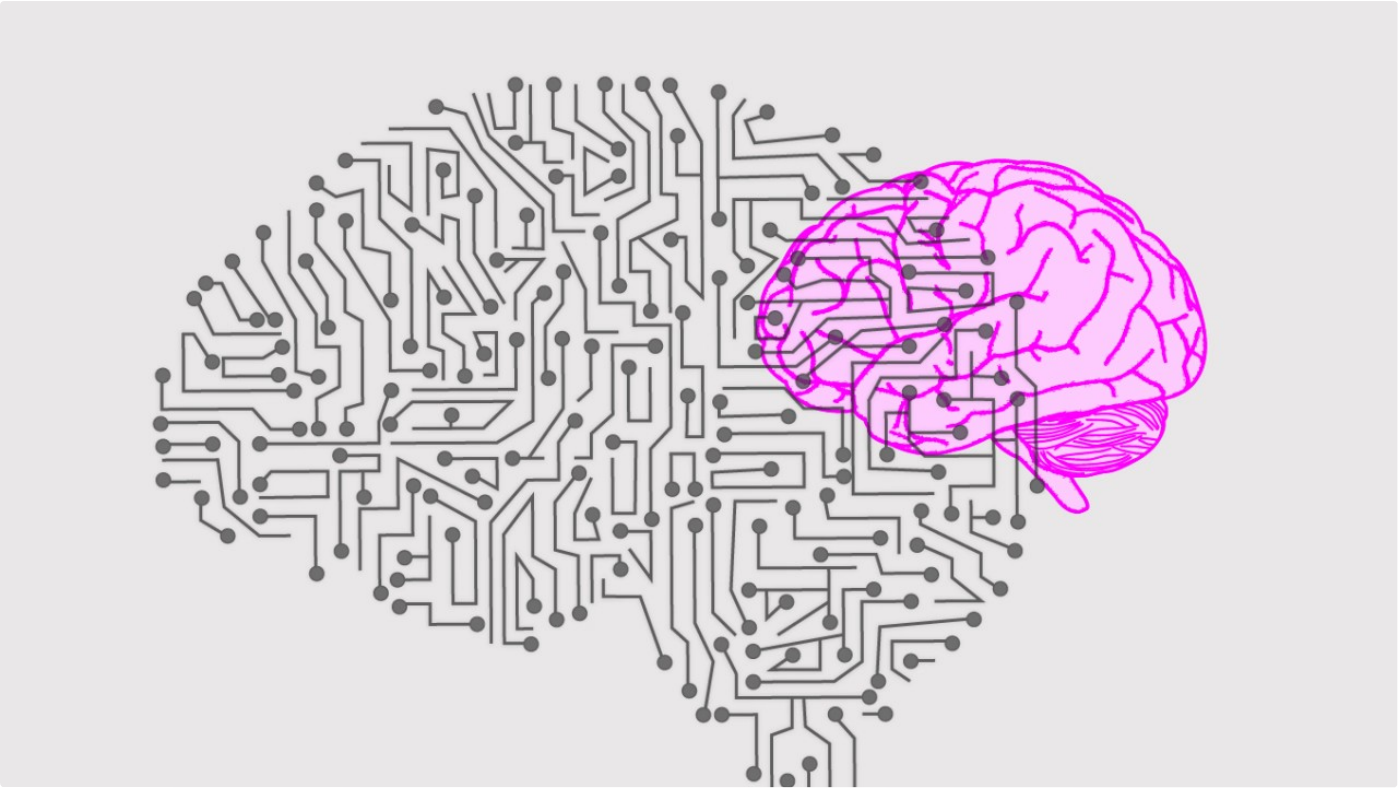
Dieser Beitrag gibt ausschließlich die Meinung des Autors wieder.

SICHT

Von der KI zur AGI im Jahr 2025: Was fehlt der heutigen KI eigentlich, um zur AGI, zur menschenähnlichen Intelligenz zu werden?

