

Große Sprachmodelle

Whitepaper

Löser, A., Tresp, V. et al.
AG Technologische Wegbereiter
und Data Science

Kurzfassung



Sie schreiben Hausarbeiten, Gedichte oder Programmiercodes – große Sprachmodelle wie ChatGPT, BARD oder BLOOM verändern mit einer rasanten Entwicklung, wie wir arbeiten und kommunizieren und wie wir mit Information und Wissen umgehen. Diese KI-Modelle sind flexibler und leistungsfähiger als ihre Vorgänger. Als Schlüsseltechnologie sind sie der Kern vieler wichtiger Anwendungen: Sie erkennen, produzieren, übersetzen und verarbeiten Sprache. Die zugrundeliegende KI-Technologie ist jedoch nicht auf Sprachverarbeitung beschränkt. Große Sprachmodelle bieten daher enormes Potenzial für alle Lebensbereiche und stoßen einen grundlegenden und dauerhaften Wandel nicht nur in vielen Branchen an, sondern auch in Forschung und Entwicklung.

Expertinnen und Experten der Arbeitsgruppe Technologische Wegbereiter und Data Science der Plattform Lernende Systeme geben mit dem Whitepaper einen Überblick, auf welchen Grundlagen große Sprachmodelle beruhen, was ihre besonderen Eigenschaften sind und welche Handlungsfelder sich dabei für Forschung und Entwicklung ergeben. Dabei werden vor allem Chancen, Herausforderungen sowie Perspektiven der Forschung in Deutschland aufgezeigt.

Übersicht zu Sprachmodellen am Beispiel ChatGPT:

ChatGPT kann ...

- *Natürlich klingende Konversationen führen*
- *Texte aller Art erstellen (Aufsätze, Gedichte, Zusammenfassungen, Kochrezepte etc.)*
- *Gewünschten Stil imitieren (sachlich, poetisch etc.)*
- *Programmcode generieren*
- *Texte übersetzen u. v. m.*

basiert auf ...

- *Allgemeinen Textdaten aus dem Internet (Wikipedia etc.)*
- *Büchern*

funktioniert ...

- *Auf Basis von Wahrscheinlichkeiten*
- *Mittels KI-Algorithmen und mit aufwändigem Training unter Nutzung menschlicher Bewertungen von generierten Inhalten*

Grenzen

- *Liefert immer eine Antwort (auch wenn die Datenbasis nicht ausreichend Informationen zur Anfrage enthält)*
- *Erfindet bisweilen Inhalte oder Quellen („halluziniert“)*
- *Und weitere ...*

Quelle: Eigene Zusammenstellung.

Perspektiven aus Forschung und Entwicklung

Für Forschung und Entwicklung ergibt sich durch die rasante und disruptive Entwicklung und Verbreitung von großen Sprachmodellen ein dynamisches Handlungsfeld mit vielen Potenzialen, Forschungsfeldern und -bedarfen, aber auch Herausforderungen. Die Forschung aus Deutschland ist an vielen Stellen in die Entwicklung großer KI-Modelle engagiert und leistet einen gewichtigen Beitrag beispielsweise zu Themen wie Sprachanwendungsdienste nach europäischen Standards, Design und Augmentierung für medizinische Sprachmodelle, Erklärbarkeit, Benchmarks zur Evaluation und vielem mehr. Da für die deutsche Sprache bisher nur wenige große Sprachmodelle existieren, sollte die Forschung und (Weiter-)Entwicklung von modernen, offenen multilingualen Modellen sowie Modellen für die deutsche Sprache in Deutschland und Europa vorangetrieben und unterstützt werden. Damit ergeben sich bedeutende Forschungsfelder:

