Deutscher Bundestag **Drucksache 21/4312**

21. Wahlperiode 23.02.2026

Antwort

der Bundesregierung

auf die Kleine Anfrage der Abgeordneten Ruben Rupp, Robin Jünger, Alexander Arpaschi, weiterer Abgeordneter und der Fraktion der AfD – Drucksache 21/4037 –

Zum Vorhaben eines quelloffenen europäischen Sprachmodells

Vorbemerkung der Fragesteller

Die Veröffentlichung der KI-Software (KI = Künstliche Intelligenz; engl.: AI) ChatGPT des Unternehmens OpenAI im November 2022 teilt die Evolution der Künstlichen Intelligenz in ein Davor und ein Danach. Erstmals in der Geschichte dieser Technologie wird plastisch für jedermann greifbar, wozu Künstliche Intelligenz heute imstande ist. Mit Künstlicher Intelligenz ist es auch für technische Laien möglich, Texte aller Art zu generieren, Bilder, Videos und Audiodateien anzufertigen und sich in eine ausführliche Konversation über gesellschaftliche und private Themen zu begeben. Dabei kommuniziert das Programm humanoid – es ist allein aus den Antworten oft nicht mehr ersichtlich, dass sie von einer Software kommen und nicht von einem Menschen. Die human anmutende Sprachdarstellung wird von Large Language Models (LLM), also großen Sprachmodellen, ermöglicht. Hinzu kommen multimodale Modelle, die parallel mit verschiedenen Informationsarten umgehen können, etwa Texten und Bildern.

ChatGPT und vergleichbare KI-Lösungen wie Gemini, Grok, Claude, Llama, und DeepSeek werden mit großen Datenmengen, deren Herkunft, Tiefe und Struktur nicht direkt nachvollziehbar sind, auf ihren Praxiseinsatz hin trainiert. Zudem stammen die genannten LLM von großen Konzernen, die diese Produkte in ihre Wertschöpfungsketten integrieren und damit bereits vorhandene Nutzer weiter an sich binden. Dergestalt droht sich auf dem Gebiet der Künstlichen Intelligenz die globale Dominanz einiger weniger außereuropäischer Unternehmen, die bereits das Feld der Digitalisierung bestimmen, zu wiederholen. Vor dem Hintergrund einer anzustrebenden digitalen Souveränität für Deutschland ist dies nach Auffassung der Fragesteller als kritisch zu bewerten. Nicht zuletzt kommen fast alle großen Sprachmodelle aus den USA, was dazu führt, dass weniger weit verbreitete Sprachen neben Englisch nicht so filigran und detailliert repräsentiert werden.

Zum 1. Februar 2025 hat die Europäische Kommission das Projekt OpenEuroLLM (https://openeurollm.eu/) lanciert, das alle Amtssprachen der EU abdecken, die Einhaltung des AI Acts der EU gewährleisten und bei Unternehmen, Entwicklern sowie Endnutzern Akzeptanz finden soll (https://strategic-technologies.europa.eu/step-results/step-stories/openeurollm-european-family-large-language-models_en?prefLang=de&etrans=de). Das Projekt nutzt die Hoch-

---

Die Antwort wurde namens der Bundesregierung mit Schreiben des Bundesministeriums für Forschung, Technologie und Raumfahrt vom 23. Februar 2026 übermittelt.

Die Drucksache enthält zusätzlich – in kleinerer Schrifttype – den Fragetext.

leistungsrecheninfrastruktur EuroHPC, verwendet Daten europäischer Herkunft zum Training und legt den Quellcode des Programms offen (www.heise.de/news/Sprachmodell-OpenEuroLLM-soll-KI-in-der-EU-unabhaengiger-und-vielfaeltiger-machen-10269667.html).

An diesem Konsortium sind insgesamt neun Länder aus der EU beteiligt. Aus Deutschland sind die Eberhard-Karls-Universität Tübingen, das Forschungszentrum Jülich, die Fraunhofer Gesellschaft zur Förderung der angewandten Forschung sowie die Unternehmen Aleph Alpha und Ellamind involviert. Die Bundesregierung zeigt sich prinzipiell offen für die Entwicklung offener europäischer Plattformmodelle (www.koalitionsvertrag2025.de/sites/www.koalitionsvertrag2025.de/files/koav_2025.pdf, hier S. 71).

1. Welche Erkenntnisse liegen der Bundesregierung rund ein Jahr nach dem Start des Projekts OpenEuroLLM über dessen gegenwärtigen Stand ggf. vor, und wann ist nach Kenntnis der Bundesregierung mit der Veröffentlichung eines ersten großen Sprachmodells im Rahmen des Projekts zu rechnen?

