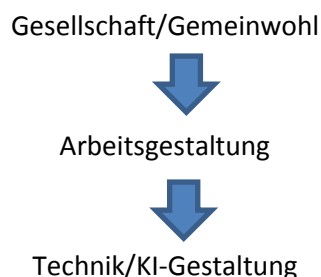
Ganz konkret sollen die Leitlinien Orientierung, Checkliste, Hilfestellung und Handlungsanleitung für diejenigen sein, die KI-Anwendungen entwickeln, einführen und nutzen. Die Leitlinien sind somit Teil eines Anforderungskatalogs (Requirement-Check oder „soziales Pflichtenheft“) für die Planung, Entwicklung, den Einkauf und die Einführung von KI-Systemen.

Darüber hinaus leisten sie einen Beitrag zur Diskussion von bereits vorliegenden ethischen Leitlinien und sozial-ethischen Standards für KI-Anwendungen. Die hier vorliegenden Leitlinien stellen eine Erweiterung der ersten ver.di-Positionierungen dar.¹ Darin ist bereits von „Klare(n) Ethikregeln und Kriterien für Gute Arbeit als Einsatzprinzipien“ die Rede. Nun geht es darum, diese „klaren Ethikregeln“ konkreter zu benennen.

Das Besondere an diesen von ver.di formulierten ethischen Leitlinien ist die gewerkschaftliche Sicht auf KI. Deshalb sind darin die drei wichtigsten Dimensionen: Gute Arbeit, Gemeinwohl und Gute Dienstleistungen. Sie haben einen ausgeprägt sozialen Charakter. Die Einhaltung solcher sozial-ethischen Leitlinien kann ganz entscheidend zu einer Akzeptanz von KI bei den Erwerbstätigen beitragen.

Ziele von KI-Anwendungen

Künstliche Intelligenz ist Werkzeug, Mittel zum Zweck. KI-Anwendungen sollen der Verbesserung von Arbeits- und Lebensqualität dienen. Dazu braucht es gute Dienstleistungen, eine stark ausgeprägte Gemeinwohlorientierung und Gute Arbeit. Die Ziele, die auf gesellschaftlicher Ebene demokratisch zu diskutieren und festzulegen sind, müssen dann bei der Arbeits- und wie auch KI-Gestaltung maßgeblich sein.



¹ innovation-gute-arbeit.verdi.de/themen/digitale-arbeit/beschluesse-und-positionen

Es ist also einerseits festzuschreiben, wofür KI-Systeme entwickelt, aber auch für was sie keinesfalls eingesetzt werden sollen. Dafür sind „rote Haltelinien“ einzuziehen: Es darf keine KI-Anwendungen geben, die Menschen schaden, die gegen Menschen- und Grundrechte verstoßen. Dabei sollte nicht erst ein Schaden nachgewiesen werden müssen, sondern auch bei einem zu hohen Risiko, wenn ein Schaden nicht ausgeschlossen werden kann, sind KI-Systeme nicht einzusetzen. Beispiele hierfür sind autonome Waffensysteme wie auch autonome Medizintechnik wie bspw. bestimmte Robotiksysteme für Operationen. Auch Profiling-Systeme - also Verfahren, die autonom messen, detektieren, klassifizieren, zählen und prüfen - sind grundsätzlich kritisch zu hinterfragen.

Bestehendes Recht und geltende Gesetze sind einzuhalten. Verwiesen sei hier beispielsweise auf die Menschenrechte, das Grundgesetz bzw. die Verfassung, die Tarifautonomie, das Arbeitsschutzgesetz, das Betriebsverfassungsgesetz und die Personalvertretungsgesetze, das Urheberrecht sowie die Datenschutzgrundverordnung. Zu prüfen ist, ob die besonderen Risiken bestimmter KI-Entwicklungen eine Erweiterung bestimmter Gesetze wie das Betriebsverfassungsgesetz nahelegt, weil angesichts der sich abzeichnenden gesellschaftlichen wie betrieblichen Umwälzungen in Zukunft mehr Mitbestimmung in den Unternehmen nötig sein wird.

