Empfehlungen für die Nutzung von „Large Language Models“ (LLM) wie ChatGPT, Google Bard und BingAI

Tünde Fülöp, Walter Hötzendorfer & Andreas Czák

Was können LLM?

LLM sind sogenannte „Basismodelle“ (foundation model / generative AI), die für verschiedene Anwendungen geeignet sind. Sie werden auf Grundlage von großen Mengen an Text trainiert, der zum Teil ohne Rücksicht auf die Herkunft oder Legalität dieser Daten verwendet wurde („data scraping“). Die Algorithmen errechnen beim Training die statistische Häufigkeit der Aufeinanderfolge von Worten und „erraten“ bei der Beantwortung von Fragen oder Generierung von Text, was die nächste statistisch wahrscheinlichste Wortfolge im spezifischen Kontext ist. „Wahr“ oder „falsch“ spielen dabei keine Rolle, eine Antwort kann reine Erfindung sein, auch wenn sie plausibel klingt.

Datenbasis – Diskriminierung, Bias, Rechtsverletzungen

Ob ein LLM für den beabsichtigten Zweck mehr oder weniger geeignet ist, hängt auch davon ab, auf welcher Datenbasis es trainiert wurde – ein LLM, das ausschließlich auf Grundlage von peer-reviewten wissenschaftlichen Artikeln trainiert wurde, wird andere Ergebnisse liefern als ein anderes, das wahllos mit im Internet verfügbaren Texten trainiert wurde.

Antworten / Ergebnisse können, je nach Datengrundlage für das Training, diskriminierend oder verzerrt („biased“) sein, beispielsweise wenn mit Daten einer bestimmten, engen Personengruppe trainiert wurde, oder wenn menschliche Entscheidungen, die nicht diskriminierungsfrei waren, die Grundlage für das Training darstellen. In manchen Fällen wie Entscheidungen über die Auswahl von Jobbewerber*innen oder richterliche Entscheidungen kann die Verwendung unzulänglich trainierter KI eine Grundrechtsverletzung darstellen.

Es kann vorkommen, dass KI-Systeme einzelne Informationen aus den Trainingsdaten 1:1 in Antworten wiedergeben: Insbesondere bei Wiedergabe personenbezogener Daten kann das zu Rufschädigung, Identitätsdiebstahl oder anderen Risiken für betroffene Personen führen. Für die Verwendung falscher Inhalte können Sie unter Umständen haftbar gemacht werden.

Software, die von LLM generiert wird, kann (ohne umfangreiche Prüfung) nicht urheberrechtlich geschützt werden: Das Training erfolgt zum Teil auf der Grundlage von offen verfügbarem Source-Code, der unter Open Source Lizenz im Internet verfügbar ist. Sie können auch durch den Import von eingeschleustem Code, der Viren enthält, Probleme bekommen.

Nachvollziehbarkeit

Aufgrund der Art des Trainings der LLM ist kaum nachvollziehbar, wie diese zu einem bestimmten Ergebnis kommen oder welches Material die Grundlage für eine bestimmte Antwort darstellt. Im kreativen Kontext (Gedichte, Geschichten oder Einladungen zum Kindergeburtstag) spielt das keine Rolle. In anderen Zusammenhängen ist es wichtig, die Quellen zu kennen und den Wahrheitsgehalt einschätzen zu können. Manche LLM weisen die Herkunft der Informationen aus – solche Antworten und Ergebnisse lassen sich leichter nachprüfen (Google Bard und Bing AI beispielsweise beziehen Informationen aus einer Internet-Recherche in Antworten ein, Bing gibt zusätzlich auch Quellen an). Oft wird jedoch der Aufwand für die Überprüfung den Zeitgewinn durch die Verwendung des LLM übersteigen. Im akademischen Bereich kann die Verwendung von LLM-generierten Texten, die die Herkunft der Texte oder Inhalte nicht ausweisen, zu ethischen oder rechtlichen Konsequenzen (bis zur Aberkennung eines Titels) führen.

Verwendung von Eingaben („prompts“) für das laufende Training der LLM-Modelle

Bei unentgeltlicher Nutzung stimmen Sie üblicherweise zu, dass Ihre Eingaben für die Verbesserung des Modells verwendet werden – Sie trainieren das KI-System weiter. Das ist ein Problem, wenn Sie schützenswerte Inhalte eingeben: Beispielsweise Geschäftsgeheimnisse, urheberrechtlich oder durch das Recht auf geistiges Eigentum geschützte Inhalte, personenbezogene Daten, proprietärer Code, sicherheitsrelevante oder finanzielle Informationen. Diese Eingaben können in das System einfließen, was unter Umständen zur Verletzung Ihrer Rechte (oder der Rechte von Dritten) führen kann. Unterschiedliche Versionen für die „private“, unentgeltliche Nutzung und für die kostenpflichtige Variante haben oft unterschiedliche Nutzungsbedingungen und Modalitäten: So versichert z.B. Open AI, der Anbieter von ChatGPT, eingegebene Daten nicht selbst weiter zu verwenden, wenn entweder die kostenpflichtige Business Variante verwendet wird, die eine Anbindung via API (Application Programming Interface) erfordert, oder wenn der Chatverlauf deaktiviert ist. Ob das den Tatsachen entspricht und ein „Rückfluss“ von Daten tatsächlich ausgeschlossen werden kann, ist umstritten.

