

Supercomputer Fugaku hilft bei der Entwicklung von verteilten Trainings für Large Language Models

Wien, 11. August 2023 – Fujitsu, das Forschungsinstitut RIKEN, die Tohoku University sowie das Tokyo Institute of Technology (Tokyo Tech) haben im Mai 2023 eine gemeinsame Initiative zur Entwicklung von verteilten Trainings für Large Language Models (LLMs) ins Leben gerufen. Basis für das Projekt über generative KI-Anwendungen ist der Supercomputer Fugaku, dessen Nutzung im Rahmen der von der japanischen Politik definierten Initiative zum Einsatz des Supercomputers stattfindet.

LLMs sind KI-basierte Deep Learning-Modelle und bilden den Kern generativer KI, wie etwa bei ChatGPT. Die vier an der Zusammenarbeit beteiligten Parteien wollen die Erstellung und das Training von LLMs verbessern, die bereits jetzt breite Anwendung in Universitäten und Unternehmen finden. Darüber hinaus wird die KI-Forschung in Japan vorangetrieben, und der Nutzungsbereich von Fugaku im universitären wie im kommerziellen Bereich, verbreitert.

Hintergrund

LLMs und generative KI nehmen eine grundlegende Rolle bei der Erforschung und Entwicklung von Technologien für die Sicherheit, die Wirtschaft und die Gesellschaft insgesamt. Zu beachten ist, dass die Weiterentwicklung und Verfeinerung dieser Modelle leistungsstarke Rechenressourcen erfordert, die große Datenmengen effizient verarbeiten können.

Fujitsu, RIKEN, die Tohoku University und Tokyo Tech wollen diese Ressourcen gemeinsam zur Verfügung stellen und sich dabei auf verteilte Trainings von LLMs konzentrieren. Die Implementierung der Initiative zur Fugaku-Nutzung seitens der japanischen Politik, ist im Mai 2023 gestartet und läuft bis März 2024.

Rolle und Funktionen der einzelnen Beteiligten

Die Technologie, die im Rahmen des Projekts zum Einsatz kommt, ermöglicht den Organisationen ein effizientes Large-Scale Language Model Training in der Computing-Umgebung von Fugaku. Dabei übernehmen die Partner im Einzelnen folgende Rollen:

- **Fujitsu:** Beschleunigung der LLMs.
- **RIKEN:** Verteiltes Rechnen durch Parallelisierung und Beschleunigung der LLM-Kommunikation sowie Beschleunigung der LLMs.
- **Tohoku University:** Aggregation der Lerndaten, Auswahl der Modelle.
- **Tokyo Institute of Technology:** Gesamtübersicht über sämtliche Prozesse, Parallelisierung und Beschleunigung von LLMs.

Weitere Pläne

Um japanische Forscher und Ingenieure künftig bei der Entwicklung von LLMs zu unterstützen, planen die vier Partner, die Forschungsergebnisse, die im Rahmen, der durch die japanische Politik definierten Initiativen zur Nutzung von Fugaku erzielt wurden, im Geschäftsjahr 2024 auf GitHub (3) und Hugging Face (4) zu veröffentlichen. Es wird zudem erwartet, dass zahlreiche Forscher und Ingenieure an der Optimierung des Basismodells sowie neuen Forschungsansätzen teilhaben werden – und damit neue, effiziente Methoden erarbeiten können, mit denen sich noch viel mehr erreichen lässt.

„Durch die Zusammenarbeit aller Beteiligten werden die Parallelisierung und Beschleunigung großer LLMs verbunden. Dies geschieht im Zusammenspiel mit ‚Fugaku‘. Hierfür arbeiten Tokyo Tech und RIKEN, sowie Fujitsus Abteilung für High-Performance Computing-Infrastruktursoftware zusammen. Zusätzlich wird das Performance Tuning von KI-Modellen und die Natural Language Processing-Technologie der Tohoku University eingesetzt. Mit Fujitsu wollen wir künftig auch unser gemeinsames Forschungslabor ‚Fujitsu Collaborative Research Center for Next Generation Computing Infrastructure‘ nutzen“, kommentiert Toshio Endo, Professor, Global Scientific Information and Computing Center am Tokyo Institute of Technology. „Wir freuen uns sehr auf die gemeinsame Arbeit, denn sie wird die KI-Forschung in Japan entscheidend voranbringen.“

„Unser Ziel ist ein großes Open Source-Language-Modell zur Verfügung zu stellen, das kommerziell genutzt werden kann und vor allem auf japanischen Daten basiert – einschließlich einer bestimmten Transparenz bei den Trading-Informationen“, ergänzt Kentaro Inui, Professor der Graduate School of Information Sciences an der Tohoku University. „Indem wir die Nachverfolgbarkeit der Lerndaten gewährleisten, werden wir die Forschung auch in Peripheriebereichen wie Bias, Falschinformation und dem so genannten Halluzinations-Phänomen der KI vorantreiben. Die Deep Learning Insights aus dem Natural Language Processing an der Tohoku University ermöglichen den Aufbau echter Large-Scale-Modelle. Wir sind zuversichtlich, durch das Teilen von Informationen und Erkenntnissen entscheidend zur Entwicklung von KI-Kapazitäten beitragen zu können.“

