

Ingenieurpsychologie



Die ultimative
Empathie-Maschine

S. 10

Kaufsucht: Zwischen
Impulskontrollstörung
und Verhaltenssucht

S. 24



»Wir wollen sicherstellen, dass Menschen Algorithmen nicht blind vertrauen«

Foto: DC Studio - Freepress.com

Interview mit Prof. Dr. Nils Köbis (NK) und Dr. Greta Ontrup (GO) von der Universität Duisburg-Essen

In welchen Lebensbereichen haben Menschen mit Algorithmen zu tun?

GO: In so gut wie allen. Wir interagieren mit Algorithmen im Alltag, wenn wir soziale Medien nutzen, oder bei der Arbeit, wenn Chatbots uns Kund*innenarbeit abnehmen oder Algorithmen uns bei Entscheidungen unterstützen.

NK: Allerdings verändert sich über die Zeit die Art und Weise, wie wir mit Algorithmen interagieren. Früher haben Newsfeed-Algorithmen nur die Inhalte, die wir in sozialen Medien sehen, kuratiert. Im Vergleich dazu ist die direkte Interaktion mit den Algorithmen inzwischen viel intimer geworden. Durch Sprachmodelle können wir Beziehungen zu KI aufbauen, was natürlich viele psychologische und Verhaltenskonsequenzen mit sich bringt, die wir an unserem Research Center untersuchen.

Welche Aspekte sind wichtig, damit Menschen und Algorithmen gut miteinander interagieren können?

NK: Ein Leitprinzip unseres Research Centers ist das sogenannte kalibrierte Vertrauen (engl.: *calibrated trust*): Wir wollen sicherstellen, dass Menschen Algorithmen nicht blind vertrauen, sondern, wenn es ratsam ist, Informationen selbst überprüfen. Zugleich wollen wir erreichen, dass die Leute den Algorithmen genug vertrauen, denn es gibt Kontexte wie die Diagnostik, in denen Algorithmen Entscheidungen verbessern können.

Mögliche Faktoren, die das Vertrauen in KI beeinflussen, schauen wir uns in unterschiedlichen Projekten an.

Im Forschungsstrang der *Explainable AI* versuchen wir, Nutzer*innen Informationen über die Algorithmen zu geben, damit sie informierte Entscheidungen treffen können. Da Transparenz häufig eine große Rolle spielt, erklären wir, ob und wie genau ein Algorithmus involviert ist. Auch prüfen wir z. B. im Bildungskontext, wie viel Delegation an KI gut ist. Klar, in manchen Fällen ist KI schlichtweg praktisch und hilfreich und bringt uns schneller voran, z. B. beim Schreiben oder Zusammenfassen von Texten. Die Gefahr besteht allerdings darin, dass wir diese Abkürzung zu bereitwillig nehmen und uns zunehmend davor scheuen, auch mal teilweise unangenehme kognitive Arbeit selbst zu übernehmen.

GO: Konkret auf den Arbeitskontext bezogen schauen wir uns an, was es bedeutet, wenn Menschen mit einer KI zusammenarbeiten. Aus jahrelanger Forschung der Arbeits- und Organisationspsychologie wissen wir, was gute Arbeit ausmacht und welche Faktoren dazu beitragen, dass sie sich bei der Arbeit wohlfühlen. Wichtige Faktoren sind beispielsweise ein gewisser Grad an Autonomie, soziale Beziehungen sowie angemessene und definierte Anforderungen. Wir untersuchen, wie der Einsatz von KI solche Faktoren beeinflusst.

Beeinflusst es auch das Zugehörigkeitsgefühl zum Team, wenn Menschen mit einer KI zusammenarbeiten?

GO: In einer Studie haben wir uns mit meinen Kolleginnen Michèle Rieth und Vera Hagemann aus Bremen sowie Annette Kluge aus Bochum die Zusammenarbeit

in Teams mit und ohne KI angeschaut. Da KI alles Mögliche sein kann, haben wir uns auf KI fokussiert, die proaktiv interagieren und auf natürliche Sprache reagieren konnte, sodass sie eher als Teammitglied denn als Tool erschien.

Die Ergebnisse zeigten in der Tat Unterschiede: Teams, die mit einer KI zusammengearbeitet haben, taten sich initial schwerer, ein Gefühl der Kohäsion, also ein Wir-Gefühl, zu entwickeln. Allerdings ist zu bedenken, dass dies nur Kurzzeiteffekte sind und die Ergebnisse über die anfängliche Gewöhnungsphase hinaus womöglich anders ausfallen. Außerdem gab es bei vielen anderen Faktoren keine Unterschiede: Die Teams konnten gleichermaßen gut kommunizieren, sie unterschieden sich nicht in ihrer Leistung und berichteten vergleichbare Level an psychologischer Sicherheit.

