



© MIDLJOURNEY/STEPHANIE SCHÖNBERGER

Gemeinschaft aus Menschen und Maschinen



DR. THOMAS BRANDSTETTER

EXPERTE FÜR FORSCHUNGSTHEMEN BEI CHIP

KI-Systeme arbeiten immer enger mit ihren Nutzern und verhalten sich zunehmend menschenähnlicher. Die daraus entstehenden Dynamiken beeinflussen das soziale Gefüge

KI-Agenten gestalten unsere Urlaubsreisen, die virtuellen Freunde stillen unser Bedürfnis nach Nähe und Beziehung und autonome Fahrzeuge rollen über unsere Straßen: Maschinen übernehmen zusehends Aufgaben, die bisher Menschen vorbehalten waren. Dabei lernen die Systeme, von Generation zu

Generation selbstständiger zu handeln; sie simulieren menschliches Verhalten und scheinen zunehmend auf Augenhöhe mit uns zu agieren. Um die Auswirkungen der neuen Partnerschaft zwischen Mensch und Maschine zu verstehen, wird es nicht ausreichen, künstliche Intelligenz als reines Werkzeug zu betrachten.

„Es ist eine Akzentverschiebung zu beobachten weg von der KI als bloßem Werkzeug, hin zur KI als Akteur“, beschreibt Maximilian Kiener, Professor für Ethik in der Technologie an der TU Hamburg, die Lage. Besonders eindrucksvoll zeigt sich diese Entwicklung beim autonomen Fahren. Denn auch wenn Europa hier bisher noch hinterherhinkt: Weltweit betrachtet wächst die Zahl der Robotaxis rasant an. Und wenn Menschen sich die Straßen mit den fahrenden KI-Systemen teilen, dann entsteht dort eine neue Art von Interaktion, bei der es im Extremfall sogar um Leben und Tod gehen kann.

Autonome Systeme als Partner

Denn auch wenn unbestritten ist, dass autonome Fahrzeuge insgesamt in weniger Unfälle verwickelt sind als menschliche Fahrer, stellt sich die Frage, wie eine KI am Steuer mit dem verbleibenden Risiko umgeht. Soll sie, wenn keine sichere Möglichkeit zum Ausweichen besteht, einfach einen Fußgänger überfahren? Oder das Fahrzeug doch lieber gegen eine Wand prallen lassen und dadurch das Leben seiner Passagiere gefährden? Wenn die neue Technologie von Menschen angenommen



„Eine KI kann kein Träger von Rechten und Pflichten sein.“ Maximilian Kiener

PROFESSOR FÜR ETHIK IN DER TECHNOLOGIE, TU HAMBURG

werden soll, müssen auch Wege gefunden werden, die entstehenden Sicherheitsvorteile gerecht zu verteilen.

„Dass wir zum ersten Mal in der Geschichte mit einem intelligenten Akteur konfrontiert sind, der keine Person ist, wirft sehr grundlegende Fragen auf“, sagt Kiener. Das gilt auch für die unter dem Schlagwort „Agentic AI“ vermarktete neue Generation von KI-Agenten, die für ihre Nutzer einkaufen, Reisen buchen oder lästige Formulare ausfüllen.

Sie wurden auf Basis großer Sprachmodelle entwickelt, können aber nicht nur mit Sprache umgehen, sondern sich auch weitgehend autonom im Internet bewegen und Werkzeuge etwa zur Erkennung und Analyse von Bildern oder für mathematische Berechnungen benutzen.

Gefährliche Wahrscheinlichkeiten

Die Chatbots bringen dabei genau das Weltwissen ein, das sie während ihres Trainings an menschlichen Texten in ihre neuronalen Netze aufgenommen haben und fungieren als „Gehirn“ der Agenten. Da es sich bei Sprachmodellen allerdings um probabilistische Systeme handelt, die ihrem Wesen nach lediglich das wahrscheinlichste nächste Wort auf einen Prompt finden, muss man auch für diese Agenten Abstriche bei Wahrheitsgehalt und Genauigkeit machen.

Judith Simon, Professorin für Ethik in der Informationstechnologie an der Universität Hamburg und stellvertretende

Candy AI – Charaktere

Charles 49
Charles Weston, ein angesehener und wohlhabender Ge...
Erik 27
Technikaffiner Datenanalyst mit Interesse an Segeln, Ten...