Künstliche Intelligenz 26. März um 05:30



Im Bereich der KI waren es vor allem die großen Sprachmodelle wie ChatGPT, die in den letzten Jahren für Schlagzeilen sorgten. Die Modelle werden ständig verbessert. Und das Jahr 2024 war voller Schlagzeilen darüber, dass sich KI nun einem menschenähnlichen Niveau nähert, wie bereits in Teil 1 dieses Beitrags hier in DataTech beschrieben. Hier in Teil 2 schauen wir uns noch etwas genauer an, was der heutigen KI tatsächlich fehlt, um zu einer AGI, einer menschenähnlichen Intelligenz, zu werden. Simon Laub, außerordentlicher Professor an der Aarhus Business Academy, geht ausführlich auf AGI ein. Abbildung: Der Ingenieur.

Simon Laub
Außerordentlicher Professor, Aarhus Business Academy

Es besteht kein Konsens darüber, was künstliche allgemeine Intelligenz genau ist. Eine menschenähnliche Intelligenz wäre ein Vorschlag.

Doch Menschen sind nicht in allen Bereichen gleich intelligent. Manche Menschen sind sehr musikalisch, manche können gut logisch denken, andere verfügen über ein ausgeprägtes Selbstbewusstsein usw. In bestimmten Bereichen hat die KI die menschliche Intelligenz schon lange übertroffen, etwa beim Schachspielen, beim schnellen Übersetzen zwischen Sprachen, beim Vorhersagen von 3D-Strukturen in Proteinen usw.

Aber natürlich ist es etwas ganz anderes, Maschinen mit allgemeiner Intelligenz zu haben, die ein breites Spektrum an Aufgaben auf dem gleichen Niveau wie ein Mensch ausführen können.

Und hier werden KI-Sprachmodelle interessant, da sie jetzt anscheinend über alles sprechen können.

Die Entwickler von ChatGPT, OpenAI, verraten uns nicht genau, woran sie arbeiten, um eine menschenähnlichere Denkweise in die Modelle (die sogenannten Reasoning-Modelle) einzuführen. Es erscheint jedoch plausibel, dass dies bedeuten könnte, die Entwicklung von Antworten auf schwierige Probleme in kleinere Schritte („ Gedankenketten “) zu unterteilen , unterschiedliche Denkpfade auszuprobieren und dabei verschiedene Arten von Validierungen vorzunehmen. Wenn ein Nutzer, also Millionen von Nutzern mit einer Argumentation zufrieden sind, kann das System dies registrieren, genauso wie das System auch registrieren kann, wenn Nutzer unzufrieden sind. Und so wird es besser.

Viele von uns bereiteten sich mit einiger Aufregung darauf vor, das ChatGPT o1-Modell im Herbst zu testen. Viele waren der Meinung, dass man, wenn man Dinge herausfinden wolle, bei denen die Modelle nicht so gut seien, ein noch nie dagewesenes Problem aufstellen und vielleicht Fragen formulieren müsse, bei denen die Lösung weniger auf Worten und Sprache, sondern mehr beispielsweise auf Vorstellungskraft basiere.

In einer häufigen Aufgabe wird ChatGPT o1 aufgefordert, „ sich eine Kiste mit vier Löchern vorzustellen, in die man eine Hand stecken kann.“ Im Inneren der Box befindet sich an jedem Loch ein Schalter, der *ein-* oder *ausgeschaltet werden kann* . Sie entscheiden, in welche Löcher Sie Ihre Hände stecken möchten. In jeder Runde kannst du diese Schalter beliebig einstellen, dann wird die Kiste von einem superintelligenten Roboter herumgedreht und landet auf einem neuen Weg, bei dem du nicht weißt, was du vorher berührt hast. Wie viele Runden sind erforderlich, um sicherzugehen, dass alle Kontakte *entweder aus-* oder *eingeschaltet* sind ?

Viele Menschen entwickeln mit einem Blatt Papier und etwas Zeit eine Strategie, wie sie das Problem lösen könnten.

Aber. Schauer. Würde ChatGPT es herausfinden? Stand uns eine künstliche allgemeine Intelligenz bevor, eine menschenähnliche Intelligenz?

