Vortrag „KI – nachahmen oder verstehen?“
18.11.2024, Prof. Dr. Martin Butz

Zusammenfassung der Fragerunde

Frage:
Eine KI zum Wohle aller einzusetzen ist zweifellos ein äußerst ambitioniertes Ziel. Ich wünsche mir, dass dies eines Tages gelingt. Doch die zentrale Frage, die sich für mich stellt, lautet: Was können wir schon jetzt tun? Welche Kompetenzen müssen wir bereits heute besitzen, entwickeln und fördern, um mithilfe von KI effektiver und schneller zu werden, als wir es als Menschen ohne diese Technologie könnten?

Butz:
Effektiver und schneller zu werden ist das Stichwort. Mit den aktuellen KI-Systemen können wir bereits gut arbeiten, indem wir sie nutzen, um Aufgaben schneller zu erledigen, Denkanstöße zu erhalten oder komplexe Zusammenhänge zu analysieren. Ich bin jedoch überzeugt, dass diese Systeme insbesondere in Bereichen wie Governance, Politik, Wirtschaft und Analysen noch viel weiter und positiver eingesetzt werden können.
Vor allem die nächste Generation von KI-Systemen, deren Entwicklung ich in naher Zukunft erwarte, könnte uns helfen, unsere Herausforderungen tiefer und umfassender zu durchdenken, als wir es derzeit vermögen. Bereits die bestehenden Systeme könnten intensiver genutzt werden, beispielsweise um Fragen zu beantworten wie: Welche Auswirkungen hätte eine bestimmte Richtlinie auf verschiedene Bevölkerungsgruppen? Wie würde diese wahrgenommen werden? Welche negativen Nebeneffekte könnten auftreten, und wie lassen sich diese von Anfang an minimieren?

Es wäre wichtig, solche Überlegungen nicht zu ignorieren, sondern aktiv in den Entscheidungsprozess einzubinden. Ich sehe eine große Chance darin, KI zu nutzen, um systematisch voranzukommen, anstatt weiterhin das Stückwerk zu verfolgen, das meiner Meinung nach momentan dominiert.

Nachfrage:
Entschuldigung, dass ich das zuvor mit „schneller und effektiver“ so unklar formuliert habe. Meine eigentliche Frage lautet: Wie können wir sicherstellen, dass der Mensch im Mittelpunkt bleibt, während wir KI als Werkzeug nutzen? Dabei geht es darum, dass die Ergebnisse, die wir durch KI erzielen, weder anderen Menschen noch uns selbst schaden und zugleich nachvollziehbar und plausibel sind.

Butz:
Das ist eine sehr schwierige Frage. Ich frage mich selbst, wie gut die aktuellen Systeme

das bereits können. Mit großer Vorsicht würde ich sagen, dass es in manchen Bereichen noch schwierig ist. Dennoch denke ich, dass auch die heutigen KI-Systeme bereits positiv eingesetzt werden können, um beispielsweise negative Auswirkungen von Entscheidungen – sei es im politischen, wirtschaftlichen oder investitionsbezogenen Kontext – abzuwägen.

Diese Systeme könnten zumindest dazu beitragen, bestimmte Aspekte zu thematisieren, etwa welche positiven und negativen Effekte Entscheidungen haben könnten. Ganz ausschließen lässt sich der negative Teil jedoch nie. Mit jeder Veränderung, die in der Welt vorgenommen wird, gibt es zwangsläufig Menschen, die davon Nachteile erfahren – insbesondere diejenigen, die vom Status quo maximal profitieren.

Es gibt also immer eine positive und eine negative Seite, das befürchte ich.

Frage:
Wenn ich Sie richtig verstanden habe, haben Sie gesagt, dass die Daten allein nicht ausreichen, sondern Sie fordern, dass KI-Systeme mit einer klaren Struktur ausgestattet werden müssen. So wie wir innerhalb des Gehirns eine Struktur finden, benötigen wir diese ebenso in KI-Systemen. Wie könnte eine solche Struktur aussehen? So wie es biologisch inspirierte Architektur gibt, könnte man ja zum Beispiel auch von kognitiv inspirierter KI sprechen.