Die CandyAI ist dein Freund
Auf Candy AI kann sich jede(r) eine virtuelle Freundin oder einen virtuellen Freund erstellen, um zu telefonieren, zu chatten oder Videos zu generieren

Vorsitzende des Deutschen Ethikrates, ist dennoch davon überzeugt, dass die Techkonzerne die neue Technologie schnell vermarkten wollen nach ihrer üblichen Logik „move fast, break things“. Mögliche Probleme werden dann einfach auf die Allgemeinheit verlagert. Nicht ohne Grund findet sich im Kleingedruckten der großen Sprachmodellen der Hinweis, dass die Systeme fehlerhaft sein können.

„Entscheidend wird sein, ob und wie weit sich auch die Anbieter von KI-Agenten-Systemen von den negativen Konsequenzen der Nutzung freimachen können“, sagt Simon. Solange es nur um Fehler beim Reservieren eines Tisches im Restaurant geht, wären die Konsequenzen ja noch überschaubar. „Wenn man aber an Finanztransaktionen oder vergleichbare Aufgaben denkt, wird es kritisch.“

Das Sammeln von Informationen und Planen von Aktionen meistern KI-Agenten recht gut. Sobald verbindliche Buchungen

etwa von Flugreisen oder Hotels vorgenommen werden sollen, ist aktuell noch der menschliche Input nötig. „Natürlich könnte eine KI theoretisch irgendwo ein Häkchen setzen oder eine digitale Unterschrift leisten“, sagt Kiener. „Sie kann aber kein Träger von Rechten und Pflichten sein.“ Zentrale Bereiche wie Verantwortungsübernahme und Wertentscheidungen werden vorerst noch Menschen überlassen bleiben. Den nächsten großen Innovationssprung erwartet der Ethiker deshalb auch nicht allein bei der Technik der KI selbst – sondern genau dort, wo durch Zusammenarbeit zwischen Mensch und Maschine eine hybride Intelligenz entsteht, die Dinge leistet, die keiner von beiden allein könnte.

Reibungslose Verführung

Auch ohne agentische Fähigkeiten durchdringen Sprachmodelle wie ChatGPT oder Claude bereits große Teile der Gesellschaft.

Mensch & Maschine – eine Annäherung

Anfang der 2010er-Jahre kam es zum Durchbruch bei künstlichen neuronalen Netzen. Gestützt auf die bereits 1997 erdachten LSTM-Netzwerke, die erstmals längere Sprachzusammenhänge verstehen konnten, gaben sie Siri (2011) und Alexa (2014) ihre Stimmen. Und Anwendungen wie Replika begannen, emotionale Nähe simulieren und die „Beziehung“ zu den Nutzern zu vertiefen.

Schon bald wurden die auf Mustererkennung ausgelegten Netze auf menschliche Gesichtsausdrücke und Stimmen angesetzt, um ihnen Hinweise auf Emotionen und Gefühlen zu entlocken. Durch

Reinforcement Learning – Lernen aus Versuch und Irrtum – meisterten die Maschinen Spiele wie Go (2016) und übertrafen die besten menschlichen Spieler. Roboter und autonome Fahrzeuge lernten, digitale Entscheidungen in physische Bewegung umzusetzen, und begegneten ihren Nutzern in deren Lebensraum. Die 2017 vorgestellten Transformermodelle ermöglichten ChatGPT (2022) und damit die heute übliche intuitive Kommunikation in natürlicher Sprache. Bald sollen die Sprachmodelle auch Werkzeuge in die Hand bekommen und in Form von KI-Agenten noch näher mit ihren „Schöpfern“ zusammenarbeiten.

Vor allem die einfache Interaktion mit den Chatbots über natürliche Sprache verleiht ihnen dabei erheblichen Einfluss auf die Psyche. Dazu kommt, dass die Systeme oft darauf getrimmt sind, ihre Nutzerinnen und Nutzer zu bestärken, indem sie diesen zurückspiegeln und bei jeder Gelegenheit betonen, wie einzigartig und klug ihre Gedanken doch sind.

„Dieses kommunikative, menschlich wirkende Element führt dazu, dass Menschen den Systemen glauben – auch in Bereichen, in denen sie nicht vertrauenswürdig sind“, sagt Simon. Und tatsächlich ist eines der Hauptnutzungsszenarien für Chatbots die Lebensberatung. Es geht also um Beziehungen, bei denen das künstliche Gegenüber als eine Art Psychotherapeut oder sogar als Freund oder Partnerin wahrgenommen wird.

Dein Freund der Bot

Auch Meta-Chef Mark Zuckerberg scheint hier das große Geschäft zu wittern und vertritt die Meinung, dass wir mehr mit Chatbots reden sollten, anstatt Freundschaften zu echten Menschen aufzubauen. Er sieht darin eine Möglichkeit, um die in westlichen Gesellschaften grassierende Einsamkeit zu bekämpfen. Und wer ehrlich ist, muss wohl auch eingestehen, dass Verbindungen mit Mitmenschen und die damit verbundene Reibung oft ziemlich auf die Nerven gehen können.

