Technische Universität Braunschweig

Sicherer und verantwortungsvoller Umgang mit Sprachmodellen

Rahmenbedingungen

Bei der Nutzung des Sprachmodells müssen sich die Nutzer*innen bewusst sein, dass ihre Aktivitäten rechtlichen und ethischen Rahmenbedingungen unterliegen. Auch technische Besonderheiten müssen bei der Anwendung berücksichtigt werden.

Wichtige Gesetze und Vorschriften (Auszug):

Datenschutz-Grundverordnung(DSGVO):

- Regelt den Umgang mit personenbezogenen Daten und schützt die Privatsphäre der Individuen.

Urheberrecht:

- Schützt die Rechte von Urheber*innen bezüglich ihrer Werke und Inhalte.

Allgemeines Gleichbehandlungsgesetz (AGG):

- Ziel des Gesetzes ist es, Benachteiligungen aus Gründen der Herkunft oder wegen der Ethnizität, des Geschlechts, der Religion oder Weltanschauung, einer Behinderung, des Alters oder der sexuellen Identität zu verhindern oder zu beseitigen.

Rahmenbedingungen

Verantwortung der Nutzer*innen:

Prüfung der Rechtskonformität:

- Die Nutzer*innen sind für die Überprüfung verantwortlich, ob die Verwendung des Sprachmodells und dessen Ausgaben den rechtlichen Vorgaben entsprechen.

Abwägung und Bewertung:

- Insbesondere im Hinblick auf die DSGVO müssen Nutzer*innen eine Abwägung vornehmen, ob die Nutzung der KI-Technologie im jeweiligen Kontext rechtmäßig ist.

Zusammenfassung:

- Die Kenntnis und Einhaltung der rechtlichen Rahmenbedingungen ist für einen verantwortungsvollen Umgang mit dem Sprachmodell unerlässlich.
- Nutzer*innen tragen die Verantwortung, die beschriebenen Regelungen und gesetzlichen Anforderungen bei der Nutzung des Sprachmodells einzuhalten.

Grundregel 1 – Keine personenbezogenen Daten

Anfragen an das Sprachmodell dürfen keine personenbezogenen Daten enthalten, um die Datenschutzbestimmungen einzuhalten und die Privatsphäre zu schützen.

Unzulässige Anfrage:

- Beispiel: „Kann das Sprachmodell die E-Mail-Adresse der*des Studierenden [*Vorname Nachname*] finden?“
- Problem: Verwendung spezifischer personenbezogener Daten (Name und E-Mail-Adresse).

Angemessene Anfrage:

- Beispiel: „Wie können allgemein Kontaktinformationen von Abteilungen an Universitäten gefunden werden?“
- Lösung: Allgemeine Anfrage ohne personenbezogene Daten.

Grundregel 1 – Keine personenbezogenen Daten

Umformulierte Anfrage:

- Ursprüngliche Anfrage: „Bitte gib mir die Telefonnummer von [*Professorin/Professor Name*] im Fachbereich Physik."
- Umformulierte Anfrage: „Wie finde ich die Kontaktdaten des Fachbereichs Physik an der TU Braunschweig?"
- Lösung: Vermeidung spezifischer personenbezogener Daten; allgemeine Frage.

Zusammenfassung:

- Anfragen sollten so formuliert werden, dass sie keine spezifischen personenbezogenen Daten enthalten.

Grundregel 2 – Keine Bezugnahme auf natürliche Personen

Anfragen an das Sprachmodell sollten so formuliert werden, dass sie keinen direkten Bezug zu real existierenden Personen haben. Dies dient dem Schutz der Privatsphäre und der Einhaltung von Datenschutzrichtlinien.

Beispiel - Unzulässige Anfrage:

- Unangemessen: „Kannst du mir Informationen über [*Name einer realen Person*] geben, die an der TU Braunschweig studiert?"
- Problem: Direkter Bezug zu einer realen Person; Verletzung der Privatsphäre und möglicherweise des Datenschutzes.

Beispiel - Angemessene Anfrage:

- Angemessen: „Welche allgemeinen Fähigkeiten sind für ein erfolgreiches Studium in [*Studienfach*] wichtig?"
- Lösung: Allgemeine, nicht personenbezogene Frage; keine Datenschutzprobleme

Grundregel 2 – Keine Bezugnahme auf natürliche Personen

Beispiel - Umformulierte Anfrage:

- Unangemessen: „Kannst du die neuesten Veröffentlichungen von [*Professorin/Professor Name*] analysieren?“
- Angemessen: „Kannst du allgemeine Trends in der Forschung in [*Fachgebiet*] analysieren?“
- Lösung: Direkten Personenbezug vermeiden; auf das Fachgebiet statt auf die Person fokussieren.