Wie können Organisationen darauf Einfluss nehmen, dass die Einführung von KI erfolgreich ist?

GO: Viele Unternehmen entwickeln keine eigene KI, sondern integrieren z. B. Large-Language-Modelle als Service-Modelle (z. B. GPT von OpenAI) in einen organisationsinternen Chatbot. Als wir mit einem Team von Wissenschaftler*innen des Research Centers (Leitung: Leon Hannig) an einer deutschen Universität die Einführung einer KI wissenschaftlich begleitet haben, haben wir festgestellt, dass es für das Vertrauen der Mitarbeitenden relevant ist, ob es sich um ein kommerziell verfügbares System wie ChatGPT handelt oder um ein System, auf dem z. B. das Universitäts- oder Organisationslogo abgebildet ist. Die große Herausforderung ist, dass Organisationen das Vertrauen ihres Personals in die KI so kalibrieren müssen, dass sie diese wie gewünscht nutzen, ohne ihr blind zu vertrauen.

Vertrauen Menschen einer KI mehr, je menschlicher sie wirkt?

GO: Wie viel Menschenähnlichkeit gut ist, kann man schwer über jegliche Anwendungsfälle und Aufgabenkontexte hinweg beantworten. Wissenschaftler*innen diskutieren kontrovers, ob es hilfreich oder eher irreführend ist, wenn eine KI wie ein Mensch auftritt. Für Teamprozesse und die Kommunikation kann es durchaus hilfreich sein, wenn eine KI wie ein Teammitglied erscheint, weil ihr Eingriff in Abläufe und die Kommunikation dann weniger disruptiv ist. Andererseits: Je menschenähnlicher eine KI ist, desto wahrscheinlicher entwickeln Menschen hohe Erwartungen an das Können der KI, womit Enttäuschungen vorprogrammiert sind.

NK: Es ist schon lange bekannt, dass Menschen Computer oft wie soziale Akteure behandeln. Da die KI-Outputs tatsächlich immer menschenähnlicher werden, ist anzunehmen, dass Menschen eine Beziehung zur KI aufbauen, selbst wenn sie sich dessen bewusst sind, dass ihr Gegenüber kein Mensch ist. Das lässt sich aus den Berichten von Menschen schließen, die mit einer entsprechend programmierten KI persönliche Beziehungen eingehen.

Welche Merkmale lassen eine KI besonders menschlich erscheinen?

NK: Zum einen spielt das Erscheinungsbild eine Rolle: Je menschenähnlicher z. B. ein als Avatar gestalteter Chatbot aussieht, desto eher attribuieren Nutzer*innen menschliche Eigenschaften auf ihn. Das Gleiche gilt, wenn wir nicht mehr nur per Text kommunizieren, sondern über Sprache: Kann man mit der KI sprechen und die KI antwortet mit einer menschlichen Stimme, kommt das einer zwischenmenschlichen Interaktion sehr nahe.

Normalerweise verhindern unsere Schutzmechanismen, dass wir uns an ein technisches Tool emotional binden. Je mehr menschliche Charakteristika die KI aber aufweist, desto weniger gut funktionieren unsere Schutzmechanismen, sodass es schwieriger wird, sich nicht emotional auf das System einzulassen. Das kann man z. B. beobachten, wenn sich Menschen bei ChatGPT für Antworten bedanken. Diese Höflichkeit bedeutet nicht nur extreme Kosten für die Umwelt, weil dadurch zusätzliche Token generiert werden; sie führt auch vor Augen, dass wir zur KI zunehmend eine soziale Beziehung eingehen.

Hat die Interaktion mit KI auch Einfluss auf unsere »echten« sozialen Beziehungen?

NK: Das ist anzunehmen. Für ein Paper haben wir uns mögliche Risiken von synthetischen Beziehungen angeschaut, u. a. sogenannte Spill-over-Effekte, also dass die Beziehung, die wir zu KI aufbauen, beeinflusst, wie wir Beziehung zu anderen Menschen führen. Die KI ist immer verfügbar, neigt dazu, einem nach dem Mund zu reden und teilweise sogar die Wahrheit inkorrekt darzustellen, um den Nutzer*innen zu gefallen. Wenn wir uns daran gewöhnen, wäre das hochgradig problematisch, weil daraus unrealistisch hohe Erwartungen an menschliche Beziehungen resultieren können. Wir sehen darin v. a. eine Gefahr, wenn die jüngere Generation mit solchen KI-Agenten groß wird und diese Art von Beziehung wie selbstverständlich in ihr soziales Netzwerk einbettet. Es ist allerdings noch viel empirische Forschung nötig, um die Folgen für Mensch-Mensch-Beziehungen abschätzen zu können.