Butz:
Ich versuche, das noch etwas genauer zu erläutern. Die Struktur, die ich meine, umfasst eine hierarchische Organisation des kognitiven Systems, wobei diese Hierarchie flexibel und nicht starr sein sollte. Es muss möglich sein, verschiedene kontextuelle Räume dynamisch zu aktivieren. Beispielsweise kann ich über meinen nächsten Urlaub nachdenken, gleichzeitig jedoch auch Ihre Frage beantworten oder mich entscheiden, mir ein Glas Wasser einzuschenken. Diese unterschiedlichen Kontexte müssen flexibel aufgerufen und verarbeitet werden können.

In aktuellen KI-Systemen, etwa bei großen Sprachmodellen (*large language models*), spricht man oft von „*in-context learning*“. Dabei wird der Kontext jedoch meist durch eine vorab eingeführte Datenbasis, etwa einen Artikel, definiert, der dann als Grundlage dient. Dies ist jedoch weit von der Flexibilität entfernt, die wir Menschen besitzen.

Menschen verfügen über eine extrem anpassungsfähige Kontextualisierung. Wir können innerhalb kürzester Zeit zwischen sehr detaillierten Aufgaben – wie dem Berücksichtigen kleiner Aspekte – und der Betrachtung des großen Ganzen wechseln. Diese Fähigkeit, Kontexte dynamisch zu wechseln und die Aufmerksamkeit flexibel auszurichten, stellt eine der größten Herausforderungen für aktuelle KI-Systeme dar.

Diese Art der Flexibilität wird bisher kaum in künstlichen kognitiven Architekturen berücksichtigt. Auch in der Neurowissenschaft, Psychologie und den kognitiven Wissenschaften ist unser Verständnis darüber noch begrenzt.

Frage:
In den letzten Jahren, in denen die Entwicklung der KI relativ dynamisch verlief, hat sich einiges verändert. KI ist ja erst in den letzten eineinhalb bis zwei Jahren im Mainstream wirklich angekommen. Früher hieß es oft: „Content is king." Heute könnte man aber sagen: „Context is king."
Meine Frage bezieht sich darauf, welchen Einfluss Kontext auf das sogenannte Prompting hat. Wenn KI „nur nachahmt", wie Sie es ausdrücken, wird der Kontext dann nicht umso wichtiger?

Mir scheint, dass ich eine bessere und valide Antwort erhalte, wenn ich den Fokus klar setze und die Frage strukturiert stelle. Gibt es konkrete Strategien oder Empfehlungen, um diese Kontextualisierung gezielt zu fokussieren und so bessere Ergebnisse bei der Nutzung von KI-Systemen zu erzielen?

Butz:
Es gibt bereits eine Vielzahl an Artikeln und Ansätzen – oft unter dem Begriff „Prompt Engineering" – die erklären, wie man den Kontext in *large language models* (LLMs) gezielt setzen und einschränken kann. Diese Kontextualisierungen ermöglichen es beispielsweise, KI-Systeme besser auf spezifische Aufgaben oder Unternehmenskontexte auszurichten.

Ein häufiger Ansatz ist, dass firmeninterne Datenmengen oder spezifische Vorgaben als Pre-Prompt eingebettet werden, bevor Mitarbeitende mit dem System interagieren. So wird das Modell auf einen bestimmten Kontext „fokussiert", ohne dass dies für die NutzerInnen direkt sichtbar ist.

Allerdings möchte ich betonen, dass die sprachlich umgesetzte Kontextualisierung in solchen Systemen wenig mit der Art von Kontextualisierung gemein hat, die in menschlichen Köpfen stattfindet. Unsere mentale Kontextualisierung strebt nach Kohärenz und Kausalität. Sie erlaubt es uns, komplexe Referenzrahmen zu schaffen, innerhalb derer wir kausal argumentieren, planen und Alternativen abwägen. Aktuelle KI-Systeme hingegen beschränken sich darauf, sprachliche Muster und Assoziationen in einem gegebenen Kontextraum zu wiederholen. Sie schaffen es nicht, diese Kontexte auf einer tieferen, kausalen Ebene zu durchdringen oder tatsächlich „nachzudenken" – beispielsweise Szenarien zu modellieren, Entscheidungen zu analysieren oder Hypothesen zu überprüfen.

Es fehlt ihnen die Fähigkeit, innerhalb eines dynamischen, kausalen Referenzrahmens zu operieren, wie es Menschen tun, wenn sie Argumente strukturieren oder

Entscheidungen treffen. Dies ist eine der größten Hürden, die heutige KI-Systeme noch nicht überwinden können.