2. Gibt es nach Kenntnis der Bundesregierung bereits erste Pilotprojekte, mit denen die zu entwickelnden offenen europäischen Sprachmodelle getestet werden können, und wenn ja, welche Projekte existieren (bitte auflisten)?

Die Fragen 1 und 2 werden im Zusammenhang beantwortet.

OpenEuroLLM wird durch die Europäische Kommission im Rahmen des Digitalen Europa-Programms finanziert. Zu aktuellen Projektfortschritten und zum gegenwärtigen Stand des Projektes wird auf die Internetseite des Projektes verwiesen.

3. Sind nach Einschätzung der Bundesregierung die veranschlagten 20,65 Mio. Euro der EU (vgl. https://strategic-technologies.europa.eu/step-results/step-stories/openeurollm-european-family-large-language-models_en?prefLang=de&etrans=de) ausreichend, um ein wettbewerbsfähiges großes Sprachmodell zu entwickeln, und werden nach Einschätzung der Bundesregierung mit dieser Summe lediglich die Entwicklungskosten eines großen Sprachmodells abgedeckt oder auch zusätzlich die Kosten zum Training der Modelle (siehe Vorbemerkung der Fragesteller; bitte ausführen)?

Ziel der Entwicklung von Sprachmodellen ist das Training des Systems mit Hilfe sehr umfangreicher Daten. Die Kosten fallen je nach Modell und Ziel unterschiedlich aus.

Der Bundesregierung liegen zu den Details des Projektes der Europäischen Union (EU) keine Angaben vor, die eine belastbare Bewertung des Projekts über Evaluierungen der Europäischen Kommission hinaus erlauben würden.

4. Ist die Bundesregierung direkt finanziell am Projekt OpenEuroLLM beteiligt (siehe Vorbemerkung der Fragesteller; bitte ausführen), wenn ja, in welcher Höhe und über welchen Einzelplan des Bundeshaushalts, und wenn nein, plant die Bundesregierung künftig eine finanzielle Beteiligung am Projekt OpenEuroLLM, die über die indirekte Förderung über die EU hinausgeht?

Die Bundesregierung ist nicht direkt an dem EU-Projekt finanziell beteiligt. Entscheidungen über die künftige Förderung von KI-Modellen über die Festlegungen in der Hightech-Agenda Deutschlands hinaus liegen nicht vor.

5. Welche deutschen Hoch- und Höchstleistungsrechner sind nach Kenntnis der Bundesregierung Teil der Recheninfrastruktur, auf der die zu schaffenden großen Sprachmodelle des Projekts OpenEuroLLM trainiert werden?

Das Projekt OpenEuroLLM wird nach Kenntnis der Bundesregierung auf den Supercomputern des Gemeinsamen Unternehmens EuroHPC gerechnet werden. Zu den Rechnern gehört der Supercomputer „Jupiter“ am Forschungszentrum Jülich.

6. Welche Herausforderungen erkennt die Bundesregierung ggf. bei der Bereitstellung ausreichender Rechenleistung für das multilinguale Training der Sprachmodelle des Projekts OpenEuroLLM, und wie stellen sich diese Herausforderungen dar, verglichen mit der Recheninfrastruktur, auf die Anbieter proprietärer Sprachmodelle zurückgreifen können?

Laut Angaben des Projektes OpenEuroLLM hat das Projekt garantierten Zugang zu über zehn Millionen GPU-Stunden. Die Anbieter proprietärer Sprachmodelle machen zumeist keine öffentlichen Angaben zu z. B. GPU-Stunden, die für das Training eines Sprachmodells eingesetzt wurden. Vergleiche sind daher nicht möglich.

7. Welche Maßnahmen plant die Bundesregierung ggf., um deutschen Startups, kleinen und mittleren Unternehmen (KMU), Behörden und Forschungsinstituten den Zugang zu Rechenressourcen im Rahmen von OpenEuroLLM zu erleichtern?

Die Bundesregierung geht davon aus, dass die Rechenressourcen des EU-Projektes OpenEuroLLM durch EuroHPC ausreichend sein werden.

8. Aus welchen Quellen stammen nach Kenntnis der Bundesregierung die digitalen Daten, mit denen die zu entwickelnden Sprachmodelle des Projekts OpenEuroLLM trainiert werden, und werden dabei personenbezogene Daten, soweit sie in den Quellen vorkommen, pseudoanonymisiert oder gar anonymisiert (siehe Vorbemerkung der Fragesteller)?

