

MEDIEN

ZWISCHEN INNOVATION UND HALLUZINATION

Wie künstliche Intelligenz Berichterstattung und Medienkonsum verändert

Claire Barthelemy

Künstliche Intelligenz schlägt große Wellen im Journalismus: Sie revolutioniert investigative Recherchen, deckt Fake News auf und vereinfacht Arbeitsabläufe. Doch populäre KI-Chatbots interpretieren journalistische Arbeit oft falsch oder liefern sogar erfundene Informationen. Welche Auswirkung hat das auf Journalist*innen und Leser*innen?

Diversität, mentale Gesundheit, Frauen – dies sind einige der fast 200 Begriffe, die Donald Trumps Regierung aus offiziellen Dokumenten streichen ließ. Das deckte eine Recherche der New York Times auf, bei der Journalist*innen Tausende von Regierungsmemos, öffentlichen Dokumenten und internen Richtlinien untersuchten. Mithilfe von Künstlicher Intelligenz analysierten sie Unmengen Screenshots von Regierungswebseiten, die vor und nach Trumps Amtseinführung gemacht wurden. Ein KI-Tool extrahierte die Texte aus den Bildern und ein Large Language Model (LLM) suchte anschließend nach Veränderungen. So zeichnete sich aus der Datenmenge das Bild eines ideologischen Wechsels. Eine Recherche, die ohne KI viel zeitaufwendiger und vielleicht sogar unmöglich gewesen wäre.

Hilfe beim Recherchieren

Dylan Freedman, ein Ingenieur für maschinelles Lernen und Mitglied des „A.I. Initiative Teams“ der New York Times stellte das Projekt auf der Konferenz „AI and the Future of News“ in London vor. Organisiert wurden die Diskussionsrunden von der Nachrichtenagentur Reuters. Hier wurden weitere Beispiele vorgestellt, bei denen KI im Newsroom benutzt wird. Ein gemeinsames Projekt der BBC und Oxford University kann zum Beispiel Deepfake-Fotos erkennen, während die Financial Times ein KI-Tool benutzt, das Artikel vorliest und den Leser*innen eine Zusammenfassung aus Stichpunkten anbietet.

„KI eröffnet tatsächlich eine völlig neue Kategorie des Journalismus, die wir zuvor gar nicht in Betracht ziehen konnten – ich spreche von Recherchen, die zehntausende Seiten

unstrukturierter Dokumente, hunderte Stunden Videomaterial oder sämtliche bundesstaatlichen Gerichtsakten umfassen“, sagte Zach Seward, Redaktionsleiter für KI-Initiativen bei der New York Times gegenüber dem New Yorker.

AI kann Journalist*innen beim Recherchieren, Überprüfen und Zusammenfassen helfen und könnte sogar eine neue Ära des Investigativjournalismus herbeiführen. Für die Nachrichtenverbreitung und das Leseverhalten stellen sich jedoch ganz andere Fragen. Chatbots wie ChatGPT und Deepseek, die auf LLMs basieren, sind für viele Menschen keine Neuheit mehr. In Großbritannien nutzen bereits 40 Prozent der Bevölkerung LLMs, laut einer Studie des KI-Forschungszentrums Ada Lovelace Institute.

„Das ist ein plötzlicher und bedeutender Wandel gegenüber zwei Jahren, als es diese Produkte noch überhaupt nicht gab“, so Andrew Strait, der als Ethik- und Politikforscher bei Deepmind, Googles KI-Forschungslabor, arbeitete. Er ist nun Leiter für Gesellschaftliche Resilienz im AI Security Institute, einer Sparte des britischen Ministeriums für Wissenschaft, Innovation und Technologie. „Ich glaube, wir sehen Zeichen, dass traditionelle Suchmaschinen ersetzt werden“, sagte er. Dabei müsste man bedenken, dass sich die Technologien von LLMs und Suchmaschinen stark voneinander unterscheiden, so Strait. „LLMs sind aufgrund ihrer Struktur und ihres Designs „stochastisch und probabilistisch“, so der Forscher. „Das heißt, wenn Sie hundertmal die gleiche Frage stellen, erhalten Sie 100 leicht unterschiedliche Versionen der Antwort.“ Dies sei zum Beispiel problematisch bei medizinischen Fragen, die rund 7 Prozent der täglichen Google-Suchen ausmachen.