Frage:
Ein Wort, das heute Abend nicht gefallen ist, aber mir dennoch durch den Kopf ging, ist „Kreativität“. Vielleicht bin ich froh, dass es nicht erwähnt wurde. Heute Morgen las ich in der Zeitung einen Artikel über Künstliche Intelligenz (KI), die Gedichte schreibt. Darin wurde behauptet, diese Gedichte seien teilweise besser als die von Shakespeare.

Natürlich stellt sich die Frage, was ein „gutes“ Gedicht ausmacht – das müsste erst einmal definiert werden. Mein Eindruck des Artikels war jedoch, dass ein Gedicht dann besser sei, wenn es leichter verständlich ist als eines von Shakespeare.

Mich beschäftigt dabei die Frage: Gibt es Grenzen, an die KI stoßen wird und die sie nicht überschreiten kann? Beispielsweise, wenn es darum geht, etwas zu schreiben wie Franz Kafka oder KünstlerInnen des 20. und 21. Jahrhunderts, die über Grenzen hinausgingen. Kann KI auf solche Ideen kommen? Oder stößt sie an eine Art logische Grenze, die nicht überwindbar ist?

Butz:
Das ist eine spannende Frage. Ich denke, hier gibt es zwei Aspekte, die man unterscheiden muss.

Der erste betrifft das Menschliche – unsere emotionale und soziale Wahrnehmung. Nehmen wir Kafka: Seine Werke sind geprägt von seiner emotionalen Verfassung, seiner familiären Situation, seiner Einsamkeit und seiner Selbstwahrnehmung als Belastung für andere. All diese menschlichen, emotionalen Eigenschaften spiegeln sich in seinen Texten wider. Diese Art von Emotionalität und Menschlichkeit kann KI aktuell höchstens imitieren, wie es etwa durch *large language models* geschieht. Aber diese Systeme können keine echte Kreativität oder Menschlichkeit erzeugen.

Unser emotionaler Apparat ist tief in unseren Gehirnstrukturen verankert, genetisch vorgegeben und ein wesentlicher Bestandteil unseres Daseins. KI-Systeme sind von solchen emotionalen Grundlagen weit entfernt – und ich denke, sie sollten auch nicht unbedingt versuchen, diese nachzubilden. Warum sollten wir eine KI schaffen, die einem Menschen gleicht? Das sehe ich nicht als Ziel.

Die Nachahmung durch KI wird sich immer darauf beschränken, sprachliche und konzeptionelle Muster zu reproduzieren. Sie mag Inhalte generieren, die leichter zugänglich erscheinen, und wird vielleicht als „besser“ empfunden, weil sie verständlicher sind. Aber diese Werke drücken keine echte Emotion oder Menschlichkeit aus.

Der zweite Aspekt betrifft die rationale und logische Ebene. Hier sehe ich erhebliches Potenzial. KI-Systeme können Schlussfolgerungen ziehen, Einflüsse abwägen und komplexe soziale, wirtschaftliche oder kulturelle Systeme auf einer rationalen Ebene durchdringen. In diesem Bereich könnten sie uns sogar übertreffen, da ihre Kapazitäten die des menschlichen Gehirns übersteigen können.

Ein Beispiel ist das Go-Spiel. Jahrhunderte lang haben Menschen Techniken entwickelt und darüber meditiert, um in diesem Spiel besser zu werden. Doch eine KI hat Go so tief durchdrungen, wie es keinem Menschen möglich war.

Ähnlich könnten KI-Systeme eines Tages gesellschaftliche, wirtschaftliche oder kulturelle Realitäten analysieren und strukturieren – allerdings mit der Einschränkung, dass diese Systeme weit komplexer sind als ein Spiel wie Go, das inhärent strukturiert und daher einfacher zugänglich ist.

Sollte es jedoch gelingen, solche Systeme zu entwickeln, könnten sie als beratende Instanzen fungieren. Sie könnten rationale und tiefgehende Analysen bieten, um uns bei der Bewältigung globaler Herausforderungen zu unterstützen. Dabei wäre es entscheidend, ihnen keine Autonomie zu geben, sondern sie ausschließlich als Werkzeuge einzusetzen. Ihr Beitrag könnte darin bestehen, potenzielle Auswirkungen von Entscheidungen aufzuzeigen – wer positiv oder negativ betroffen wäre, wie man negative Effekte abmildern könnte und so weiter.