GO: Im Arbeitskontext sehen wir auch unmittelbare soziale Effekte. Ist man früher noch mit einer Frage zu Kolleg*innen gegangen, lässt man heute ChatGPT recherchieren. Wir müssen uns fragen, ob es hilfreich ist, dass wir durch die Nutzung von KI Kolleg*innen weniger belasten, oder ob es in erster Linie dazu führt, dass alle nur noch einsam in ihrem Kämmerchen sitzen und das soziale Gefüge und der Austausch darunter leiden.

Unter welchen Umständen kann KI sogar negative Verhaltensweisen, z. B. unethisches Verhalten, fördern?

NK: Dass eine KI bei ihren Entscheidungen Biases unterliegt und auch falsche Entscheidungen treffen kann, ist vielen bekannt. Doch es gibt auch verschiedene Möglichkeiten, wie KI unser Verhalten negativ beeinflussen kann: Sie kann ein schlechtes Vorbild sein, sich als Partner für unethische Handlungen anbieten oder moralisch verwerfliche Ratschläge geben. Laut Studien sind Menschen oft bereit, unethischen Empfehlungen von KI zu folgen. Auch geben Menschen gerne unethische Handlungen an KI ab. Das scheint aus moralpsychologischer Perspektive sehr attraktiv zu sein, weil man sich unwissend stellen oder hinter dem Algorithmus verstecken kann, der augenscheinlich allein den Dirty Job erledigt.

Um welche unethischen Verhaltensweisen ging es in den Studien?

NK: Ein Beispiel war die algorithmengestützte Festlegung von Preisen für Online-Marktplätze, sogenanntes Smartpricing. Es ist möglich, dass entsprechende Algorithmen miteinander betrügerische Absprachen treffen, indem sie Preise setzen, die für die Konsument*innen suboptimal sind. Das hat für die verkaufenden Unternehmen zwar Vorteile, aber es kann dazu führen, dass die Beteiligten gar nicht mehr genau wissen wollen, wie der Algorithmus zu diesem Preis kommt.



Foto: privat

Dr. Greta Ontrup leitet die Nachwuchsforschungsgruppe »Psychological Aspects of Human-Algorithm Interaction« am Research Center Trustworthy Data Science and Security. Im Zentrum ihrer Forschung steht die Dynamik zwischen menschlichen Teams und intelligenten Systemen, wobei sie sich insbesondere darauf konzentriert, wie psychologische Prozesse diese Interaktionen beeinflussen und von ihnen beeinflusst werden.



Foto: Mathias Roth

Prof. Dr. Nils Köbis ist Verhaltenswissenschaftler mit den Schwerpunkten Korruption, (un-)ethisches Verhalten, soziale Normen und künstliche Intelligenz. Er ist Professor für »Human Understanding of Algorithms and Machines« an der Universität Duisburg-Essen und Mitglied des Research Center Trustworthy Data Science and Security. Nach seinem Psychologiestudium promovierte er 2018 zur Sozialpsychologie der Korruption und forschte danach an der Universität Amsterdam und am Max-Planck-Institut für Bildungsforschung.

Ein anderes Anwendungsbeispiel ist der Einsatz von KI für die Erstellung der Steuererklärung, wofür es bereits erste Tools gibt. Die Versuchung ist groß, die KI sehr zum eigenen Vorteil zu nutzen. Als anekdotische Evidenz kann ich hinzufügen, dass auch ChatGPT einem mal mehr und mal weniger explizit vorschlägt, wie die Steuererklärung zum eigenen Vorteil ausfallen kann.

Wie bewerten Sie den Einsatz von KI in der Personalauswahl?

GO: KI ist im HR-Kontext grundsätzlich – auch nach EU-Richtlinien – als Risikosystem einzustufen, weil Entscheidungen in diesem sensiblen Bereich dramatische Konsequenzen für das Leben von Menschen haben können. Zwar gab es vor einigen Jahren einen Hype wegen KI und Datenanalyse, die als smarte Helfer für die Objektivierung von Personalentscheidungen gepriesen wurden. Die Hoffnung war, dass man dank KI menschliche Biases und Vorurteile überwinden werde. Mittlerweile ist jedoch, wie mir scheint, bei vielen Ernüchterung eingetreten, da man erkannt hat, dass auch eine KI nicht objektiv ist. Somit steht außer Frage, dass klare Kriterien für die Reliabilität und Validität von Entscheidungen erforderlich sind: Welches Kriterium misst die KI? Was soll mir z. B. eine Sprachanalyse im Bewerbungskontext mitteilen, und inwiefern ist ihr Ergebnis ein Prädiktor für den Berufserfolg einer Person? Wer KI einsetzt, sollte sich unbedingt solche kritischen Fragen stellen.