- **Vortrainierte große Sprachmodelle für die deutsche Sprache:** Auch für kleinere Sprachgemeinschaften ist es sinnvoll, große Sprachmodelle zu entwickeln, so auch für die deutsche Sprache. Insbesondere für sensible Domänen wie Medizin, Justiz oder Sicherheit ermöglicht die Weiterentwicklung eine Umsetzung unter Berücksichtigung des deutschen und europäischen Rechts sowie hiesiger Werte.
- **Multimodalität:** Große Sprachmodelle profitieren von zusätzlicher Kontextinformation, wie sie durch Multimodalität integriert werden kann, das heißt, wenn jenseits von Textdaten auch auf Bilder, Ontologien, Mengen, Tabellen, zeitvariante Daten, Bewegungstrajektorien etc. zurückgegriffen wird.
- **Multilingualität:** Multilinguale Modelle (Trainingsdaten stammen aus mehreren Sprachfamilien) können fehlendes Kontextwissen in der einen durch vorhandenes Kontextwissen in der anderen Sprache unter Umständen kompensieren. Das ist besonders hilfreich in Domänen mit einer globalen Nomenklatur bzw. einem globalen Standard oder Grounding (zum Beispiel Medizin).
- **Kombinierte KI bzw. hybride KI:** Bedeutende Fortschritte konnten bei Sprachmodellen durch die Kombination von mehreren Methoden erreicht werden, die unter anderem zusätzliche Kontexte zur Verfügung stellen. Potenzial liegt in der Kombination von wissensbasierter KI und Sprachmodellen, etwa wenn Wissensgraphen zur schnelleren Anpassung an Fakten und Gegebenheiten einbezogen werden.
- **Kontext in langen Texten verstehen:** Typische Aufgaben von Sprachmodellen sind es, Informationen übergreifend zu verdichten und Zusammenhänge zu erkennen. Daher sind Sprachmodelle für bestimmte Domänen der deutschen Sprache, die besonders lange Texte und Kontexte, beispielsweise aus den Prompts, gut verarbeiten können, interessant.
- **Erklärbarkeit:** Die Forschung zur Erklärbarkeit von KI zielt darauf ab, nachzuvollziehen, wie ein Modell zu seinen Ergebnissen kommt, was bei der Verarbeitung der Daten passiert und, ob Ergebnisse beispielsweise auf unerkannten Bias oder Korrelationen beruhen. Mit dieser Methode können die Modelle verbessert und weiterentwickelt sowie das Vertrauen der Nutzenden in KI gestärkt werden.
- **Differenzierbare Tokenizer:** Tokenizer sind ein Verfahren zur Zerlegung von Zeichenketten und Multi-Wörtern in logische und zusammengehörige Einheiten und sind daher wichtig als Grundlage für große Sprachmodelle. Sie sind stark abhängig von der Domäne und erfolgen momentan oft regelbasiert, was die Anpassung an Domänen erschwert.

- **Benchmarks:** Zur Evaluation großer Sprachmodelle werden in der Forschung sogenannte Benchmarks verwendet, um etwa ihre Grenzen sowie die Emergenz von Fähigkeiten untersuchen zu können oder sie hinsichtlich ihrer Leistung bei bestimmten Aufgaben zu testen: beim logischen Schlussfolgern, dem Lösen mathematischer Aufgaben oder dem Beantworten von Fragen. Für zahlreiche wichtige Domänen für die deutsche Sprache (Gesundheits-, Industrie-, Dienstleistungs- wie Bildungssektor, Maschinen- und Fahrzeugbau oder das Rechtswesen) fehlen solche Benchmarks.

Aus den Limitationen solcher Modelle ergeben sich zudem auch weitere Forschungsfelder, die es in den Blick zu nehmen gilt, um das **Sprachverstehen** solcher Modelle zu verbessern: beispielsweise **multiple Texte** korrekt in eine **Beziehung zueinander zu stellen**, das **Verständnis von Zeit** bzw. zeitlichen Abfolgen oder **Wissen über Wissen** bzw. Wissen über Nicht-Wissen sowie **Modularität**, also die Trennung von Wissen über Sprache und Schlussfolgerungen einerseits und Faktenwissen andererseits, und vieles mehr (siehe hierzu auch Goldberg, Y. 2023)¹.