Das zentrale Ziel einer Verbesserung der Arbeits- und Lebensqualität durch KI soll im Folgenden durch entsprechende ethische Leitlinien und klare Kriterien konkretisiert werden. Neben geltendem Recht sollen diese Kriterien bei der KI-Gestaltung, nicht zuletzt im Rahmen tarifvertraglicher Regelungen, berücksichtigt und eingehalten werden.

Ethische Leitlinien / Kriterien

Sinnstiftung/Nützlichkeit

KI soll für möglichst viele Menschen Sinn stiften und nützlich sein. Sie soll die Lebens- und Arbeitsbedingungen verbessern, indem sie bspw. die Gesundheit schützt, den Arbeitsalltag erleichtert, die Persönlichkeit und den gesellschaftlichen Zusammenhalt fördert.

Barriere-/Zugangsfreiheit

KI-Anwendungen sollen grundsätzlich für alle Menschen zugänglich sein.

Diskriminierungsfreiheit / Inklusion / Geschlechtergerechtigkeit

KI-Systeme sind so zu gestalten, dass durch ihren Einsatz keine Menschen von gesellschaftlichen Prozessen ausgeschlossen und keine Menschen benachteiligt werden.

Nachhaltigkeit

Der Einsatz von Ressourcen (u.a. seltene Erden) und ihr Energieverbrauch müssen in einem angemessenen Verhältnis zu ihrem Nutzen stehen. Ihre Logik muss auf nachhaltige Entscheidungen ausgerichtet sein. Es sind alle Anstrengungen zu unternehmen, die ressourcensparendste KI einzusetzen. Stets ist zu prüfen, ob die aus Umweltgesichtspunkten beste KI eingesetzt wird.

Sicherheit / Robustheit / Kennzeichnung von KI

KI-Systeme müssen sicher sein und dürfen den Menschen und der Umwelt nicht schaden. Ihr Einsatz ist zu kennzeichnen bzw. anzuzeigen, damit Menschen wissen, wenn sie mit KI-Systemen interagieren.

Beschäftigungssicherung / Qualifikationsmöglichkeiten

Bei der Entwicklung von KI-Systemen ist frühzeitig abzuschätzen, ob und welche Tätigkeiten durch sie ersetzt werden, und wie betroffene Erwerbstätige auf neue Aufgaben vorbereitet werden müssen. Hier spielt insbesondere die Frage, welche Qualifikationen in welchen Zeiträumen erworben werden müssen, eine zentrale Rolle. Gerade die Entscheider in den Unternehmen sollen verpflichtet werden, Entlassungen zu verhindern und Beschäftigten Qualifizierung anzubieten (Freiwilligkeit), deren Arbeitsplätze von KI bedroht sind – besonders aus den durch die KI generierten Gewinne.

Handlungs- und Entscheidungsspielräume

Ganz entscheidend wird es sein, KI-Systeme so zu gestalten, dass die Handlungs- und Entscheidungsspielräume der Erwerbstätigen wie auch anderer Anwender erweitert und nicht reduziert werden. Statt Tätigkeiten zu entwerten, sind sie durch gezielte Qualifizierung aufzuwerten. Dies kann durch Tarifverträge flankiert und durch neue Stellenzuschnitte unterstützt werden.

Gesundheit

KI-Systeme sollen zu einer Entlastung führen und Freiräume schaffen. Sie sind ergonomisch zu gestalten. Es soll ausgewiesen werden, in welcher Form sie tatsächlich Entlastung bringen.

