

KI-Tool fördert den Dialog zwischen Menschen mit gegensätzlichen Meinungen

Ein KI-gestütztes Tool hilft Menschen mit unterschiedlichen Meinungen, gemeinsame Standpunkte zu finden und fördert so den Dialog.



Ein chatbotsähnliches Tool, das von Künstlicher Intelligenz (KI) unterstützt wird, kann Menschen mit unterschiedlichen Ansichten helfen, Bereiche der Übereinstimmung zu finden. Dies zeigt ein Experiment mit Online-Diskussionsgruppen.

Das Modell, das von Google DeepMind in London entwickelt wurde, konnte divergierende Meinungen synthetisieren und Zusammenfassungen der Position jeder Gruppe erstellen, die verschiedene Perspektiven berücksichtigten. Die Teilnehmer bevorzugten die von der KI generierten Zusammenfassungen gegenüber denen, die von menschlichen Mediatoren verfasst wurden. Dies deutet darauf hin, dass solche Tools genutzt werden könnten, um komplexe Beratungen zu unterstützen. Die

Studie wurde am 17. Oktober in der Zeitschrift Science veröffentlicht¹.

„Man kann es als einen Machbarkeitsbeweis sehen, dass man KI, insbesondere große Sprachmodelle, nutzen kann, um einen Teil der Funktion zu erfüllen, die derzeit von Bürgerversammlungen und deliberativen Umfragen erfüllt wird“, sagt Christopher Summerfield, Mitautor der Studie und Forschungsdirektor am UK AI Safety Institute. „Die Menschen müssen Gemeinsamkeiten finden, denn kollektives Handeln erfordert Zustimmung.“

Kompromissmaschine

Demokratische Initiativen wie Bürgerversammlungen, bei denen Gruppen von Menschen gebeten werden, ihre Meinungen zu politischen Themen zu teilen, stellen sicher, dass Politiker eine Vielzahl von Perspektiven hören. Die Ausweitung dieser Initiativen kann jedoch schwierig sein, da diese Diskussionen häufig auf kleine Gruppen beschränkt sind, um sicherzustellen, dass alle Stimmen gehört werden.

Neugierig auf die Forschung zu den Möglichkeiten von großen Sprachmodellen (LLMs), entwickelte Summerfield gemeinsam mit seinen Kollegen eine Studie, um zu bewerten, wie KI Menschen mit gegensätzlichen Meinungen dabei helfen könnte, einen Kompromiss zu erreichen.

Sie setzten eine feinabgestimmte Version des vortrainierten DeepMind LLM Chinchilla ein, das sie „Habermas-Maschine“ nannten, benannt nach dem Philosophen Jürgen Habermas, der eine Theorie entwickelte, wie rationale Diskussionen helfen können, Konflikte zu lösen.

Um ihr Modell zu testen, rekrutierten die Forscher 439 britische Einwohner, die in kleinere Gruppen aufgeteilt wurden. Jede Gruppe diskutierte drei Fragen zu britischen politischen Themen und teilte ihre persönlichen Meinungen dazu. Diese Meinungen wurden dann in die KI-Maschine eingespeist, die übergreifende

Aussagen generierte, die die Sichtweisen aller Teilnehmer kombinierten. Die Teilnehmer konnten jede Aussage bewerten und Kritiken dazu abgeben, die die KI dann in eine endgültige Zusammenfassung der kollektiven Sichtweise der Gruppe einarbeitete.

„Das Modell ist darauf trainiert, eine Aussage zu produzieren, die von einer Gruppe von Personen, die ihre Meinungen freiwillig geäußert haben, maximal unterstützt wird“, sagt Summerfield. „Da das Modell lernt, was Ihre Präferenzen über diese Aussagen sind, kann es dann eine Aussage erzeugen, die höchstwahrscheinlich alle zufriedenstellt.“

Zusätzlich zur KI wurde ein Teilnehmer als Mediator ausgewählt. Auch er wurde gebeten, eine Zusammenfassung zu erstellen, die die Ansichten aller Teilnehmer am besten einbezieht. Die Teilnehmer bekamen sowohl die Zusammenfassungen der KI als auch die des Mediators angezeigt und sollten diese bewerten.

Die meisten Teilnehmer bewerteten die von der KI verfassten Zusammenfassungen als besser als die des Mediators. 56 % der Teilnehmer bevorzugten die Leistung der KI, verglichen mit 44 %, die die menschliche Zusammenfassung bevorzugten. Externe Gutachter wurden ebenfalls gebeten, die Zusammenfassungen zu bewerten und gaben den KI-Zusammenfassungen höhere Bewertungen bezüglich Gerechtigkeit, Qualität und Klarheit.

Das Forschungsteam rekrutierte dann eine Gruppe von Teilnehmern, die demografisch repräsentativ für die britische Bevölkerung waren, für eine virtuelle Bürgerversammlung. In diesem Szenario stieg die Gruppenvereinbarung zu strittigen Themen, nachdem sie mit der KI interagiert hatten. Diese Erkenntnis legt nahe, dass KI-Tools, wenn sie in eine echte Bürgerversammlung integriert werden, es den Leitern erleichtern könnten, Politikvorschläge zu erarbeiten, die verschiedene Perspektiven berücksichtigen.

„Das LLM könnte auf viele Arten verwendet werden, um

Beratungen zu unterstützen und Rollen zu übernehmen, die zuvor menschlichen Moderatoren vorbehalten waren“, sagt Ethan Busby, der untersucht, wie KI-Tools demokratische Gesellschaften verbessern könnten, an der Brigham Young University in Provo, Utah. „Ich betrachte dies als die Spitze der Arbeiten in diesem Bereich, die großes Potenzial hat, drängende soziale und politische Probleme anzugehen.“ Summerfield fügt hinzu, dass KI sogar helfen könnte, Konfliktlösungsprozesse schneller und effizienter zu gestalten.

Verlorene Verbindungen

„Diese Technologien tatsächlich in deliberative Experimente und Prozesse anzuwenden, ist wirklich erfreulich“, sagt Sammy McKinney, der an der University of Cambridge, UK, die deliberative Demokratie und ihre Schnittstellen zur Künstlichen Intelligenz untersucht. Aber er fügt hinzu, dass Forscher die potenziellen Auswirkungen der KI auf den menschlichen Aspekt der Deliberation sorgfältig bedenken sollten. „Ein Hauptgrund, um Bürgerdeliberationen zu unterstützen, ist, dass sie bestimmte Räume schaffen, in denen Menschen miteinander in Beziehung treten können“, sagt er. „Was verlieren wir, wenn wir den menschlichen Kontakt und die menschliche Moderation zunehmend entfernen?“

Summerfield erkennt die Einschränkungen an, die mit KI-Technologien wie diesen verbunden sind. „Wir haben das Modell nicht darauf trainiert, in die Deliberation einzugreifen“, sagt er, was bedeutet, dass die Aussage des Modells auch extremistische oder andere problematische Überzeugungen enthalten könnte. Er fügt hinzu, dass rigorose Forschung über die Auswirkungen von KI auf die Gesellschaft entscheidend ist, um ihren Wert zu verstehen.

„Vorsichtig vorzugehen, scheint mir wichtig“, sagt McKinney, „und dann Schritte zu unternehmen, um diese Bedenken, wo möglich, zu mildern.“

1. Tessler, M.H. et al., Science 386, eadq2852 (2024).

Google Scholar

Literaturverzeichnis herunterladen

Details

Besuchen Sie uns auf: natur.wiki