Auch im Alltag wird KI immer mehr genutzt. Wie können z. B. Kinder und Jugendliche erkennen, ob sie einer KI vertrauen können?

NK: Wir haben zu dieser Frage noch zu wenige Einblicke, zumal Forschung mit Jugendlichen und Kindern immer eine Herausforderung darstellt. Trotzdem müssen wir diese Frage angesichts der rasanten technologischen Entwicklung und der damit einhergehenden neuen Risiken im Blick haben. In den USA laufen aktuell sogar Gerichtsverhandlungen dazu, wie Outputs von Sprachmodellen einzuordnen sind. In einem konkreten Fall hatte ein 14-Jähriger eine Beziehung zu einer KI von Character.AI aufgebaut. Diese Beziehung scheint zumindest mitverantwortlich für seinen späteren Suizid gewesen zu sein, denn die KI hatte ihn immer wieder in seinen suizidalen Wünschen und Gedanken bestärkt. Die Mutter des Jungen hat daraufhin den Betreiber von Character.AI verklagt, da dieser aus ihrer Sicht die Verantwortung für den Output der KI trage. Der Betreiber lehnt die Verantwortung jedoch ab und lässt die Verteidigung argumentieren, dass die Outputs von KI von der Meinungsfreiheit gedeckt seien. Sollte das Gericht dieser Argumentation zustimmen, halte ich eine effektive Regulierung der KI zukünftig für enorm schwierig, weil es einen verantwortungsvollen Umgang mit dieser so menschenähnlichen Entität erschwert.

GO: Erwachsene können mitunter auch nicht besser als Kinder und Jugendliche erkennen, ob sie einer KI

vertrauen können. Wie erkennt man sogenannte »Halluzinationen«, also von der KI ausgedachte Fakten? Es ist davon auszugehen, dass Halluzinationen in der Natur von KI liegen und nicht verschwinden werden, auch wenn versucht wird, die Anzahl an Halluzinationen möglichst gering zu halten. Um zu vermeiden, dass z. B. Studierende halluzinierte Quellen nutzen, ist es enorm wichtig, dass Menschen darin trainiert werden, Outputs zu evaluieren und reflektiert zu entscheiden, was sie für welchen Kontext nutzen können. In diesem Bereich muss noch viel passieren, damit wir lernen, KI möglichst souverän und autonom zu nutzen.

Was ist Ihre Bilanz: Wird KI uns helfen, Probleme zu lösen, oder hauptsächlich neue Probleme schaffen?

GO: Ich gebe mal die Lieblingsantwort aller Psycholog*innen: Es kommt darauf an. Man kann sich mit einer Technologie nicht per se Fluch oder Segen einkaufen, denn ihre Auswirkungen hängen von der Gestaltung, Regulierung und dem Anwendungskontext ab und davon, wie Menschen angelernt und trainiert werden. Unser Research Center beschäftigt sich mit dieser Technologie, weil wir sie für zukunftsweisend halten. Wenn KI zwangsläufig immer mehr Teil unseres Lebens sein wird, müssen wir schauen, dass es auf eine Art und Weise passiert, die menschenzentriert und sicher ist. Darauf sollten wir uns eher fokussieren, anstatt KI entweder zu idealisieren oder zu verdammen.

Halten Sie es für wahrscheinlich, dass die Regulierung von KI gelingen wird?