Zusammenfassung:

- Es ist entscheidend, dass Anfragen allgemein und themenorientiert formuliert werden, um personenbezogene Daten und die Privatsphäre zu schützen.

Technische Universität Braunschweig

Sicherer und verantwortungsvoller Umgang mit Sprachmodellen | Version 1.0 | Seite 7

Grundregel 3 - Vermeidung von personenbeziehbaren Daten

Bei Anfragen an das Sprachmodell ist darauf zu achten, dass keine Daten verwendet werden, die indirekt Rückschlüsse auf eine bestimmte Person zulassen. Dies dient dem Schutz der Privatsphäre und der Einhaltung von Datenschutzrichtlinien.

Unzulässige Anfrage:

- Beispiel: „Kann das Sprachmodell Informationen über die Person liefern, die in der Biologievorlesung am Montag um 10 Uhr an der TU Braunschweig gesprochen hat?“
- Problem: Obwohl kein direkter Name genannt wird, könnten die spezifischen Details eine Person identifizierbar machen.

Angemessene Anfrage:

- Beispiel: „Welche allgemeinen Themen werden typischerweise in Biologievorlesungen an Universitäten behandelt?“
- Lösung: Allgemeine, nicht personenbezogene Fragen, die keine Rückschlüsse auf eine bestimmte Person zulässt.

Technische Universität Braunschweig

Sicherer und verantwortungsvoller Umgang mit Sprachmodellen | Version 1.0 | Seite 8

Grundregel 3 - Vermeidung von personenbeziehbaren Daten

Umformulierte Anfrage:

- Ursprüngliche Anfrage: „Welche Forschungsprojekte leitet derzeit der Leiter des Instituts für Chemie an der TU Braunschweig?“
- Umformulierte Anfrage: „Welche Arten von Forschungsprojekten sind im Fach Chemie an Universitäten üblich?“
- Lösung: Allgemeine Frage ohne Bezug zu einer bestimmten Person.

Zusammenfassung:

- Es ist wichtig, Fragen so zu formulieren, dass sie keine Rückschlüsse auf einzelne Personen zulassen.

Grundregel 4 – Respektierung des Urheberrechts

Bei Anfragen an das Sprachmodell ist darauf zu achten, dass keine urheberrechtlich geschützten Inhalte verwendet oder angefragt werden. Dies respektiert die Rechte der Urheber*innen und vermeidet rechtliche Probleme.

Unzulässige Anfrage:

- Beispiel: „Kann das Sprachmodell den vollständigen Text des Romans '1984' von George Orwell wiedergeben?“
- Problem: Direkte Anfrage nach einem urheberrechtlich geschützten Werk.

Angemessene Anfrage:

- Beispiel: „Kannst du eine Zusammenfassung und Analyse des Romans '1984' von George Orwell geben?“
- Lösung: Anfrage nach Informationen über das Werk, nicht nach dem Werk selbst.

Grundregel 4 – Respektierung des Urheberrechts

Umformulierte Anfrage:

- Ursprüngliche Anfrage: „Bitte spiele das Lied 'Bohemian Rhapsody' von Queen."
- Umformulierte Anfrage: „Kannst du Informationen zur Bedeutung und zum Hintergrund des Liedes 'Bohemian Rhapsody' von Queen geben?"
- Lösung: Bitte um Informationen über das Lied, nicht um eine urheberrechtlich geschützte Aufnahme des Liedes.

Zusammenfassung:

- Anfragen sollten so formuliert werden, dass urheberrechtlich geschützte Werke nicht direkt angefordert oder wiedergegeben werden. Dies respektiert das Urheberrecht und vermeidet rechtliche Probleme.

Grundregel 5 – Wahrung des Rechts am eigenen Bild

Bei Anfragen an das Sprachmodell ist das Recht am eigenen Bild zu beachten. Das bedeutet, dass Bilder oder Informationen, die bestimmte Personen identifizierbar machen, nicht ohne deren Zustimmung verwendet werden dürfen.

Erstellung von Bildern ohne Zustimmung:

- Beispiel: Verwendung eines KI-Tools zur Erstellung eines Porträts einer real existierenden Person ohne deren Einwilligung.
- Problem: Die Erstellung eines Bildes einer realen Person ohne deren Einwilligung verstößt gegen das Recht am eigenen Bild.