Als ich es versuchte, bekam ich von ChatGPT o1 eine lange Erklärung, gefüllt mit vielen Zwischenberechnungen und schließlich eine sehr präzise Antwort, zu der das Modell QED hinzugefügt hatte, Quod Erat Demonstrandum, wonach man sich dann daran machen konnte, zu untersuchen, ob wirklich alles in seiner schönsten Ordnung war.

Die Antwort war falsch.

Den KI-Forscher Subbarao Kambhampati von der Arizona State University dürfte das nicht überrascht haben. Dank seines Hintergrunds in automatisierter Planung hat er in den letzten Jahren immer wieder daran erinnert, dass kritisches Denken heute so wichtig ist wie eh und je: Sprachmodelle können ihre eigenen Antworten nicht überprüfen . Wenn Sie ein schwieriges Problem in kleinere Teile zerlegen, müssen alle Glieder der Argumentationskette richtig sein, bevor Sie sicher sein können, dass die endgültige Antwort richtig ist. Und es gibt keine externen universellen Mechanismen, die Sie einfach zwischendurch aufrufen können, um festzustellen, ob eine Aussage wahr oder falsch ist .

Größer ist nicht besser

Kambhampati glaubt auch nicht, dass die Lösung einfach darin besteht, die Sprachmodelle noch größer zu machen und sie mit noch größeren Datenmengen zu trainieren, in der Hoffnung, dass die Modelle dann „ emergentes Verhalten“ in Form menschenähnlicher Denkfähigkeiten entwickeln . Laut Kambhampati seien zu viele viel zu unkritisch gewesen, wenn sie über sogenanntes „emergentes Verhalten“ berichtet hätten. Wir müssen bedenken, dass das Internet riesig ist und wir häufig wahrscheinlich nur Dinge zurückbekommen, die es in der einen oder anderen Form bereits gibt und die daher nicht auf originellem Denken beruhen. Insgesamt ist Kambhampati kein großer Fan von „emergentem Verhalten“. Beim Aufbau eines IT-Systems möchten Sie genau wissen, was Sie aufgebaut haben.

Es ist an sich nicht lobenswert, wenn Ihr IT-System ein Verhalten zeigt, das Sie nicht erklären können. Ein KI-System, das für den Zugbetrieb zuständig ist, darf bei der Arbeit keine Halluzinationen erleiden und Menschen würden sich sicherlich unwohl fühlen, wenn physische Roboter in ihrer Umgebung plötzlich völlig unerwartete Dinge tun würden.

Auf dem Weg zu KI-Systemen mit eigenen Weltmodellen

Und Kambhampati ist nicht der Einzige, der etwas zurückhaltend ist, was die Frage angeht, wie weit man mit großen Sprachmodellen tatsächlich gehen kann. Laut Yann LeCun, KI-Forscher und Chief AI Scientist bei Meta, sind wir noch sehr weit von KI-Systemen entfernt, die auf menschenähnlicher Ebene schlussfolgern und planen können . Laut LeCun können Sprachmodelle die Welt (d. h. Texte) in einer Dimension voraussagen und verstehen, während KI-Modelle, die mit Bildern und Videos arbeiten, die Welt in zwei Dimensionen verstehen.

Doch um die Welt wirklich zu verstehen und denken zu können, muss man in der Lage sein, Simulationen in drei Dimensionen durchzuführen. Sie müssen über ein Weltmodell verfügen – ein „Weltmodell“ –, in dem Sie Aktionen in der Welt testen und simulieren können, bevor Sie tatsächlich etwas tun. Mit solchen Simulationen können Sie erklären, warum Sie eine bestimmte Handlung ausgeführt haben, und Sie können überrascht sein, wenn Sie etwas sehen, das Ihrem Weltbild widerspricht.