Persönlichkeitsrechte

Viele KI-Systeme basieren auf großen Datenmengen. Dies darf nicht zu einer Gefährdung der Persönlichkeitsrechte führen, deren Wahrung ist bereits bei der Planung und Konzeptionierung der Anwendung zu berücksichtigen. Die Zweckbindung der Daten ist sicherzustellen. Es ist eine Datensouveränität anzustreben. Dabei sind die besonderen Herausforderungen bei Beschäftigten zu beachten. Ein Beschäftigtendatenschutzgesetz ist überfällig.

Beherrschbarkeit

KI-Systeme sind stets so zu konzipieren und zu entwickeln, dass sie beherrschbar bleiben. Letztlich muss stets der Mensch die Entscheidung treffen.

Nachvollziehbarkeit / Erklärbarkeit

Viele der genannten Kriterien sind nur einzuhalten, wenn die KI-Systeme durchschaubar bleiben. D.h. nicht nur die Qualität der Daten ist sicherzustellen, sondern auch ihre Herkunft ist auszuweisen.

Transparenz

Die KI-Verfahren, die Funktionsmechanismen und die vollständigen Entscheidungswege und -parameter sind offenzulegen. Außerdem muss es einen Whistleblower-Schutz geben, um zu gewährleisten, dass Missstände eben gerade von denen aufgedeckt werden können, die am meisten Einsicht in die Systeme haben. Zur Transparenz gehört auch, die Einhaltung der hier genannten Kriterien zu dokumentieren. Es muss entsprechende Rechenschaftspflichten geben.

Qualität der Daten

Da die Daten für die Lern- bzw. Trainingsprozesse der KI von entscheidender Bedeutung sind, ist hier besondere Sorgfalt und Transparenz geboten. Die Qualität der Daten entscheidet bspw. mit darüber, ob die KI tatsächlich diskriminierungsfrei ist. Persönlichkeitsrechte der Erwerbstätigen und Bürger*innen sind zu wahren. Zudem sollte eine demokratische Kontrolle von Daten vor allem in öffentlichen Bereichen forciert werden.

Verantwortung / Haftung

Wo Unklarheiten zu Verantwortlichkeiten und Haftung bestehen, sind diese festzulegen. Sie dürfen nicht auf die Technik übertragen werden. Der Mensch bleibt in der Verantwortung.

Prozesse der Entwicklung und des Einsatzes von KI

Um die Einhaltung der Kriterien zu sichern, ist die Beteiligung von Nutzer*innen/Anwender*innen und Beschäftigten sowie ihrer Interessenvertretungen, von Technikfolgenabschätzer*innen und Arbeitsgestalter*innen auch bei Planung und Entwicklung von KI zu gewährleisten.

Um die Einhaltung der Kriterien zu prüfen, sind für die Beteiligten v.a. folgende Fragen hilfreich:

- Wer ist Auftraggeber*in eines KI-Systems mit welchem Ziel?
- Wem und wie vielen bringt es welchen Nutzen?
- Wem schadet es, und wer hat Nachteile?
- Wer wird durch die KI-Anwendung von welchen (gesellschaftlichen) Prozessen ausgeschlossen?
- Wie wurde das ermittelt: Wer wurde wie gefragt?
- Wem gehören die Daten, wer hat Zugriff / Nutzungsrechte?
- Werden Nutzer*innen, Anwender*innen und Beschäftigte bei der KI-Konzeptionierung beteiligt?

Um eine angemessene Beteiligung zu gewährleisten, kann an folgende Prozesse angeknüpft werden, auf deren Einhaltung mindestens zu dringen ist:

- Privacy by Design, also Datenschutz durch Technikgestaltung (Datenschutz-Grundverordnung)
- Technikfolgenabschätzung (v.a. mit Blick auf die Robustheit / Sicherheit der KI-Systeme)
- vorausschauende Arbeitsgestaltung (v.a. vollständige Gefährdungsbeurteilung)
- kontinuierliche Interventions- und Lernmöglichkeiten schaffen
- frühzeitige Mitbestimmung (v.a. durch prozessorientierte Vereinbarungen und Regelungen)