Doch auch auf Fragen über aktuelle Nachrichten erwarten Benutzer akkurate Antworten. Die BBC kam in einer Studie über KI-Assistenten und Nachrichteninhalte zu erschreckenden Ergebnissen. Mehr als die Hälfte der KI-Antworten auf Fragen zu Nachrichten hatten „erhebliche Probleme“, während 19 Prozent der Antworten,

die BBC-Inhalte zitierten, falsche Aussagen, Daten und Zahlen angaben. In manchen Fällen erfand das KI-Modell sogar Zitate, die in den jeweiligen Artikeln überhaupt nicht vorkamen. Forscher nennen das Phänomen KI-Halluzinationen.

Wenn ein KI-Modell sich bei einer Frage nicht sicher ist, „füllt es Lücken“ anhand von Kontexten aus seinen Trainingsdaten, so die ForscherInnen Anna Choi und Katelyn Xiaoying Mei in einem Artikel für die Website The Conversation. Der Grund dafür können unausreichende oder verzerrte Trainingsdaten sein. Genau wie menschliche Halluzinationen, nehme das System dann Dinge wahr, die nicht existieren. Dies sei eine der größten Herausforderungen der Technologie. „Halluzinationen werden uns erhalten bleiben“, sagte Daniel Ho, ein Professor der Stanford University gegenüber dem Magazin Wired. „Wir haben noch keine fertigen Methoden, um diese wirklich zu eliminieren“, so Ho weiter.

Halluzinierte Schlagzeilen

Martin Bernklau, ein Journalist aus Tübingen, wurde Opfer einer solchen Halluzination. Als er seinen Namen und Wohnort bei Copilot, dem KI-Chatbot von Microsoft angab, lieferte das KI-Modell erschreckende Resultate: Es beschrieb den Journalisten unter anderem als Kinderschänder und Witwenbetrüger. Bernklau war in der Tat öfters im Landgericht Tübingen – und zwar als Gerichtsreporter. Das KI-Modell machte aus dem Berichterstatte also einen Verurteilten.

Doch es sind nicht nur Chatbots, die unter diesen Halluzinationen leiden. Im Dezember beschwerte sich die BBC beim Tech-Giganten Apple, nachdem eine Nachrichtenzusammenfassung von Apple News folgende Meldung verkündete: „Luigi Mangione shoots himself.“ Diese Nachricht stand auf iPhone Bildschirmen neben dem Logo der BBC. Die Meldung stimmte natürlich nicht – Luigi Mangione ist zwar Hauptverdächtiger im Mordfall um den Chef einer Krankenversicherung, er selbst lebt jedoch noch.

Die Zusammenfassungen wurden von Apple Intelligence erstellt, einem



Während Journalist*innen recherchieren und versuchen, faktisch richtige Artikel zu schreiben, werden Nachrichten durch KI-Tools oft falsch wiedergeben. Wofür werden die Leser*innen sich am Ende entscheiden?

Tool, das Verbesserungsbedarf habe, wie die Firma betonte. „Die Apple Intelligence-Funktionen befinden sich in der Betaphase und wir arbeiten kontinuierlich mit der Hilfe von Nutzerfeedback an Verbesserungen“, so der Tech-Gigant in einem Statement gegenüber der BBC.

Doch die britische Journalistengewerkschaft National Union of Journalists (NUJ) forderte von Apple eine vollständige Entfernung der KI-Funktion. „In einer Zeit, in der der Zugang zu akkurater Berichterstattung wichtiger denn je ist, darf die Öffentlichkeit nicht in die Lage versetzt werden, an der Richtigkeit der Nachrichten zweifeln zu müssen“, so Laura Davison, Generalsekretärin des NUJ im Guardian.

In der Tat sind Falschnachrichten und Deep Fakes immer schwieriger zu erkennen. Während in Redaktionen der großen Medienhäuser KI-Tools unter Aufsicht von ausgebildeten Journalist*innen benutzt werden, um Fake News zu erkennen und aus riesigen Datenmengen sinnvolle Informationen herauszufiltern, verlieren Nachrichteninhalte in den Händen der Tech-Giganten ihre Glaubwürdigkeit. Wenn sich die Nutzer*innen der Chatbots dieser Schwächen der KI-Tools jedoch bewusst sind, könnte das dem Journalismus wiederum nützen.

„Während sich der Rest des Internets mit KI-generiertem Mist füllt und es immer schwieriger wird, die Herkunft des Gelesenen zu erkennen, wird es immer wertvoller sagen zu können: ‚Dies wurde von den Reportern berichtet und geschrieben, deren Gesichter man in der Autorenzeile sieht‘“, sagte Zach Seward dem New Yorker. Die Last der Überprüfung werde dabei nie auf Leser*innen abgewälzt, so Seward. Darin ist die alte Profession des Journalismus der modernen Technologie der KI-Chatbots weit voraus.

FOTO: CLAIRE BARTHELEMY