„Es ist aufregend, eine solche Chance für den zielgerichteten Einsatz von Fugaku zu bekommen. Seine Kapazitäten im Hinblick auf parallele Computing-Ressourcen haben einen Beschleunigungseffekt für die KI-Forschung und die Entwicklung von LLMs“, so Seishi Okamoto, EVP, Head of Fujitsu Research bei Fujitsu. „Im weiteren Verlauf wollen wir die Ergebnisse dieses Projekts in unsere neue KI-Plattform ‚Kozuchi‘ integrieren, um revolutionäre Anwendungen zu entwickeln, die auch und vor allem zu einer insgesamt höheren Nachhaltigkeit in der Gesellschaft beitragen.“

„Der A64FX Prozessor von Fujitsu verfügt bereits über eine KI-Beschleunigungsfunktion, die so genannte SVE. Die Softwareentwicklung und -optimierung ist maßgeblich, damit die CPU für KI-Anwendungen genutzt werden kann“, resümiert Satoshi Matsuoka, Director des RIKEN Center for Computational Science. „Unsere kollektiven Anstrengungen sorgen für einen Zusammenschluss aller relevanten LLM-Experten und Computerwissenschaftler in Japan einschließlich der R-CCS-Forscher und -Ingenieure bei RIKEN. So können spezifische Technologien zum Bau von LLMs auf dem ‚Fugaku‘ entwickelt werden. Zusammen mit unseren Partnern leisten wir so einen Beitrag für die Society 5.0.“

Weitere Informationen über Fujitsu erhalten Sie über

- Fujitsu Aktuell: <http://www.fujitsu.com/de/fujitsuaktuell>
- Twitter: http://twitter.com/Fujitsu_DE
- LinkedIn: <http://www.linkedin.com/company/Fujitsu>
- Facebook: <http://www.facebook.com/Fujitsu.DE>
- Fujitsu Bildmaterial und Media Server: https://mediaportal.ts.fujitsu.com/pages/dk_home.php
- Fujitsu Newsroom: <https://www.fujitsu.com/de/about/resources/news/press-releases/>

Pressekontakt:

Fujitsu Central Europe
Melanie Saß
Tel. +49 (0) 174 3091033
Melanie.sass@fujitsu.com

FleishmanHillard Germany GmbH
Lisa Lehrke
Tel.: +49 (0) 174 8130659
Fujitsu-PR@fleishmaneuropa.com

Über Fujitsu

Fujitsu ist ein globaler Partner für die Digitale Transformation und hat sich zum Ziel gesetzt, die Welt nachhaltiger zu gestalten und mit Innovationen das Vertrauen in die Gesellschaft zu fördern. Mit rund 124.000 Mitarbeiter unterstützt Fujitsu seine Kunden in über 100 Ländern, einige der größten Herausforderungen der Menschheit zu lösen. Das Dienstleistungs- und Lösungsportfolio für nachhaltige Transformation basiert auf fünf Schlüsseltechnologien: Computing, Networks, AI, Data & Security und Converging Technologies. Im Geschäftsjahr 2022 (zum 31. März 2023) erzielte Fujitsu Limited (TSE:6702) mit Hauptsitz in Tokio, Japan, einen konsolidierten Jahresumsatz von 3,7 Billionen Yen (28 Milliarden US-Dollar) und bleibt, gemessen am Marktanteil, das führende Unternehmen für digitale Dienstleistungen in Japan. <http://www.fujitsu.com/>

Fujitsu in Deutschland, Österreich und der Schweiz

Fujitsu unterstützt als führender Informations- und Telekommunikations-Komplettanbieter seine Kunden bei allen Aspekten der digitalen Transformation. Dafür kombiniert das Unternehmen IT-Dienstleistungen und Produkte mit zukunftsweisenden digitalen Technologien – wie Künstlicher Intelligenz (KI), dem Internet der Dinge (IoT), Analytics, Digital Annealing, Container Technologie sowie Multicloud-, SAP-, ServiceNow- und Sicherheitslösungen – und schafft zusammen mit seinen Kunden und Partnern neue Werte. Das Produkt- und Service-Angebot kann passgenau auf die individuellen Bedürfnisse zugeschnitten werden – von der Konzeption bis hin zur Implementierung, dem Betrieb und der Orchestrierung von digitalen Ökosystemen. In Deutschland, Österreich und der Schweiz hat Fujitsu über 4.000 Beschäftigte und verfügt mit über 10.000 Channel-Partnern allein in diesen drei Ländern zudem über eines der leistungsfähigsten Partnernetzwerke der Branche. <http://www.fujitsu.com/de/>