

The Public Voice

Allgemeine Richtlinien für die künstliche Intelligenz (Universal Guidelines on Artificial Intelligence - UGAI),

Vorgelegt am 23. Oktober 2018 in Brüssel, Belgien

Neue Entwicklungen in der Künstlichen Intelligenz verändern die Welt. Die Veränderungen betreffen viele Bereiche, von der Wissenschaft und Industrie bis hin zu Verwaltung und Finanzen. Der Bedeutungszuwachs von KI-Entscheidungen berührt die grundlegenden Rechte auf Fairness, Rechenschaftspflicht und Transparenz. Die Ergebnisse der modernen Datenanalyse haben für die Menschen erhebliche praktische Folgen. Sie wirken sich in den Bereichen Beschäftigung, Wohnen, Kredit, Handel und Strafverfolgung aus. Viele dieser Techniken sind völlig undurchsichtig, so dass der Einzelne nicht weiß, ob er überhaupt Gegenstand einer KI-Entscheidung war und ob die Entscheidung richtig und fair getroffen wurde.

Wir schlagen diese allgemeinen Leitlinien vor, um über das Design und die Verwendung von KI zu informieren und diese zu verbessern. Die Leitlinien zielen darauf ab, den Nutzen der KI zu maximieren, das Risiko zu minimieren und den Schutz der Menschenrechte zu gewährleisten. Diese Leitlinien sollten in ethische Standards einfließen. Sie sollten in nationales Recht und internationale Vereinbarungen übernommen, in die Praxis umgesetzt und beim Entwurf von Systemen berücksichtigt werden. Wir stellen klar, dass die Hauptverantwortung für KI-Systeme bei den Institutionen zu liegen hat, welche solche Systeme finanzieren, entwickeln und einsetzen.

1. **Recht auf Transparenz.** Jeder Einzelne hat das Recht, die Grundlage einer KI-basierten Entscheidung zu kennen, die ihn betrifft. Dazu gehört die Kenntnis der in die Entscheidung einfließenden Faktoren, der Verarbeitungslogik und der entscheidungsrelevanten Techniken.
2. **Recht auf menschliche Bestimmung.** Jeder Einzelne hat das Recht, dass die ihn betreffenden Entscheidungen letztlich von einem Menschen getroffen werden.
3. **Identifikationspflicht.** Die für das KI-System verantwortliche Institution muss der Öffentlichkeit bekannt gemacht werden.
4. **Verpflichtung zur Fairness.** Die für den KI-Einsatz verantwortliche Institution muss sicherstellen, dass das KI-System keine ungerechten Verhältnisse widerspiegelt, verzerrte Ergebnisse liefert und keine unzulässig diskriminierenden Entscheidungen trifft.
5. **Folgenabschätzung und Rechenschaftspflicht.** Ein KI-System sollte nur eingesetzt werden, nachdem die Zwecke, die Ziele, der Nutzen sowie die Risiken in angemessener Weise bewertet wurden. Institutionen, die KI-Systeme einsetzen, müssen für dessen Entscheidungen verantwortlich sein
6. **Richtigkeit, Zuverlässigkeit und Gültigkeit.** Die Institutionen müssen die Richtigkeit, Zuverlässigkeit und Validität von Entscheidungen gewährleisten.
7. **Verpflichtung zur Datenqualität.** Die Institute müssen festlegen, welche Daten in die KI-Algorithmen einfließen und sie müssen sicherstellen, dass deren Qualität und Relevanz gewährleistet ist.
8. **Verpflichtung zur Gewährleistung der öffentlichen Sicherheit.** Die Institutionen müssen die Risiken für die öffentliche Sicherheit bewerten, wenn KI-Systeme Geräte steuern oder kontrollieren. Sie haben die erforderlichen Kontrollmechanismen zu implementieren, um die Sicherheit zu gewährleisten.
9. **Verpflichtung zur Cybersicherheit.** Institutionen müssen KI-Systeme vor Bedrohungen schützen, die sich aus ihrer Vernetzung ergeben.

10. **Verbot der geheimen Profilerstellung.** Keine Institution darf ein System zur heimlichen Erstellung von Profilen einrichten oder betreiben.
11. **Verbot des umfassenden Scorings.** Kein Staat darf eine allgemeine Bewertung seiner Bürger oder Einwohner einrichten oder betreiben.
12. **Abschaltspflicht.** Jede Institution, die ein KI-System betreibt, hat die positive Verpflichtung zur Abschaltung des Systems, wenn die menschliche Kontrolle über das System nicht länger gewährleistet ist.