Haftung

Die Frage der Haftung für Rechtsverstöße (z.B. Urheberrecht, Datenschutz) ist derzeit nicht eindeutig geregelt. In den USA gibt es Urheberrechtsklagen gegen ChatGPT. Europäische Datenschutzbehörden sehen sich die DSGVO-Konformität diverser LLM an. Die Verwendung von LLM kann daher insbesondere im geschäftlichen oder behördlichen Bereich rechtliche Unsicherheit und potenzielle Rechtsverletzungen nach sich ziehen.

Empfehlungen:

1. Prüfen Sie, was das LLM kann und wofür es nicht geeignet ist. Wenn Sie korrekte Ergebnisse benötigen: Überprüfen Sie den Output von LLM anhand verlässlicher Quellen.
2. Achten Sie darauf, auf welcher Datenbasis das LLM trainiert wurde. Hinweise dazu finden Sie in den Medien. Alle großen US-Modelle (Google Bard, LLaMA von Meta, ChatGPT und Bing AI,) wurden „wahllos“ auf Basis von „data scraping“ trainiert. Das hat Auswirkungen auf die Outputs und auf die legale Verwendung der Ergebnisse.
3. Prüfen Sie, ob Ihre Eingaben („prompts“) für das laufende Training des LLM verwendet werden. Seien Sie vorsichtig bei der Eingabe geschützter Inhalte (oder verzichten Sie auf diese). Geschäftsgeheimnisse, Information die Verschwiegenheitspflichten unterliegt, personenbezogene Daten, proprietäre Daten, urheberrechtlich oder durch das Recht auf geistiges Eigentum geschützte Inhalte, finanzielle oder sicherheitsrelevante Informationen – verwenden Sie diese für die Text- oder Code-Generierung nur, wenn ausgeschlossen ist, dass sie vom System weiterverwendet werden.
4. Prüfen Sie die Anwendbarkeit der Datenschutz-Grundverordnung: Wenn Sie personenbezogene Daten verwenden, können Sie – insbesondere bei Nutzung im geschäftlichen Kontext – der DSGVO unterliegen. Nehmen Sie Fragen zur Nachvollziehbarkeit, Transparenz, Rechtsmäßigkeit, Rechte betroffener Personen (z.B. Löschung oder Richtigstellung), ungewollte (illegale) Offenlegung personenbezogener Daten und Auftragsverarbeitung bzw. gemeinsame Verantwortlichkeit ernst. Holen Sie sich wenn nötig kompetente Unterstützung, wenn Sie hohe Geldbußen vermeiden wollen: Die Datenschutzbehörden werden sich mit diesen Fragen intensiv beschäftigen.
5. Im geschäftlichen Kontext: Überlegen Sie genau, wofür sich ein Einsatz von LLM lohnt (bzw. welche Variante, unentgeltlich oder entgeltlich), einschließlich der Frage, ob Sie Ihren Compliance-Pflichten, z.B. aus der DSGVO oder anderen EU-Verordnungen wie dem Digital Services Act (mit angemessenem Aufwand) nachkommen können. Überprüfen Sie durch LLM generierten Code auf allfällige Open-Source Elemente, wenn Sie eine entgeltliche Verwendung beabsichtigen. Bedenken Sie, dass Sie für Rechtsverletzungen (einschließlich bei ungewollter Verwendung geschützter Inhalte) haftbar sein können.

6. Als (Mitarbeitende einer) Behörde müssen Sie bei Verwendung von LLM einige rechtliche Vorgaben bedenken: Achten Sie zusätzlich zu den bereits genannten Punkten u.a. darauf, ob Sie eine gesetzliche Grundlage für die Verwendung von LLM und ähnlichen Systemen benötigen. Automatisierte Entscheidungsfindung unterliegt den Vorgaben von Art. 22 DSGVO. Zudem unterliegen Sie dem Amtsgeheimnis, Rechenschaft- und Dokumentationspflichten und weiteren Dienstpflichten.
7. LLM und andere KI-Systeme werden ständig besser und spezialisierter – suchen Sie sich das Tool, das am besten für Ihren Bedarf geeignet ist.

Wenn Sie unsere Empfehlungen beachten, können Sie sich von LLM sehr gut die Arbeit des Ausformulierens von Texten abnehmen lassen – aber nur dann, wenn Sie in der Lage sind, hinterher die Richtigkeit der darin enthaltenen Informationen zu prüfen (außer, der Wahrheitsgehalt des Textes spielt im Einzelfall keine Rolle).

Wir beraten Sie gerne und hoffen, dass die Verwendung von Systemen künstlicher Intelligenz, mit Blick auf Verfahren vor Datenschutzbehörden, neue EU-Gesetze wie AI Act oder Digital Services Act und drohende Klagen, zunehmend rechtskonform und damit sicherer in der Handhabung wird. Diese Empfehlungen sind gedacht einen raschen Einblick in dieses Thema zu bieten. Sollten Sie eine Firma oder Behörde sein können wir ihnen auch eine detailliertere Beratung anbieten.

Disclaimer

Dieses Dokument wurde auf Basis jener Informationen erstellt, die den Autor*innen als für den Zweck des Dokuments relevant erschien. Die Autor*innen übernehmen jedoch keine Haftung/Gewähr für Vollständigkeit und Richtigkeit der in diesem Dokument zur Verfügung gestellten Informationen. Die Angaben in diesem Dokument können nicht als Zusicherung oder Garantie verstanden werden. Die in diesem Dokument enthaltenen Informationen können sich im Laufe der Zeit verändern oder zum Übergabezeitpunkt bereits verändert haben. Technische Änderungen vorbehalten.