

Ethik-Briefing

Leitfaden für eine verantwortungsvolle Entwicklung und Anwendung von KI-Systemen

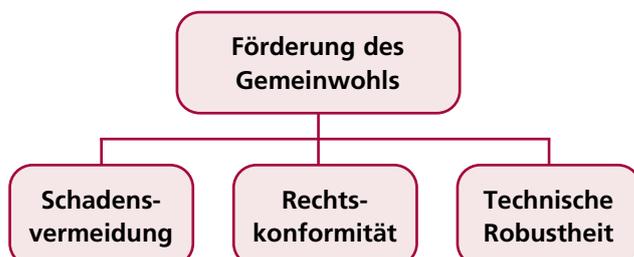
Whitepaper von Jessica Heesen et al.
Arbeitsgruppe IT-Sicherheit, Privacy, Recht und Ethik



Kurzfassung

Ethische Werte und Prinzipien spielen für viele Menschen auch im wirtschaftlichen Kontext eine große Rolle. Dies gilt ganz besonders für die Entwicklung und Anwendung von Künstlicher Intelligenz (KI). Auch wenn die Möglichkeiten, die der Einsatz von KI-Systemen in den verschiedenen Anwendungsbereichen bringt, von vielen positiv bewertet werden, bestehen bezüglich dieser technologischen Innovation auch noch einige Bedenken. Mangelndes Vertrauen ist ein Hauptgrund für die fehlende Akzeptanz von KI-Systemen. Die Arbeitsgruppe IT-Sicherheit, Privacy, Recht und Ethik der Plattform Lernende Systeme hat einen Leitfaden dazu entwickelt, welche Herausforderungen bei der verantwortungsvollen Entwicklung und Anwendung von KI-Systemen beachtet werden müssen. Der Leitfaden skizziert auch Maßnahmen, um einen ethisch reflektierten Technikentwicklungs- und -anwendungsprozess umzusetzen.

Die Entwicklung und der Einsatz von KI-Systemen sollen immer das Ziel verfolgen, zur Förderung des Gemeinwohls beizutragen. Daraus ergibt sich, dass KI-Systeme so eingesetzt werden sollen, dass **Schaden** für Einzelpersonen, Gemeinschaften und die Umwelt **vermieden wird**, die **Rechtskonformität** von KI-Systemen und Akteuren gewährleistet ist und das eingesetzte KI-System die notwendige **technische Robustheit** erfüllt, sodass es zu keinem Zeitpunkt ein unannehmbares Sicherheitsrisiko darstellt.



Darauf aufbauend lassen sich Kriterien für eine verantwortungsvolle Entwicklung und Anwendung von KI-Systemen definieren. Dabei können drei übergeordnete Kategorien ethischer Werte unterschieden werden: **Selbstbestimmung**, **Gerechtigkeit** und **Schutz der Privatheit und der Persönlichkeit**. An diese Werte schließen weiterführende ethische Prinzipien sowie notwendige Voraussetzungen zur Realisierung an.



Selbstbestimmung begründet die Forderung, dass KI-Systeme menschliche Akteure bei der Entscheidungsfindung unterstützen sollen. Notwendige Voraussetzungen zur Realisierung sind **Transparenz** und **Nachvollziehbarkeit** sowie **offene Schnittstellen** und **Systemoperabilität** in Bezug auf Selbstbestimmung auf dem Markt. Das Prinzip der **Gerechtigkeit** fordert **Gleichheit vor dem Gesetz**, **Diskriminierungsfreiheit** und **Chancengerechtigkeit**. **Fairness** in der Verteilungsgerechtigkeit und **Nachhaltigkeit** als intergenerationale Gerechtigkeit sind konkrete Anwendungsbeispiele. Voraussetzung für die Realisierung sind Rechenschaftspflichten. **Privatsphäre** bezeichnet einen Raum, der sich von der Öffentlichkeit abgrenzt und der geschützt ist vor äußeren Eingriffen. Mit zunehmendem Einsatz von KI wird der Schutz der Privatsphäre eine große Herausforderung. **Datenminimierung** und **Transparenz** bei Datenerhebung und -verarbeitung sind Voraussetzungen zum Schutz der Privatsphäre. Diese ethischen Werte können durch konkrete Maßnahmen von unterschiedlichen Akteuren realisiert werden, die in folgendem Leitfaden vorgestellt werden:

Handlungsoptionen für Entwickelnde von KI-Systemen

- | | |
|-------------------------|---|
| Selbstbestimmung | <ul style="list-style-type: none"> ■ Entwicklung von KI-Systemen als Erklärbare KI ■ Transparente Kommunikation des Einsatzes und Autonomiegrades der KI ■ Entwicklung von KI-Systemen, die den Menschen als Letztentscheider vorsehen |
| Gerechtigkeit | <ul style="list-style-type: none"> ■ Vermeidung von diskriminierenden Variablen in Datensätzen ■ Etablierung von Kontrollmechanismen |
| Privatsphäre | <ul style="list-style-type: none"> ■ Verwendung anonymisierter Datensätze ■ Etablierung hoher Sicherheitsstandards |