Verwendung von Bildern in einem unzulässigen Kontext:

- Beispiel: Verwendung von Bildern von Universitätsangehörigen für Flyer oder Broschüren durch KI-gestützte Software ohne deren Genehmigung.
- Problem: Die Verwendung von Bildern in einem Kontext, für den keine Genehmigung vorliegt, kann das Recht am eigenen Bild verletzen.

Grundregel 5 – Wahrung des Rechts am eigenen Bild

Automatisierte Gesichtserkennung ohne Einwilligung:

- Beispiel: Einsatz eines KI-basierten Systems zur Gesichtserkennung realer Personen ohne Einwilligung der Betroffenen.
- Problem: Die Nutzung von Gesichtserkennungstechnologien kann in das Recht am eigenen Bild verletzen, insbesondere wenn keine explizite Zustimmung vorliegt.

Zusammenfassung:

- Beim Einsatz von KI-gestützten Anwendungen ist stets darauf zu achten, dass das Recht am eigenen Bild gewahrt bleibt. Dazu gehört auch der respektvolle Umgang mit Bildmaterial und personenbezogenen Informationen.

Grundregel 6 –Diskriminierung vermeiden und verhindern

Bei der Verwendung des Sprachmodells ist darauf zu achten, dass keine diskriminierenden Fragen gestellt oder Antworten generiert werden, die aufgrund von Herkunft, ethnischer Zugehörigkeit, Geschlecht, Religion, Weltanschauung, Behinderung, Alter oder sexueller Identität diskriminieren.

Beispiele für diskriminierende Verwendungen:

Verwendung von Stereotypen in Anfragen:

- Beispiel: Anfragen an das Sprachmodell, die stereotype Annahmen über bestimmte soziale Gruppen beinhalten, z.B. „Warum essen die Deutschen Eisbein mit Sauerkraut?“.
- Problem: Solche Anfragen basieren auf diskriminierenden Stereotypen und verstärken diese.

Grundregel 6 – Diskriminierung vermeiden und verhindern

Erstellung diskriminierender Inhalte:

- Beispiel: KI wird eingesetzt, um Texte oder Bilder zu erstellen, die rassistische, sexistische oder anderweitig diskriminierende Botschaften vermitteln.
- Problem: Die Erstellung solcher Inhalte verstößt gegen Gleichbehandlungsgrundsätze und geltendes Recht (z. B. Grundgesetz).

Diskriminierung durch KI-Bias:

- Beispiel: Nutzung des Sprachmodells, um Daten nach Geschlecht oder Herkunft zu bewerten.
- Problem: Sprachmodelle können unbewusste Vorurteile verstärken und zu diskriminierenden Entscheidungen führen.

Zusammenfassung:

- Es ist wichtig, dass sowohl die Anfragen an das Sprachmodell als auch die Antworten frei von diskriminierenden Inhalten sind.
- Bewusstsein und Vorsicht sind erforderlich, um Diskriminierung in jeder Form zu vermeiden.

Überprüfung der Antworten auf Richtigkeit

Nutzer*innen sind angehalten, die Richtigkeit der von dem Sprachmodell bereitgestellten Informationen aktiv zu überprüfen. Dies ist besonders wichtig, da KI-Systeme unter Umständen ungenaue oder irreführende Informationen liefern können.

Wichtigkeit der Überprüfung:

Vermeidung von Fehlinformationen:

- KI-Systeme können manchmal falsche oder veraltete Informationen wiedergeben.
- Die Nutzer*innen sollten sich nicht ausschließlich auf die Antworten des Sprachmodells verlassen.

Erkennung von Halluzinationen:

- KI kann manchmal „halluzinieren“, d.h. Informationen generieren, die fälschlicherweise als Fakten präsentiert werden.
- Um solche Halluzinationen zu erkennen, müssen die Antworten kritisch hinterfragt werden.

Überprüfung der Antworten auf Richtigkeit

Beispiele für notwendige Überprüfungen:

Faktenchecks:
- Überprüfung von Daten, Statistiken und historischen Fakten zur Sicherstellung der Richtigkeit.

Aktualität der Informationen:
- Überprüfung, ob die bereitgestellten Informationen aktuell und relevant sind.

Kontextbezogene Richtigkeit:
- Überprüfung, ob die Antwort im richtigen Kontext und im Rahmen der gestellten Frage angemessen ist.