Erläuterndes Memorandum und Referenzen

Kontext

Die Universal Guidelines on Artificial Intelligence (UGAI) weisen auf die wachsenden Herausforderungen durch intelligente Computersysteme hin. Sie enthalten konkrete Empfehlungen zur Verbesserung des Designs von KI-Systemen. Im Kern sollen UGAI die Transparenz und Rechenschaftspflicht für diese Systeme voranbringen und sicherstellen, dass die Menschen die Kontrolle über die von ihnen geschaffenen Systeme behalten. Nicht alle Systeme fallen in den Anwendungsbereich dieser Richtlinien. Unser Anliegen sind jene Systeme, die sich auf die Rechte der Menschen auswirken. Vor allem dürfen diese Systeme keinen Schaden anrichten.

Die Erklärung kommt noch zur rechten Zeit. Regierungen in aller Welt entwickeln politische Vorschläge und schaffen öffentliche und private Institutionen, um die Forschung und Entwicklung von „KI“ zu fördern. Dies hat enorme Auswirkungen auf die Gesellschaft, unabhängig davon, ob und inwieweit die Öffentlichkeit an der Gestaltung und Entwicklung solcher Systeme mitwirkt. Die UGAI sind eine gesellschaftliche Reaktion auf diese Herausforderungen.

Die UGAI wurden auf der Internationalen Datenschutzkonferenz 2018 veröffentlicht, einem der weltweit bedeutendsten Treffen von Technologieführern und Datenschutzexperten.

Die UGAI bauen auf der bisherigen Arbeit von wissenschaftlichen Gesellschaften, Think Tanks, NGOs und internationalen Organisationen auf. Die UGAI beinhalten Elemente der Menschenrechtslehre, des Datenschutzrechts und ethischer Richtlinien. Die Leitlinien bauen auf etablierten Grundsätzen für die KI-Governance auf und sie schlagen neue Prinzipien vor, die bisher nicht in politischen Rahmenwerken enthalten sind.

Terminologie

Der Begriff "Künstliche Intelligenz" (KI) ist weit und unpräzise zugleich. Er beinhaltet Aspekte des maschinellen Lernens, regelbasierte Entscheidungsfindung und andere Computertechniken. Es gibt sogar unterschiedliche Meinungen darüber, ob Künstliche Intelligenz überhaupt möglich ist. Die UGAI räumen lediglich ein, dass der Begriff KI im allgemeinen Sprachgebrauch ein breites Spektrum verwandter Themen abdeckt und sie verwenden den Begriff, um die aktuelle Debatte anzuregen. Die Leitlinien versuchen nicht, die begrifflichen Grenzen von KI zu definieren, außer dass die KI einen gewissen Grad an automatisierter Entscheidungsfindung erfordert. Der Begriff "Leitlinien" folgt der Praxis politischer Festlegungen, die sich in erster Linie an Regierungen und private Unternehmen richten.

Die UGAI enthalten Verpflichtungen für "Institutionen" und Rechte der "Einzelnen". Dies ergibt sich aus den fairen Informationspraktiken im Bereich des Datenschutzes. Die UGAI basieren auf dem Schutz des Einzelnen als grundlegendem Ziel. Unter öffentlichen und privaten Institutionen werden diejenigen verstanden, die KI-Systeme entwickeln und einsetzen. Der Begriff "Institution" wurde anstelle des vertraulichen Begriffs "Organisation" gewählt, um den dauerhaften und nachhaltigen Charakter der in den Leitlinien festgelegten Verpflichtungen zu unterstreichen. Ein Prinzip richtet sich an "nationale Regierungen". Der Grund dafür wird im Folgenden erläutert.

Anwendung

Diese Leitlinien sollten in ethische Normen aufgenommen, in nationales Recht und internationale Vereinbarungen übernommen und in die Gestaltung von Systemen integriert werden.

Die Prinzipien

Die Elemente des **Transparenzprinzips** finden sich in mehreren modernen Datenschutzgesetzen, darunter dem US-Datenschutzrecht, der EU-Datenschutzrichtlinie, der Datenschutzgrundverordnung und in dem Übereinkommen 108 des Europarates. Dieses Prinzip soll eine unabhängige Rechenschaftspflicht für automatisierte Entscheidungen ermöglichen, wobei der Schwerpunkt auf dem Recht der Einzelnen liegt, die Gründe von für sie nachteiligen Entscheidungen zu kennen. Auch wenn es einem Einzelnen in der Praxis vielleicht nicht immer möglich ist, die Grundlagen einer bestimmten Entscheidung richtig zu interpretieren, ist es nicht überflüssig, eine solche Erklärung zu ermöglichen.

Das **Recht auf eine menschliche Bestimmung** bekräftigt, dass Menschen und nicht Maschinen für automatisierte Entscheidungen verantwortlich sind. In vielen Fällen, wie beispielsweise beim Betrieb eines autonomen Fahrzeugs, wäre es nicht möglich oder praktikabel, eine menschliche Entscheidung vor einer automatisierten Entscheidung einzufügen. Das ändert aber nichts an der Notwendigkeit, die Rechenschaftspflicht zu gewährleisten. Wenn also ein automatisiertes System versagt, sollte entsprechend diesem Prinzip eine menschliche Beurteilung des Ergebnisses vorgenommen werden.