Handlungsoptionen für Anbietende von KI-gestützten Produkten und Dienstleistungen

- | | |
|-------------------------|---|
| Selbstbestimmung | <ul style="list-style-type: none"> ■ Offene Kommunikation der Verwendung von KI-Systemen ■ Anbieten von Schulungen für Beschäftigte |
| Gerechtigkeit | <ul style="list-style-type: none"> ■ Überprüfung von Verletzungen von Gerechtigkeitsprinzipien durch Formen der Selbstregulierung ■ Etablierung von Erkennungs- und Reaktionsmuster für Schäden |
| Privatsphäre | <ul style="list-style-type: none"> ■ Einholen von Einwilligungserklärungen zur Erhebung und Verarbeitung von Daten und Anbieten von leicht verständlichen Widerrufsmöglichkeiten zu diesen ■ Verwendung anonymisierter Datensätze |

Handlungsoptionen für Nutzende von KI-Systemen

- | | |
|-------------------------|--|
| Selbstbestimmung | <ul style="list-style-type: none"> ■ Verwendung von Produkten, die den Einsatz von KI-Systemen transparent kennzeichnen ■ Wahl von Produkten, die die Marktpluralität fördern |
| Gerechtigkeit | <ul style="list-style-type: none"> ■ Vermeidung des Einsatzes von KI in Bereichen mit hohem Diskriminierungspotential ■ Stichprobenartige Überprüfung der Empfehlungen von KI-Systemen auf Plausibilität |
| Privatsphäre | <ul style="list-style-type: none"> ■ Einholen von Einwilligungserklärungen zur Erhebung und Verarbeitung der Daten ■ Verwendung anonymisierter Datensätze |

Handlungsoptionen für Betroffene von KI-Systemen

- | | |
|-------------------------|---|
| Selbstbestimmung | <ul style="list-style-type: none"> ■ Bereitschaft, bestehende Informationsangebote zu KI-Systemen anzunehmen ■ Aktives Einfordern von Informationsangeboten |
| Gerechtigkeit | <ul style="list-style-type: none"> ■ Verzicht auf die Angabe von Daten, die zu Diskriminierung führen könnten (sofern möglich) |
| Privatsphäre | <ul style="list-style-type: none"> ■ Bereitstellung von Daten nur gegenüber seriösen Anbietern |

Dieser Leitfaden soll als Orientierung für Entwickelnde, Anbietende, Nutzende und Betroffene dienen, die einen ethisch reflektierten Entwicklungs- und Anwendungsprozess von KI-Systemen verfolgen. Deshalb ist es wichtig, dass auch die Rückkopplung in die Praxis von Unternehmen in Deutschland erfolgt. Ein ethisch reflektiertes Design und ein ebensolcher Einsatz sind notwendig, aber nicht immer hinreichend, um alle Risiken, die mit KI für Anwendungen einhergehen, auszuschließen. So kann es durch menschliches oder technisches Versagen sowie durch Missbrauch der KI-Systeme zu nicht intendierten Folgen kommen, die trotz eines umsichtigen Entwicklungsprozesses nicht abzusehen sind.

Die Frage nach ethischen Kriterien für die Anwendung und den Einsatz eines KI-Systems steht in engem Zusammenhang mit der Diskussion um Regulierung und Zertifizierung im Allgemeinen sowie der Kritikalität von KI-Systemen im Besonderen. Allgemein gilt: Je höher die Risiken durch den Einsatz eines Systems sind, desto mehr Gewicht soll die Einhaltung ethischer Prinzipien bereits im Entwicklungsprozess haben. Gleichwohl sollten diese im Optimalfall auch bei Anwendungen mit geringem Schadenspotenzial beachtet werden.

Impressum

Herausgeber: Lernende Systeme – Die Plattform für Künstliche Intelligenz | Geschäftsstelle | c/o acatech | Karolinenplatz 4 | D-80333 München | kontakt@plattform-lernende-systeme.de | www.plattform-lernende-systeme.de | Folgen Sie uns auf Twitter: @LernendeSysteme | Stand: Oktober 2020 | Bildnachweis: tortoon/AdobeStock/Titel

Diese Kurzfassung entstand auf Grundlage des Whitepapers *Ethik-Briefing – Leitfaden für eine verantwortungsvolle Entwicklung und Anwendung von KI-Systemen*, München, 2020. Es wurde erstellt von der Arbeitsgruppe IT-Sicherheit, Privacy, Recht und Ethik der Plattform Lernende Systeme. Die Originalfassung der Publikation ist online verfügbar unter: <https://www.plattform-lernende-systeme.de/publikationen.html>



GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

 **acatech**
DEUTSCHE AKADEMIE DER
TECHNIKWISSENSCHAFTEN