Daneben sind auch **Kontext und Grounding** sowie die **Erkennung und die Behebung von Bias** stärker zu berücksichtigen. Denn angesichts der Tatsache, dass Sprachmodelle aktuell häufig selbstreferentiell und ohne Grounding sind, bedeutete dies, dass ausschließlich die Sprache (also lediglich eine Modalität) und die Referenzen der Wörter aufeinander dazu dienen, die tatsächliche Situation zu beschreiben. Um die menschliche Sprache aber besser zu verstehen und zu generieren, könnten weitere Modalitäten jenseits von Texten einbezogen werden. Tatsächliches Grounding im engeren Sinn wird allerdings erst durch den Bezug auf die Wirklichkeit im Verbund mit Wahrnehmung und Handlung erreicht werden.

Bei der Erkennung von Bias in Textsammlungen für das Training von großen Sprachmodellen sowie in den Modellen selbst, stehen der Forschung verschiedenste Methoden zur Verfügung, um Vorurteile zu beheben bzw. zu minimieren: So kann beispielsweise durch die Anpassung eines vortrainierten Modells an Textsammlungen zu spezifischen Online-Communities in sozialen Medien, Bias identifiziert werden. Ein solches Vorgehen kann bereits vor dem Einsatz eines generativen Sprachmodells für Dialoge dessen möglichen Bias beheben. Weiterhin lassen Unterschiede zwischen den Modellen Rückschlüsse auf die Texte zu, mit denen diese trainiert wurden, wodurch versteckter Bias einer Textsammlung ermittelt werden kann.

¹ Goldberg, Y. (2023): Some remarks on Large Language Models.
Online unter: <https://gist.github.com/yoavq/59d174608e92e845c8994ac2e234c8a9> (abgerufen am 21.01.2023)

Handlungsfelder für die Forschung

Die Forschung in Deutschland befindet sich in einer guten Ausgangsposition, um die Herausforderungen bei der (Weiter-)Entwicklung großer Sprachmodelle anzugehen, sie im Sinne europäischer Werte zu entwickeln und so ihre Potenziale für Deutschland und Europa auszuschöpfen. Dennoch darf sie in diesem dynamischen Forschungsfeld, das stark durch außereuropäische Akteure und große Technologieunternehmen geprägt ist, nicht den Anschluss verlieren. Auch wenn deutsche Forscherinnen und Forscher an vielen Stellen an der Entwicklung großer Sprachmodelle beteiligt sind, gilt es die dringlichsten Forschungsbedarfe anzugehen und so die Grundlagen für die Ausschöpfung des wirtschaftlichen und gesellschaftlichen Potenzials dieser KI-Technologie für Deutschland und Europa auszubauen – im Sinne europäischer Werte. Voraussetzung dafür ist ein europäisches Ökosystem für die Entwicklung großer Sprachmodelle sowie eine anwendungsnahe Forschung, die den Transfer in die industrielle Praxis erleichtert:

