

Pressemitteilung

Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS

Katrin Berkler

16.05.2024

<http://idw-online.de/de/news833764>

Forschungsprojekte, Kooperationen
Elektrotechnik, Informationstechnik
überregional



8,8 Millionen GPU-Stunden für große Sprachmodelle: Durchbruch für Generative KI-Forschung in Deutschland und Europa

Das Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS und die NLU-Gruppe von AI Sweden haben gemeinsam Rechenzeit am neuen Hochleistungsrechner MareNostrum 5 im Barcelona Supercomputing Center erhalten. Es handelt sich um eines der größten Kontingente, die seitens des Gemeinsamen Unternehmens für europäisches Hochleistungsrechnen (GU EuroHPC) für die Entwicklung europäischer großer KI-Sprachmodelle (LLMs) auf der EuroHPC-Infrastruktur gewährt wurden. Ab Ende Mai 2024 starten die Partner damit, die ersten multilingualen Modelle zu rechnen. Laufzeit des Projekts »EuroLingua-GPT« ist ein Jahr.

Das über einen EuroHPC »Extreme Scale Access« bewilligte Kontingent umfasst 8,8 Millionen GPU-Stunden auf H100-Chips und steht seit Mai bereit. »Die gewonnenen Rechenkapazitäten sind ein Meilenstein für Deutschland und Europa. Die damit trainierten Modelle werden den Einsatz generativer KI in Unternehmen massiv beschleunigen und sowohl Wirtschaft als auch Wissenschaft einen Boost geben – GenAI ‚made in Europe‘ wird damit konkret«, so Dr. Joachim Köhler, Abteilungsleiter NetMedia am Fraunhofer IAIS. Mit den neuen Rechenkapazitäten können kleine Modelle im Bereich von 7 bis 34 Milliarden Parametern, und große Modelle mit bis zu 180 Milliarden Parametern von Grund auf trainiert werden.

Eine Modellfamilie, alle europäischen Sprachen – Fraunhofer IAIS und AI Sweden bündeln ihre Expertise

Die neuen EuroLingua-Modelle bauen auf einem Trainingsdatensatz auf, der aus 45 europäischen Sprachen, Dialekten und Code besteht, einschließlich der 24 europäischen Amtssprachen. Damit liegt eine signifikante Gewichtung auf europäischen Sprachen und Werten – bisher sind multilinguale KI-Sprachmodelle noch selten. Das Training startet Ende Mai 2024. Die Veröffentlichung erster gemeinsamer Modelle soll im Laufe der kommenden Monate erfolgen.

Projektleiter Dr. Nicolas Flores-Herr, Teamleiter Conversational AI am Fraunhofer IAIS, sagt: »Das Ziel unserer Zusammenarbeit mit AI Sweden ist es, eine Familie von großen KI-Sprachmodellen von Grund auf zu trainieren, die Open Source veröffentlicht werden sollen.« Magnus Sahlgren, Head of Research NLU bei AI Sweden, ergänzt: »Sowohl der öffentliche als auch der private Sektor in der EU verlangen nach offenen, leistungsfähigen Sprachmodellen, die für europäische Sprachen trainiert sind. EuroLingua ist eine Möglichkeit, diesen Bedarf zu adressieren.«

Die Modelle, die auf der EuroHPC-Infrastruktur entwickelt werden, sollen zum einen als generalistische Basismodelle Forschung und Wissenschaft fördern, zum anderen – etwa in gemeinsamen Transfer-Projekten – auch auf bestimmte Branchen oder Themengebiete spezialisiert für den produktiven Einsatz in Unternehmen oder öffentlichen Verwaltungen trainiert werden.