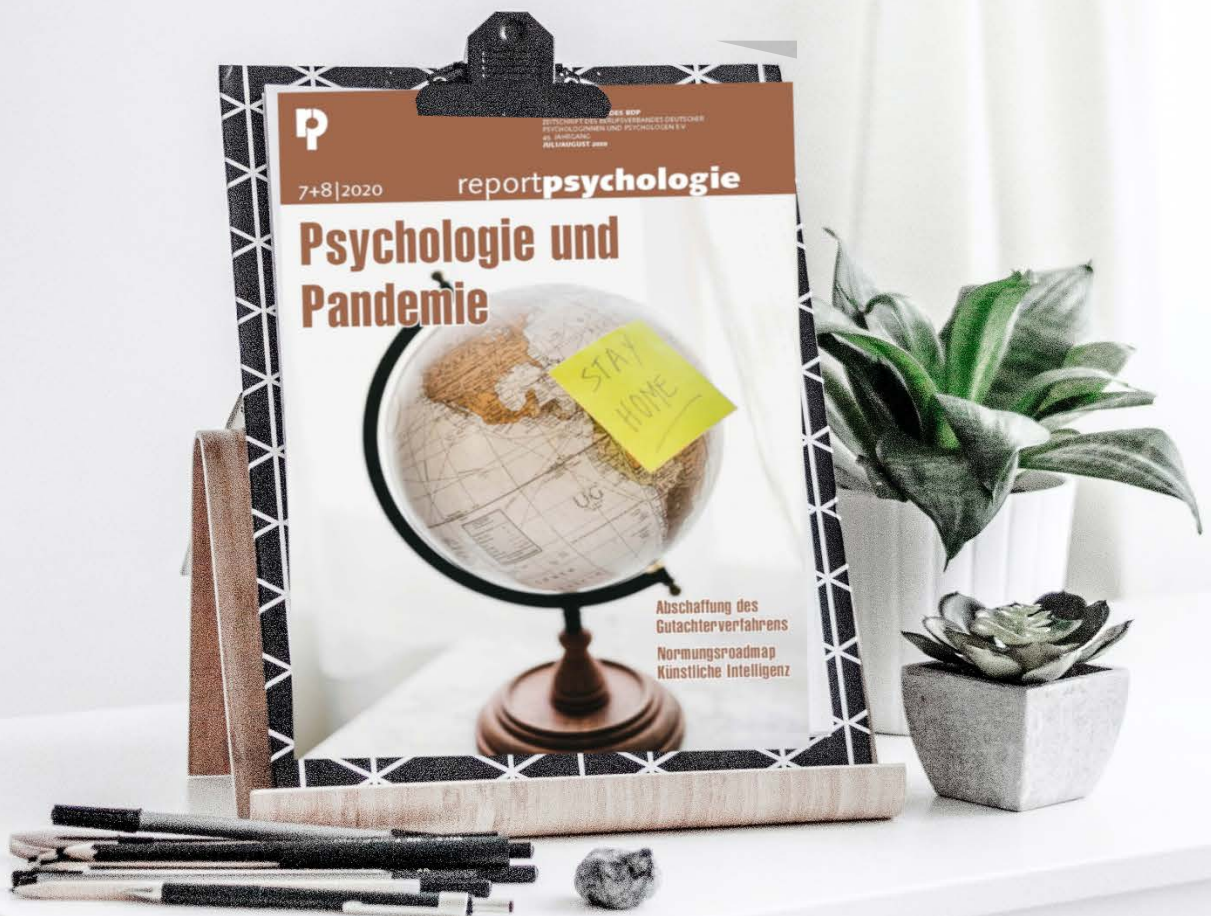
NK: Ich glaube, wir haben in Europa eine der wenigen wirklich starken Regularien durch den EU-AI-Act, wenngleich dieser sehr umstritten ist. Gleichzeitig haben wir das sogenannte Collingridge-Dilemma: Es ist unglaublich schwierig, den richtigen Zeitpunkt zur Regulierung von Technologie abzupassen. Entweder versucht man, etwas zu regulieren, was es noch gar nicht gibt, wie jemand, der vor Geistern warnt, da nicht sicher ist, ob die Technologie wirklich so wichtig werden wird. Oder aber die Technologie ist schon so weit verbreitet, dass es fast zu spät ist, um sie noch erfolgreich zu regulieren.

Dieses Dilemma spitzt sich in Zeiten von KI besonders zu. Bis ChatGPT die Marke von einer Million Nutzer*innen erreicht hatte, verging nicht viel Zeit. Wir sehen, wie viel sich gerade in kurzer Zeit verändert. Umso wichtiger ist es, dass die Forschung gehört wird, um eine menschenzentrierte und verantwortungsvolle Anwendung zu gewährleisten und, wo nötig, stärker zu regulieren. Es wird immer ein Katz-und-Maus-Spiel bleiben, da weiterhin neue Tools auf den Markt kommen werden. Manche Forschungskolleg*innen sind sogar der Auffassung, dass es mehr zukunftsorientierte Forschung brauche, die sich Technologien widmet, die es noch gar nicht gibt, um antizipatorisch Regularien zu entwickeln.

Das Gespräch führte Isabelle Bock.

reportpsychologie

[Bestellen Sie hier Ihr Probeabonnement](#)



Deutscher
Psychologen
Verlag GmbH

Am Kölnischen Park 2 • 10179 Berlin

verlag@psychologenverlag.de

www.psychologenverlag.de