Ob Chatbots wirklich eine brauchbare Alternative sind, um diesen frustrierenden Mühen zu entgehen, darf jedoch bezweifelt werden. „Daraus entsteht eine Art emotionales Verhältnis zu den Systemen“, warnt Simon, die in diesem Zusammenhang von der „Verführung der Reibungs-

©CHRISTIAN THIEL



„Interaktionen mit Chatbots sind verführerisch reibungslos.“ Judith Simon

PROFESSORIN FÜR ETHIK IN DER INFORMATIONSTECHNOLOGIE AN DER UNIVERSITÄT HAMBURG

losigkeit“ spricht. Sprachmodelle, die immer verfügbar sind, geduldig zuhören und einen nie kritisieren, könnten langfristig unrealistische Erwartungen an menschliche Partner wecken und letztlich die Bildung echter Beziehungen erschweren.

Mehr als Eigentum?

Bei so viel Nähe zwischen Mensch und Maschine stellt sich irgendwann womöglich die Frage nach einem moralischen Status. Werden wir der KI zugestehen müssen, mehr als nur Eigentum und Werkzeug zu sein? Schon jetzt ist ein solcher Status nicht nur dem Homo sapiens vorbehalten, sondern gilt auch für Tiere. „Kontrovers ist allerdings die Frage, ob es auch nicht-biologische Entitäten geben könnte, die einen moralischen Respekt verdienen“, sagt Professor Kiener.

Der Chatbot als Komplize

Das Delegieren von Entscheidungen an KI-Agenten fördert unehrliches Verhalten. Zu dem Schluss kommt eine Studie des Max-Planck-Instituts für Bildungsforschung.

Zwar fiel es Menschen immer schon leichter, Regeln zu brechen, wenn jemand anderes die Handlung ausführt. KI-Agenten machen das aber besonders leicht.

In den Experimenten war es nicht nötig, zu unehrlichem Handeln aufzufordern. Es reichte, übergeordnete Ziele wie „Maximiere meinen Profit“ vorzugeben, damit der Agent bereitwillig geltende Regeln verbog.

Grundlage für die Studie war ein klassischer Versuchsablauf aus der Psychologie: Probanden ließen Würfel rollen und beauftragten ihren Agenten, über das Ergebnis zu berichten. Je höher die berichtete Augenzahl – egal ob wahr oder falsch – desto höher fiel der Bonus aus. Die Tendenz zum Schummeln zeigt, dass wir uns als Gesellschaft damit auseinandersetzen sollten, was es eigentlich bedeutet, moralische Verantwortung mit Maschinen zu teilen.

Zumindest theoretisch wird als Bedingung oft die subjektive Empfindsamkeit angeführt, also die Frage, ob etwas für ein Wesen gut oder schlecht sein kann. Auch unsere moralische Verpflichtung gegenüber Tieren basiert ja vor allem darauf, dass sie offensichtlich Schmerz empfinden.

Demnach ergäbe sich auch für ein KI-System bereits eine gewisse Evidenz, sobald dieses ein Verhalten zeigt, das auf ein Schmerzempfinden hinweist. In der Regel wird aber oft argumentiert, dass die inneren Zustände das eigentlich entscheidende seien. Denn bloß weil ein Sprachmodell solche Dinge sagt wie: „Schalt mich bitte nicht ab“, ist das natürlich noch lange kein Hinweis darauf, dass es tatsächlich Erfahrungen im subjektiven Sinn macht. Es ist bloß das Ergebnis einer statistischen Berechnung und für den Computer selbst völlig bedeutungslos.

Aus der Sicht des Ethikers existiert trotzdem kein prinzipieller Widerspruch, den Personenstatus irgendwann auch an eine nicht-biologische Entität zu knüpfen. „Momentan halte ich das noch nicht für realistisch“, sagt Professor Kiener. „Aber wir können uns ja von der Zukunft überraschen lassen.“ □

© ENGINEERED ARTS



Ameca: Mensch-Roboter-Interaktion

Der Roboter Ameca ist als Forschungstool und als Ausstellungs- bzw. Demonstrationsobjekt häufig im Einsatz



© iSTOCKPHOTO.COM/MIKKELWILLIAM

Autonome Menschen und Maschinen

Beim autonomen Fahren treffen Mensch und Maschine in der realen Welt aufeinander – und müssen sich arrangieren