Dafür bündeln beide Organisationen ihre Expertise: Das Fraunhofer IAIS und die NLU-Gruppe von AI Sweden sind zwei der führenden LLM-Labore in Europa mit ausgewiesener Expertise und jahrelanger Erfahrung in der Entwicklung von

LLMs. So leitet Fraunhofer das vom Bundesministerium für Wirtschaft und Klimaschutz (BMWK) geförderte Konsortialprojekt OpenGPT-X, in der ebenfalls große europäische, multilinguale Open-Source-Modelle entwickelt werden. Die NLU-Gruppe von AI Sweden hat das LLM GPT-SW₃ für die skandinavischen Sprachen entwickelt. Die beiden Teams arbeiten auch gemeinsam an weiteren Open-Source-Community-Projekten. Zudem ist EuroLingua-GPT eines von drei großen laufenden EU-Projekten zu Sprachmodellen, an denen Fraunhofer IAIS und AI Sweden beteiligt sind. Die beiden anderen sind TrustLLM und Deploy AI.

URL zur Pressemitteilung: <http://www.iais.fraunhofer.de> Webseite Fraunhofer IAIS

URL zur Pressemitteilung: <http://www.ai.se/en> Webseite AI Sweden

URL zur Pressemitteilung: <http://www.bsc.es/> Webseite Barcelona Supercomputing Center

Anhang Pressemitteilung FraunhoferIAIS EuroLingua als PDF <http://idw-online.de/de/attachment102976>

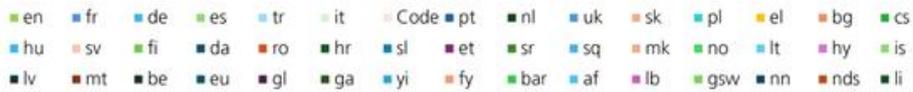
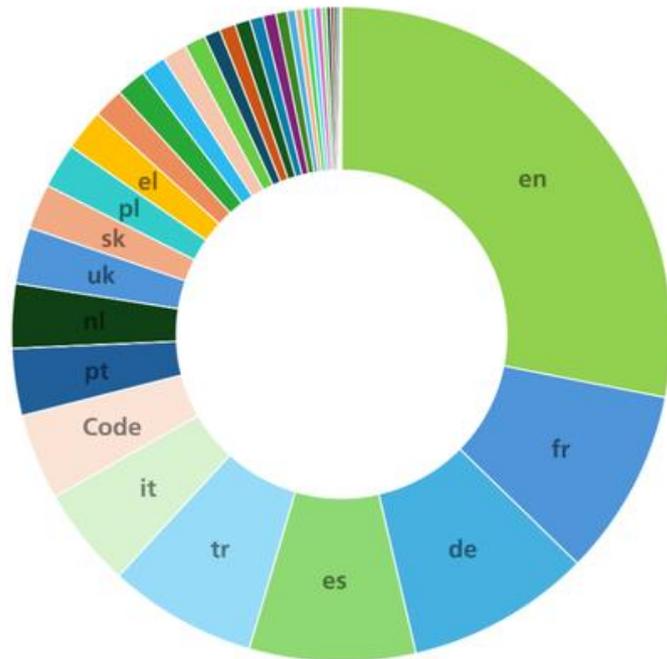


8,8 Millionen Stunden Rechenkapazität erhalten Fraunhofer IAIS und AI Sweden am neuen Hochleistungsrechner MareNostrum 5 im Barcelona Supercomputing Center.

Barcelona Supercomputing Center

By courtesy of Barcelona Supercomputing Center - www.bsc.es

45 europäische Sprachen, Dialekte und Code Verteilung der Trainingsdaten für das erste EuroLingua-Modell*



© Fraunhofer IAIS 5/2024

*Änderung der Sprachzusammensetzung im Projektverlauf vorbehalten.

Die EuroLingua-Modelle bauen auf einem Trainingsdatensatz auf, der aus 45 europäischen Sprachen, Dialekten und Code besteht, einschließlich der 24 euro-päischen Amtssprachen.

Fraunhofer IAIS
Fraunhofer IAIS