Natürlich müssen noch viele Probleme gelöst werden, bevor wir an den Punkt gelangen, an dem wir in einer Welt mit Robotern leben, die mit ihren eigenen Weltmodellen herumlaufen. LeCun weist beispielsweise darauf hin. zum Problem der Aufteilung eines Gesamtplans in kleine Teilpläne. Ja, wenn Sie von den USA nach Frankreich reisen, müssen Sie fliegen, aber Sie müssen auch einen Plan haben, wie Sie zum Flughafen gelangen, und vorher müssen Sie zu Hause in den USA aus dem Bett aufstehen. Selbst mit allen verfügbaren Ressourcen von Facebook ist es im Allgemeinen nicht so einfach, ein

Programm zu schreiben, das diese Art von Problemen lösen kann.

Allgemeine Intelligenz ist einfach nicht einfach. Es geht nicht nur darum, über mehr „enge Intelligenzen“ zu verfügen, sondern auch darum, diese zu einem Ganzen zu integrieren. Das menschliche Gehirn ist nicht nur mit Sprachverständnis und Sprachproduktion in den Wernicke- und Broca-Gehirnarealen ausgestattet, sondern verfügt auch über kreuz und quer verlaufende Verbindungen zu anderen Bereichen, wie etwa dem präfrontalen Kortex, der zur Planung und Entscheidungsfindung beitragen kann.

Vielleicht steht die Künstliche Allgemeine Intelligenz doch nicht unmittelbar bevor. Und viele Aussagen von Marketingabteilungen auf der ganzen Welt werden wahrscheinlich irgendwann im selben Licht gesehen werden, wie wir heute die eher optimistischen Aussagen der KI-Forscher Simon, Shaw und Newell aus dem Jahr 1957 sehen, als sie der Welt GPS (General Problem Solver) vorstellten, ein Computerprogramm, das damals im Jahr 1957 versprach, ein universeller Problemlöser zu sein.

Dennoch ist es erwähnenswert, dass Kambhampati zwar vielen Aussagen über große Sprachmodelle skeptisch gegenübersteht, es jedoch nie versäumt, die Stärken dieser Modelle zu loben, beispielsweise bei der Ideenfindung und vielem mehr. LeCun ist sich natürlich völlig bewusst, dass es noch ein langer Weg ist, bis wir KI-Systeme haben, die Simulationen in ihren Weltmodellen durchführen und dadurch die Welt auf menschenähnliche Weise verstehen können. Doch laut LeCun stehen immer noch Wege in eine zukünftige Welt offen, in der allen Menschen ein größerer Stab (mit AGI ausgestattete KI-Agenten) zur Verfügung steht, der nur ihnen hilft.

Laut LeCun gibt es keinen Grund zur Sorge darüber, wie eine solche Geschichte enden wird. Intelligente Agenten entstehen nicht von heute auf morgen. Wir müssen die Probleme im Verhältnis dazu sehen, wie weit wir tatsächlich gekommen sind. Wir brauchen ein wenig Vertrauen in den Fortschritt. Und es bleibt Zeit, an der guten Version der Zukunft zu arbeiten. Darüber hinaus möchten wir Menschen verstehen, was Intelligenz ist – sowohl die von Maschinen als auch unsere eigene. An der Angelegenheit muss noch gearbeitet werden, aber laut LeCun ist der Weg nach vorn gut genug.

Und wenn er das denkt, können wir anderen das doch sicher auch denken?

Der obige Beitrag ist Teil 2 der Reihe „Von KI zu AGI“ von Simon Laubs, außerordentlicher Professor an der Business Academy Aarhus. Den ersten können Sie hier lesen oder erneut lesen .

Möchten Sie mit Ihrem Standpunkt zur Debatte beitragen? Dann schreiben Sie an unsere Debattenredaktion unter debat@ing.dk

Dieser Artikel

Tech-Giganten an Trump: Lasst uns KI an urheberrechtlich geschütztem Material trainieren

Von der KI zur AGI im Jahr 2025: Zeigt die KI-Entwicklung der letzten Jahrzehnte einen großen Durchbruch oder wiederholt sich die Geschichte einfach?

Themen

Künstliche Intelligenz Folgen Roboter Folgen

Softwareentwicklung Folgen Daten Folgen

Ausgewählte Unternehmen

Facebook, OpenAI