Ethikunterricht für KI-Systeme

Künstliche Intelligenz. Roboter und Co. treffen mittlerweile selbstständig Entscheidungen, doch welche?


FIZKES/ISTOCKPHOTO.COM

Unser Alltag ist geprägt von Normen – und von digitalen Lösungen. Welchen Normen und Werten diese wiederum folgen, ist aber noch nicht geklärt

Als KI-Modell habe ich keine persönliche Meinung, Überzeugung oder Emotionen", schreibt ein bekanntes KI-basiertes Textprogramm auf die Frage, wie man es mit der Ethik hält. Die Funktion des Programms bestehe vor allem darin, Informationen bereitzustellen, und Anfragen auf Grundlage des zugrunde liegenden Trainingsdatensatzes zu beantworten. Der Haken bei der Sache – KI-basierte Systeme treffen längst schon kritische Entscheidungen. Selbstfahrende Autos könnten sich beispielsweise dafür entscheiden, nicht anzuhalten, wenn sie auf der Straße auf ein Objekt treffen, das kein Lebewesen zu sein scheint. Wie geht man damit um? Was bedeutet es auch, wenn ein KI-gestütztes Textprogramm einem Nutzer die Anleitung zum Bau einer Bombe bereitstellt? Sollte es das dürfen? Wenn intelligente Systeme den Menschen immer mehr unterstützen, in manchen Fällen sogar ersetzen sollen, müssen wir uns mit diesen Fragen beschäftigen.

Nachhilfe

Agata Ciabattoni will mit dem vom WWTF geförderten Projekt TAIGER (Training and guiding AI aGents with Ethical Rules) jene Lücke schließen. „TAIGER hat sich zum Ziel gesetzt, die Grundlagen zu schaffen, sodass KI-Agenten auf rechtlich einwandfreie, ethisch sensible und sozial akzeptable Weise arbeiten können", erklärt Ciabattoni das Forschungsvorhaben. Für sogernannte autonome Agenten, wie es etwa das selbstfahrende Auto oder ein Pflegeroboter ist, die selbstständig und ohne menschliche Aufsicht agieren sollen, „ist dieses Unterfangen besonders entscheidend, aber auch besonders schwierig", sagt Ciabattoni. Die Integration der deontischen Logik, einer speziellen Form der Logik, soll dies zusammen mit dem so genannten Reinforcement Learning ermöglichen. „Das ist eine Art des maschinellen Lernens, bei dem Computer trainiert werden, durch das Prinzip von Versuch und Irrtum Entscheidungen zu treffen. Sie erhalten dabei Rückmeldungen in Form von Belohnungen oder Bestrafungen, ähnlich wie Menschen durch Erfahrung lernen", erklärt Ciabattoni. Damit lassen sich komplexe sowie neue Situationen meistern. Eine Garantie dafür, dass autonome Agenten dadurch immer ethisch korrekt handeln, gibt es aber auch hier nicht. Ciabattoni erinnert an den Vorfall vergangenen Sommer, als ein Schachroboter seinem menschlichen Gegner den Finger brach.


LUIZA PUIU

Agata Ciabattoni
Informatikerin, TU Wien

Deontische Logik bezieht sich hingegen, im Gegensatz zur klassischen Logik, nicht auf Ist-, sondern auf Soll-Begriffe. Was soll oder was soll nicht getan werden? Fragen, mit denen man üblicherweise in der Ethik oder in der Rechtsprechung konfrontiert ist. Um über solche Entscheidungen Aussagen treffen zu können, müssen mathematische und computergestützte Hilfsmittel zum Einsatz kommen, denn nur so lässt sich Maschinen etwas „beibringen".

Logische Ethik

KI ist mittlerweile ein Synonym für maschinelles Lernen, das Daten aus der realen Welt verwendet, um den Agenten beim Erlernen neuer Verhaltensweisen zu unterstützen. Um intelligentes Verhalten zu erzeugen, gibt es jedoch auch einen anderen Ansatz, der sich auf die Verarbeitung und Manipulation von Symbolen statt von Daten konzentriert. Beide Ansätze haben ihre Stärken und Schwächen. Ciabattoni und ihre Projektpartner Ezio Bartocci und Thomas Eiter arbeiten an der Integration dieser beiden unterschiedlichen Ansätze. „Indem wir Reinforcement Learning und die deontische Logik zusammenführen, können wir den autonomen Agenten beibringen, richtig zu handeln. Damit verbinden wir das Beste aus beiden Welten", so Ciabattoni. Offen bleibt allerdings die Frage, auf Basis welcher Normen und Werte solche autonomen Agenten überhaupt programmiert werden sollen? „Das Verhalten des Agenten muss in der Tat mit einer Reihe von potenziell widersprüchlichen und mehrdeutigen Normen aus den Bereichen Recht, Ethik, Gesellschaft usw. vereinbar sein", sagt Ciabattoni. „Das Projekt TAIGER zielt darauf ab, Rahmenwerke zu entwickeln, die diese Anforderungen in die Praxis umsetzen. Aber wir verzichten darauf, Aussagen darüber zu machen, welchen Normen KI-Agenten folgen sollen. Diese heikle Frage überlassen wir Ethikern, Juristen, Philosophen und Praktikern." SG

