



Positionspapier des Deutschen Forums für Ethisches Maschinelles Entscheiden

Ethik in der Künstlichen Intelligenz: Interdisziplinäre Leitlinien

München, den 10.05.2021

Im Deutschen Forum für Ethisches Maschinelles Entscheiden e.V. (EME) gestalten führende Experten aus Unternehmen, Politik und Wissenschaft die ethischen Leitlinien der Zukunft im Umgang mit maschinengestützten Entscheidungen. Im Vordergrund stehen hierbei die Förderung des allgemeinen Bewusstseins sowie die Schaffung von Transparenz und Verständnis für die Fallstricke im Einsatz von Künstlicher Intelligenz (KI). Um den Schutz der Menschen und der Umwelt zu wahren, setzt sich EME durch den geschaffenen institutionellen Rahmen und die Stärkung des Wissenschaft-Praxis-Transfers unmittelbar ein.

Künstliche Intelligenz: Ethik in der „Black Box“

Als Black Box bezeichnet man in verschiedenen Disziplinen der Wissenschaft ein komplexes System, dessen innere Funktionsweise bewusst oder unbewusst weder untersucht noch beschrieben wird. Die Systembetrachtung beschränkt sich auf den Input und Output. Im Zusammenhang mit KI werden lernende Systeme oftmals selbst zur Black Box für Unternehmen, da kohärente Risiken nur unzureichend berücksichtigt werden. Ein leicht vermeidbares Problem, denn die Risiken im Einsatz von maschinellem Lernen sind unlängst bekannt und können sowohl enttarnt als auch durch entsprechende Maßnahmen verringert bzw. vollständig beseitigt werden. Als problemverstärkender Faktor ist hierbei insbesondere die asymmetrische Verteilung des Risikos anzuführen. Mögliche Auswirkungen wie Diskriminierung und Intransparenz zeigen sich bei Unternehmen oftmals zuerst auf Kundenseite und in der Politik auf der Seite der Gesellschaft. Ein Reputationsschaden beim Einsatz einer fehlerbehafteten KI-Lösung tritt oftmals nur zeitverzögert auf.

Ein gemäß EME akkreditierter Einsatz von KI umfasst aus diesem Grund sieben Grundprinzipien (englische Übersetzung in Klammern):

I. **Gerechtigkeit (Fairness)**

Das Verhindern von Diskriminierung und die Forderung nach Gerechtigkeit ist ein zentraler Aspekt bei maschinellen Entscheidungen. Hierbei können KI-basierte Entscheidungen als Chance angesehen werden, um objektive Entscheidungen zu treffen, die idealerweise frei von subjektiven und möglicherweise unbewusst diskriminierenden Bewertungen sind.

II. **Transparenz (Transparency)**

Der Einsatz von maschinengestützten Entscheidungen bringt immer einen gewissen Kontrollverlust mit sich. Jedoch sollten sowohl die Vorgehensweise als auch der Algorithmus selbst so offen wie möglich gestaltet werden, um Verbesserungen und Nachverfolgbarkeit von getroffenen Entscheidungen sicherzustellen. Gleichermaßen sollten Nutzer und Betroffene über den Einsatz von maschinengestützten Entscheidungen informiert und aufgeklärt werden. Dies kann beispielsweise über entsprechende Hinweise, Infomaterial oder Blogs geschehen.

Deutsches Forum für Ethisches Maschinelles Entscheiden e.V.

An den Römerhügeln 32
82031 Grünwald
E-Mail: info@df-eme.org
Seite: www.df-eme.org

Ansprechpartner

Felix Rank
Head of Strategic Partnerships
Felix.Rank@df-eme.org

Vorstandsvorsitzender

Andreas Banzerus

- III. Datenschutz (Data Privacy & Governance)**
KI basiert auf einer umfassenden Datengrundlage, je nach Einsatzfeld können hier auch Gesundheits- oder sonstige sensitive Daten Verwendung finden. Dabei muss bereits bei der Erhebung der Trainingsdaten für Algorithmen auf die Einhaltung der jeweils gültigen Datenschutzrichtlinien geachtet werden. Dasselbe gilt für die Daten, welche in den Entscheidungsprozess einfließen und auf deren Grundlage ein Ergebnis produziert wird. Durch KI kommt dem Datenschutz eine essenzielle Bedeutung bei.
- IV. Sicherheit (Security)**
Die Algorithmen und Datenmodelle, auf denen eine KI aufbaut, sind aufgrund ihrer Exposition besonders schützenswert. Neben der operativen Sicherheit spielt auch die externe Gefahrlosigkeit eine große Rolle, sobald die Entscheidungen, welche von einer KI getroffen werden, sicherheitsrelevante Bereiche betreffen.
- V. Qualität (Quality)**
Die schnelllebige Gesellschaft erhöht den Druck auf Geschäftsmodelle, sich der Geschwindigkeit der Nutzer anzupassen und die Zeit für das Treffen von Entscheidungen immer weiter zu verkürzen. Die dadurch entstehende mangelnde Qualität einer KI-Lösung führt wie ein Katalysator zu weiteren Problemen, wie beispielsweise folgereichen Fehlentscheidungen, Falschbewertungen oder auch Diskriminierung. Das basale Projektziel sollte stets eine qualitative Verbesserung der bestehenden Entscheidungsprozesse sein, welche anhand von objektiven Qualitätskriterien laufend evaluiert wird.
- VI. Verantwortung (Responsibility)**
KI-gestützte Entscheidungen dürfen keine Möglichkeit bieten, die Verantwortung für Entscheidungen an Maschinen abzugeben. Die zunehmende Digitalisierung, Vernetzung und Automatisierung erhöhen dabei die Verantwortung, welche durch einzelne Technologien und Lösungen auf den Menschen liegt. Es liegt auf der Hand, dass dabei die Qualität und Moralität auf der Strecke bleibt. Durch branchenspezifische Richtlinien für die Entwicklung von KI kann die nötige Sorgfalt für einen verantwortungsvollen Einsatz von KI sichergestellt werden.
- VII. Verzichtsmöglichkeit (Dispensability)**
EME fordert Fallback- und Alternativprozesse bei maschinellen Entscheidungen. Betroffene sollten stets die Möglichkeit einer manuellen, menschlichen Prüfung haben, insbesondere, um mögliche Diskriminierung von unterrepräsentierten Bevölkerungsgruppen zu verhindern. Gleichzeitig sollten, je nach Bedeutsamkeit, weitere Prüf- und Entscheidungsprozesse vorhanden sein, um einen möglichen Ausfall eines Systems abfangen zu können.