Identifizierungspflicht. Dieses Prinzip reagiert auf die Identifikations-Asymmetrie, die bei der Interaktion zwischen Individuen und KI-Systemen entsteht. Ein KI-System weiß typischerweise sehr viel über eine Person; die Person kennt vielleicht nicht einmal den Betreiber des KI-Systems. Die Identifizierungspflicht bildet die Grundlage für die KI-Verantwortlichkeit, die darin besteht, die Identität eines KI-Systems und der verantwortlichen Institution klarzustellen.

Die **Fairness-Verpflichtung** erkennt an, dass alle automatisierten Systeme Entscheidungen treffen, die Vorurteile und Diskriminierung widerspiegeln. Aber solche Entscheidungen sollten nicht normativ ungerecht sein. Auf die Frage, was ungerecht oder unzulässig ist, gibt es keine einfache Antwort. Die Bewertung hängt oft vom Kontext ab. Die Fairness-Verpflichtung macht jedoch deutlich, dass eine Bewertung der tatsächlichen Ergebnisse allein nicht ausreicht, um ein KI-System zu bewerten. Auch die normativen Folgen müssen bewertet werden, einschließlich derjenigen, die bereits vor dem KI-Einsatz existierten oder durch ein KI-System verstärkt werden können.

Die **Folgenabschätzungs- und Rechenschaftspflicht** bezieht sich auf die Verpflichtung, ein KI-System vor und während der Implementierung zu beurteilen. Was die Folgenabschätzung betrifft, so besteht ein zentraler Zweck dieser Verpflichtung darin festzustellen, ob ein KI-System eingerichtet werden sollte. Ergibt eine Folgenabschätzung erhebliche Risiken, wie sie in den Grundsätzen zur öffentlichen Sicherheit und Cybersicherheit beschrieben sind, so sollte das Projekt nicht weiter verfolgt werden.

Die **Verpflichtungen zur Genauigkeit, Zuverlässigkeit und Gültigkeit** legen die Hauptverantwortlichkeiten fest, die mit dem Ergebnis automatisierter Entscheidungen verbunden sind. Die Aspekte sind sowohl einzeln als auch in der Gesamtschau zu interpretieren.

Das **Prinzip der Datenqualität** folgt aus den vorhergehenden Verpflichtungen.

Die **Verpflichtung zur öffentlichen Sicherheit** reagiert darauf, dass KI-Systeme Geräte in der physischen Welt steuern. Aus diesem Grund müssen die Institutionen sowohl die damit verbundenen Risiken bewerten als auch angemessene Vorsichtsmaßnahmen ergreifen, welche den Risiken beugen.

Die **Cybersicherheitsverpflichtung** folgt aus der Verpflichtung zur öffentlichen Sicherheit und unterstreicht das Risiko, dass selbst gut gestaltete Systeme das Ziel feindlicher Akteure sein können. Wer KI-Systeme entwickelt und einsetzt, muss diese Risiken berücksichtigen.

Das **Verbot der heimlichen Profilbildung** folgt aus der Identifizierungspflicht. Ziel ist es, die bei KI-Systemen zunehmend auftretende Informationsasymmetrie zu vermeiden und die Möglichkeit einer unabhängigen Rechenschaftspflicht zu gewährleisten.

Das **Verbot des umfassenden Scorings** soll dem Risiko begegnen, dass eine Regierung dem Individuum einen einzigen, multifunktionalen Scorewert zuweist. Das Datenschutzrecht beurteilt universelle Identifikatoren, die die Profilerstellung von Personen ermöglichen, negativ. Solche Identifikatoren sind vielfach reguliert und teilweise verboten. Noch größer ist die Sorge im Hinblick auf eine umfassende individuelle Bewertung, die als "einheitlicher Scorewert" bezeichnet wird. Ein einheitlicher Score spiegelt nicht nur ein umfassendes Profil sondern auch ein vorgegebenes Ergebnis über mehrere Bereiche der menschlichen Aktivität hinweg wider. Es besteht die Gefahr, dass es auch im privaten Sektor zu einheitlichen Scores kommt. Zwar ist denkbar, solche Systeme dem Wettbewerb auf dem Markt und staatlichen Vorschriften zu unterwerfen. Aber da der Einzelne keine Möglichkeit hat, einem von einer Regierung zugewiesenen einheitlichen Score entgegenzutreten, sollten solche Scores verboten werden.

Die **Abschaltspflicht** ist das ultimative Bekenntnis zur Verantwortlichkeit für ein KI-System. Die Verpflichtung setzt voraus, dass die Systeme unter menschlicher Kontrolle bleiben müssen. Ist dies nicht mehr möglich, sollte das System abgeschaltet werden.

(Übersetzt aus dem Englischen von Peter Schaar unter Verwendung von [deepl.com](https://www.deepl.com))