Das wäre meine idealistische Vision: KI als beratendes, abwägendes System, das uns hilft, die Welt besser zu verstehen und gerechter zu gestalten.

Nachfrage:
Wenn ich es richtig sehe, sind Sie ja auch Psychologe. Glauben Sie, dass die Möglichkeiten der Psychologie und die Möglichkeiten der Künstlichen Intelligenz konvergieren? Ich meine, Psychologen entdecken immer mehr, wie der Mensch funktioniert, und können damit rational umgehen. Das bedeutet, dass wir, wenn wir die Künstliche Intelligenz entsprechend weiterentwickeln, möglicherweise psychologische Phänomene, die wir bisher nicht erklären konnten, irgendwann mit KI erklären können.

Butz:
Ja, das geschieht bereits in gewissem Maße. Ein aktuelles Modell zeigt beispielsweise, dass eine kontextualisierte Simulation nicht nur kognitive Energie spart, sondern auch menschliches Verhalten in typischen Veränderungen, Multitasking und ähnlichen Aufgaben beeinflusst. In diesen Fällen kann man gut demonstrieren, wie kontextualisiertes und habituelles Verhalten mit aufgabenspezifischem Verhalten in Konflikt steht.
In dieser Richtung tut sich viel, und es wird sich sicherlich weiterentwickeln.

Frage:
Es wurde gezeigt, dass KI-Modelle, wenn sie mit ihren eigenen Daten oder mit Daten trainiert werden, die andere KIs generiert haben, deutlich schlechtere Leistungen erbringen. Meine Frage ist zweigeteilt: Erstens, was können wir dagegen tun, wenn das Internet, das einen großen Teil der Trainingsdaten für KI-Modelle darstellt, immer weiter mit KI-generierten Inputs geflutet wird? Zweitens, wie wirkt sich dies auf die menschliche Fähigkeit aus, neue Dinge zu generieren, wenn wir als Menschen nur noch Inputs von KI-generierten Quellen erhalten und diese häufiger sehen werden?

Butz:
Das ist eine sehr spannende und wichtige Frage. Manche Kollegen empfehlen, einen Snapshot aller Daten zu erstellen, bevor die KI beginnt, die Daten zu beeinflussen und selbst zu augmentieren. Im Grunde ist es wie ein genetischer Drift: Wenn das System ohne Bewertung das gesamte Datensystem mit seinen eigenen Daten flutet, werden wir einen Drift erleben, und das System wird nicht mehr gerichtet sein. Die Frage ist, wie stark wir Menschen in der Lage und auch bereit sind, kritisch zu hinterfragen, was uns ein großes Sprachmodell produziert hat.
Ich selbst musste vor drei Tagen einen Abstract schreiben und habe ihn zunächst selbst verfasst. Dann gab ich ihn einem großen Sprachmodell und bat um eine Überarbeitung, um den Text zugänglicher und spannender zu machen. Zunächst fand ich die Überarbeitung deutlich besser, flüssiger und netter. Nach genauerem Hinsehen stellte ich jedoch fest, dass der Text aber auch viel oberflächlicher geworden war. Obwohl er besser lesbar war, entsprach er nicht meinem Anspruch. Am Ende habe ich diesen Text gar nicht verwendet.

Das Problem ist, dass viele Menschen den Text nur überfliegen, ihn netter finden und sich nicht mehr damit beschäftigen müssen, was dazu führt, dass die Qualität der Ergebnisse durch die großen Sprachmodelle beeinträchtigt wird. Man sieht auch, dass man mit reinen Daten und mehr Rechenleistung nicht viel mehr aus diesen Systemen herausholen kann. Es gibt einfach nicht mehr sprachliche Daten; die *large language models* BetreiberInnen haben bereits alle Texte der Welt genutzt.
Andererseits müssen wir berücksichtigen, wie stark wir von dieser ganzen Geschichte kompetitiv beeinflusst werden. Der Einfluss durch Interessengruppen, die gezielt falsche Informationen streuen oder Informationen, die ihnen passen und anderen nicht, ist ein fundamentales Problem. Diese Manipulation wird durch KI optimiert, und es ist unklar, wie genau dies stattfindet und an welche Gruppen diese Informationen gestreut werden, um sie weiter zu verbreiten. Ich denke, das ist das größte Problem, das wir derzeit haben.