Das Internet ist für alle da – oder?

Das WWTF Projekt „Roadmap DigiTrans" erarbeitet die Rolle der Institutionen in der Förderung des Digitalen Humanismus

Es braucht keine bestimmte Ausbildung, kein erforderliches Mindestalter oder teure Anschaffungen, um in die digitale Welt einzusteigen. Und dennoch haben nicht alle Menschen den gleichen Zugang zum Netz. Digitale Kluft nannte man früher das Phänomen, das Digital-Kundige von technisch Unerfahrenen unterschied, doch mittlerweile geht es um viel mehr. „Die Erwartung, dass im globalen Dorf plötzlich alle über ein Sprachrohr verfügen und Wissen für alle gleich zugänglich wird, hat sich in vielerlei Hinsicht nicht erfüllt bzw. wird sogar von unerfreulichen und nachteiligen Entwicklungen begleitet", sagt Bernhard Jungwirth, Geschäftsführer des Österreichischen Instituts für angewandte Telekommunikation (ÖIAT) und Leiter des Projekts „Roadmap DigiTrans".

Das Projekt des ÖIAT, das den Umgang mit der Digitalisierung durch verschiedene Initiativen fördert und ausbaut, zielt nun speziell auf die Institutionen ab. „Wir wollen Vertreterinnen und Vertreter von Non-Profit- und öffentlichen Organisationen ein besseres Verständnis von Digitalisierung vermitteln, weil sie, in ganz unterschiedlichen Rollen, die digitale Welt wesentlich mitgestalten", sagt Jungwirth.

Fokus Menschen

Im Zentrum der Roadmap stehen die Werte des Digitalen Humanismus. Ein Ansatz, der sich auf ein menschenzentriertes Verständnis von Technologie stützt. Technik wird dabei in erster Linie als ein Tool betrachtet, das dem Menschen nützlich und hilfreich ist. Wie weit wir davon im Moment entfernt sind, zeigen Phänomene wie Hass im Netz, die rasche Verbreitung von Fake News oder auch der Verlust von Privatsphäre. Neben Usern gibt es damit auch viele Loser, auch unter den Institutionen. „Viele sind überfordert von dem rasanten Fortschritt und haben einen großen Bedarf an Orientierung bei aktuellen digitalen Entwicklungen", erklärt Jungwirth. Mit dem Fahrplan sollen sie für technischen Ausbau sowie auf Ebene sozialer und ethischer Fragestellungen zukunftsfit gemacht werden.


ÖIAT

Bernhard Jungwirth
Geschäftsführer ÖIAT

Regelwerk

Die Chancen der Digitalisierung wollen genutzt werden, die Anwendungsmöglichkeiten sind vielfältig. Deshalb braucht es Regeln, die das digitale Miteinander definieren. Institutionen können hier wichtige Verbindungsglieder darstellen. Auch die Wissensvermittlung spielt dabei eine Rolle. Denn zum einen ist mit dem Internet ein gewaltiger Pool an Wissen verfügbar. Das bietet zahlreiche Möglichkeiten und Chancen. Zum anderen erfordert genau dieser Umstand von allen Nutzenden auch eine gewisse Kompetenz, um damit umgehen zu können. Die Auswahl der Information oder die Einschätzung der Glaubwürdigkeit des Inhalts stellt nicht wenige vor große Hürden. „Darüber hinaus: Die Digitalisierung des Lernens ist kein Selbstzweck. Wir sind gefordert, den tatsächlichen didaktische Mehrwert von digitaler Wissensvermittlung in der Praxis deutlich zu machen. Die einfachere Zugänglichkeit allein kann es nicht sein", betont Jungwirth.

Ein „gutes digitales Leben" ist, was es anzustreben gilt, sodass die Vernetzung eine Orientierung in einer immer komplexer werdenden Welt bietet, anstatt genau jene Verbindung durch die Schattenseiten der Digitalisierung zu kappen. SG