9. Wie wird nach Kenntnis der Bundesregierung sichergestellt, dass die Trainingsdaten für das Projekt OpenEuroLLM datenschutzkonform beschafft und verarbeitet werden, insbesondere bei der Integration sensibler Daten aus EU-Sprachräumen mit vergleichsweise wenigen aktiven Sprechern und geringen finanziellen Ressourcen (bitte ausführen)?

Die Fragen 8 und 9 werden im Zusammenhang beantwortet.

Aktuelle Angaben zu Datenquellen des EU-Projektes OpenEuroLLM liegen der Bundesregierung nicht vor.

10. Existieren nationale Initiativen oder nach Kenntnis der Bundesregierung Kooperationen mit Bibliotheken, Stiftungen und Archiven, um Trainingsdaten für das Projekt OpenEuroLLM bereitzustellen, und wenn ja, wie wird nach Kenntnis der Bundesregierung sichergestellt, dass solche Datensätze urheberrechtskonform verarbeitet werden (bitte ausführen)?

Hinsichtlich nationaler Initiativen und Kooperationen des EU-Projektes OpenEuroLLM mit Bibliotheken, Stiftungen und Archiven sowie hinsichtlich der Prozesse zur Einhaltung des Urheberrechtes liegen der Bundesregierung keine Informationen vor.

11. Plant die Bundesregierung eine nationale Adoptionsstrategie, um ein quelloffenes Sprachmodell des Projekts OpenEuroLLM in der öffentlichen Verwaltung, etwa als Chatbot, zu etablieren und so den Wettbewerb mit proprietären Anbietern zu stärken (bitte ausführen)?

Die Bundesverwaltung verfolgt bei der Entwicklung von KI-Anwendungen das Ziel, diese möglichst unabhängig von konkreten Sprachmodellen auszugestalten. Hierdurch wird gewährleistet, dass etwaige LLM kurzfristig ausgetauscht werden können. Beispielsweise wird mit dem KI-System KIPITZ für die Bundesverwaltung ein zentrales, sicheres und souveränes KI-Portal mit verwaltungsspezifischen Anwendungen für die Verwaltungsmitarbeiter angeboten. KIPITZ ist so konzipiert, dass es modellunabhängig genutzt werden kann und somit eine Abhängigkeit von einem einzelnen LLM-Anbieter vermieden wird. Mit EuroLLM-9B Instruct wird bereits ein LLM des Projektes OpenEuroLLM in KIPITZ bereitgestellt.

12. Hat sich die Bundesregierung mit den Erfolgsaussichten des Projekts OpenEuroLLM im Wettbewerb mit proprietären Modellen, insbesondere hinsichtlich Leistung und Marktakzeptanz, beschäftigt und sich ggf. dazu eine eigene Positionierung erarbeitet (wenn ja, bitte ausführen), und was wären nach Einschätzung der Bundesregierung ggf. geeignete Kriterien, um die Leistung und Marktakzeptanz eines quelloffenen Sprachmodells zu messen (bitte ausführen)?

Aussagen zu Qualität und Leistungsfähigkeit des Projektes lassen sich erst im weiteren Verlauf treffen.

13. Welche Kriterien werden nach Kenntnis der Bundesregierung bei der Vergabe von Fördermitteln an deutsche Mitglieder im OpenEuroLLM-Konsortium angelegt, und wie wird der effiziente Einsatz dieser Mittel sichergestellt (bitte ausführen)?

Details über die Förderung auf EU-Ebene liegen der Bundesregierung nicht vor. Im Übrigen wird auf die Antwort zu Frage 4 verwiesen.

14. Wie wirkt sich nach Kenntnis der Bundesregierung die Finanzierung des Projektes OpenEuroLLM auf andere nationale KI-Projekte aus?

Der Bundesregierung hat keine Kenntnis von einem vergleichbaren multilingualen Sprachmodell. Im Erfolgsfall kann ein Open Source Modell wie OpenEuroLLM wichtige Grundlagen für andere Vorhaben schaffen.

Gesamtherstellung: H. Heenemann GmbH & Co. KG, Buch- und Offsetdruckerei, Bessemerstraße 83–91, 12103 Berlin, www.heenemann-druck.de
Vertrieb: Bundesanzeiger Verlag GmbH, Postfach 10 05 34, 50445 Köln, Telefon (02 21) 97 66 83 40, Fax (02 21) 97 66 83 44, www.bundesanzeiger-verlag.de
ISSN 0722-8333