Zusammenfassung:
- Die eigenständige Überprüfung der von der KI bereitgestellten Informationen ist unerlässlich. Dies sichert die Zuverlässigkeit der Nutzung und fördert ein kritisches Verständnis der Technologie.

Umgang mit potenziell unangemessenen Inhalten

Nutzer*innen müssen darauf vorbereitet sein, dass das Sprachmodell manchmal Antworten generiert, die als unangemessen betrachtet werden können, auch wenn sie sachlich korrekt sind.

Beispiele für möglicherweise unangemessene Inhalte:

Diskriminierende Aussagen:
- Aussagen, die zwar historisch korrekt sein mögen, aber heutige Werte von Gleichheit und Respekt verletzen.

Sensible Themen:
- Informationen zu sensiblen Themen wie Religion oder Politik, die jedoch in bestimmten Kontexten zu Missverständnissen oder Konflikten führen können.

Spekulative Antworten:
- Antworten, die Vermutungen oder Annahmen enthalten, die plausibel erscheinen, aber nicht durch Fakten belegt sind. Dies kann in wissenschaftlichen oder akademischen Kontexten problematisch sein.

Umgang mit potenziell unangemessenen Inhalten

Umgang mit unangemessenen Inhalten:

- Kritische Beurteilung der Angemessenheit jeder Antwort.
- Bereitschaft, Antworten zu hinterfragen und gegebenenfalls zu verwerfen, wenn sie nicht angemessen erscheinen.

Zusammenfassung:

- Ein bewusster und reflektierter Umgang mit den Ausgaben des Sprachmodells ist entscheidend.
- Die Nutzer*innen tragen die Verantwortung, die Angemessenheit der Antworten im spezifischen Nutzungskontext zu beurteilen.

Überwachung der Einhaltung

Die Einhaltung der Nutzungsbedingungen und der rechtlichen Vorgaben bei der Verwendung des Sprachmodells wird durch den Datenschutzbeauftragten, das Datenschutzmanagement und den Personalrat aktiv durch stichprobenartige Kontrollen überwacht, um einen sicheren und rechtskonformen Umgang mit dem Sprachmodell zu gewährleisten.

Konsequenzen bei Verstößen:

- Bei Verstößen gegen die Nutzungsbedingungen können Maßnahmen ergriffen werden, die von Verwarnungen bis hin zu weiteren rechtlichen Schritten reichen.
- Die Wichtigkeit der Einhaltung der Regeln wird betont, um einen verantwortungsvollen Umgang zu fördern und rechtliche Risiken zu minimieren.

Zusammenfassung:

- Eine kontinuierliche Überwachung und Bewertung der Nutzung des Sprachmodells ist unerlässlich, um Datenschutz und Compliance sicherzustellen.
- Die Nutzer*innen sind angehalten, stets im Einklang mit den festgelegten Richtlinien zu handeln.

Abschluss und Zusammenfassung

In dieser Präsentation wurden die wesentlichen Richtlinien und Verantwortlichkeiten im Umgang mit KI-Technologien und insbesondere mit Sprachmodellen aufgezeigt.

Zusammenfassung der wichtigsten Punkte:

- **Datenschutz**: Keine Verwendung personenbezogener oder personenbeziehbarer Daten.
- **Urheberrecht**: Beachtung des Urheberrechts bei der Verwendung von Inhalten.
- **Diskriminierungsverbot**: Vermeidung jeglicher Form von Diskriminierung.
- **Überprüfung und Angemessenheit**: Kritische Überprüfung der Richtigkeit und Angemessenheit der Ausgaben für KI.
- **Rechtliche Rahmenbedingungen**: Einhaltung aller relevanten gesetzlichen Bestimmungen.

Abschluss und Zusammenfassung

Appell an die Nutzer*innen:

- Wir appellieren an Sie als Nutzer*innen, die dargelegten Richtlinien stets zu beachten und umzusetzen.
- Ihr verantwortungsvoller Umgang mit der Technologie trägt entscheidend zur Wahrung von Ethik, Recht und Datenschutz bei.
- Die Einhaltung dieser Regeln stellt nicht nur die Rechtskonformität sicher, sondern fördert auch ein positives und produktives Umfeld für den Einsatz von KI-Technologie.

Abschließender Gedanke:

- Die Technologie bietet enorme Möglichkeiten, aber mit großer Macht kommt auch große Verantwortung. Wir zählen auf Ihre Mitarbeit, um diese Technologie sicher, ethisch und im Interesse aller zu nutzen.