- **Gestaltungsoptionen für eine anwendungsfördernde Forschung:** Um die Transformation für die industrielle Anwendung moderner Sprachmodelle im Rahmen einer gezielten Transferforschung und -förderung effizient und schnell zu gestalten, sollte die Automatisierung der Prozesse zur Erstellung, Anpassung und Wartung dieser Modelle verstärkt in den Fokus der Forschung rücken. Ein wichtiger Bedarf liegt konkret in Modellen und Methoden zur kostengünstigen Anpassung von Sprachmodellen. Die Forschung sollte entsprechend sowohl an einem modell-zentrischen als auch an einem daten-zentrischen Fokus ausgerichtet sein.
- **Zugang zu öffentlich nutzbaren, deutschsprachigen Trainingsdaten:** Trainingsdaten sollten einer breiten Verteilung folgen und viele Anwendungsfälle abdecken – von domänenspezifischen Aufgaben (wie juristischen Dokumenten und Anforderungsanalysen) bis hin zu umgangssprachlichen Freiform-Dokumenten.
- **Evaluation und Testen von großen Sprachmodellen:** Sprachmodelle sollten im Hinblick auf Leistungsfähigkeit, Robustheit, Bias und Ressourcenverbrauch usw. getestet werden. Die systematisierte Sammlung solcher Tests für Sprachmodelle kann als Grundlage für eine Zertifizierung großer Sprachmodelle dienen.
- **Vernetzung und Stärkung der Community zu großen KI-Modellen:** Um die Entwicklung großer Sprachmodelle über Modalitäten hinweg im Sinne europäischer Werte und Normen sowie anwendungsnahe Forschung, die den Transfer in die industrielle Praxis erleichtert, voranzutreiben, ist ein europäisches Ökosystem notwendig, bei dem die interdisziplinäre Zusammenarbeit im Mittelpunkt steht: Dies um die technischen Herausforderungen der KI-Technologie anzugehen, das wirtschaftliche Potenzial voll auszuschöpfen und an die großen Akteuren aus den USA und China aufzuschließen.

Zusammenfassung – Chancen nutzen und Herausforderungen begegnen

Chancen nutzen

- Anpassbarkeit als besondere Eigenschaft dieser Modelle nutzen, indem effiziente Methoden entwickelt werden, die den Aufwand der Domänenanpassung erleichtern und damit den Transfer in die Anwendung: Anwenden von KI-Modellen auf neue Domänen oder Aufgaben.
- Deutsch als vergleichsweise kleine Sprach-Community über multilinguale Modelle besser repräsentieren.
- Eine schnellere Anpassung an neue Faktenlagen und Gegebenheiten über hybride Verfahren angehen, wie etwa durch die Verbindung von Sprachmodellen und Wissensgraphen.
- Technische Lösungsansätze zur Erkennung von Bias umsetzen und weiterentwickeln.
- Sprachverständnis der Modelle mit multimodalen Verfahren verbessern.
- Das Grounding der Sprachmodelle durch die Verbindung des Sprachmodells mit Wahrnehmung und Handlung verbessern.

Herausforderungen begegnen

- Methoden zur Erstellung effektiver Trainingsdatensätze effizienter gestalten, sodass möglichst wenig menschliche Annotationsarbeit nötig wird.
- „Halluzination“ bei generierten Inhalten begegnen.
- Logische Fähigkeiten der Modelle verbessern und Schlussfolgerungen absichern.
- Ansätze zu Erklärbarkeit, Konsistenz und Kohärenz weiterentwickeln.
- Recency Bias bei Modellen begegnen.
- Problemstellen, die bei multilingualen Modellen aufgrund von spezifischen Charakteristika der deutschen Sprache entstehen, identifizieren und ausräumen.
- Energieeffizienz des Trainings der Modelle verbessern.

Quelle: Eigene Zusammenstellung.

Impressum

Herausgeber: Lernende Systeme – Die Plattform für Künstliche Intelligenz | Geschäftsstelle | c/o acatech | Karolinenplatz 4 | D-80333 München | kontakt@plattform-lernende-systeme.de | www.plattform-lernende-systeme.de | Folgen Sie uns auf Twitter: @Lernende Systeme | Stand: Mai 2023 | Bildnachweis: AREE/Adobe Stock/Titel

Diese Kurzfassung entstand auf Grundlage des Whitepapers: *Große Sprachmodelle – Grundlagen, Potenziale und Herausforderungen für die Forschung*. München, 2023. Es wurde erstellt von Mitgliedern der Arbeitsgruppe Technologische Wegbereiter und Data Science. https://doi.org/10.48669/pls_2023-3

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

acatech
DEUTSCHE AKADEMIE DER
TECHNIKWISSENSCHAFTEN