

Entwicklung multilingualer Sprachmodelle

Sprachverarbeitung hat durch die Verbreitung von Open-Source-Sprachmodellen erheblich an Bedeutung gewonnen. Das 2022 gestartete BMWK-Projekt OpenGPT-X zielt auf den mehrsprachigen Ausbau dieser Technologie. Da die meisten verfügbaren Benchmarks zur Evaluierung von Sprachmodellen überwiegend für die englische Sprache existierten, will das OpenGPT-X-Konsortium die Sprachvielfalt verbessern – für eine gerechtere und effektivere Sprachtechnologie. Dazu führte das OpenGPT-X-Team umfangreiche multilinguale Trainingsläufe durch und testete die entwickelten KI-Modelle anschließend u. a. auf logisches Denken, Commonsense-Verständnis, Multitasking-Lernen, Wahrheitsgehalt und Übersetzungsfähigkeiten. Jetzt hat das Projektteam das multilinguale “European LLM Leaderboard“ veröffentlicht, eine Rangliste, die eine Reihe multilingualer Sprachmodelle mit jeweils rund 7 Mrd. Parametern vergleicht. Es ist geplant, über dieses Leaderboard die Evaluierung von Modellen aus der zentralen KI-Plattform Hugging Face zu automatisieren, um die Vergleichbarkeit und Reproduzierbarkeit der Ergebnisse zu gewährleisten. Das ZIH der TU Dresden stellt dem Projekt die notwendige Infrastruktur für die umfangreichen Evaluierungen bereit. Weitere Informationen: tu-dresden.de/zih/die-einrichtung/news/european-llm-leaderboard-of-opengptx. (Kontakt: [Klaudia-Doris Thellmann](mailto:Klaudia-Doris.Thellmann@zih.scaDS.AI), ZIH/ScaDS.AI)

Datenkompetenz für Universitätssammlungen

Das Verbundprojekt „SODa – Sammlungen, Objekte, Datenkompetenzen“ dient dem Aufbau eines Datenkompetenzzentrums für die über 1.200 wissenschaftlichen Sammlungen an deutschen Universitäten und Hochschulen. Als einzigartige Wissensressource für aktuelle sowie zukünftige Forschungsfragen stellen diese Sammlungen zugleich einen bedeutenden Datenspeicher mit enormen Potenzialen dar. SODa vermittelt die für die Forschung in und mit wissenschaftlichen Sammlungen notwendige Datenkompetenz. Darüber hinaus entwickelt es diese transdisziplinär sowie praxisbezogen weiter und vernetzt relevante Akteure. Unter der Leitung der Koordinierungsstelle für wissenschaftliche Universitätssammlungen an der Humboldt-Universität zu Berlin wird das Zentrum gemeinsam mit der FAU Erlangen-Nürnberg sowie dem Germanischen Nationalmuseum Nürnberg aufgebaut. Der Fokus liegt auf der Diskussion und Fortentwicklung von Standards, bewährten Praktiken und neuen Verfahren des sammlungsbezogenen Forschungsdatenmanagements sowie

der Datenanalyse und datenbasierten Sammlungs- und Objektforschung. SODa ist eines von insgesamt 11 Datenkompetenzzentren, deren Aufbau durch das BMBF im Rahmen der Datenstrategie der Bundesregierung und des BMBF-Aktionsplans Forschungsdaten seit November 2023 für 3 Jahre gefördert wird. Weitere Informationen: sammlungen.io. (Kontakt: [Martin Stricker](mailto:Martin.Stricker@hu-berlin.de), HU Berlin)

Effiziente I/O-Operationen für Exascale-Computing

Mithilfe eines globalen, hochauflösenden Erdsystemmodells untersucht das Projekt „TOols für raPides und effizientes IO für Exascale (TOPIO)“ die Lese- und Schreibgeschwindigkeiten bei der Verarbeitung großer Datenmengen auf Höchstleistungsdateisystemen. Ein Fokus liegt auf der Datenkomprimierung, um die Datenmenge zu reduzieren ohne wesentlichen Informationsverlust zu verursachen. Im Projekt werden zwei Hauptaspekte der I/O-Operationen optimiert: Die Menge der Daten, die geschrieben wird, und die Effizienz dieses Vorgangs. Ersteres wird über Kompressionsmethoden verbessert. Die Effizienzsteigerung wird über einen Auto-Tuning-Ansatz verfolgt, der die Komplexität mehrerer I/O-Schichten vor den Benutzern verbirgt und gleichzeitig die I/O-Leistung verbessern soll. Dabei wird das Model for Prediction Across Scales (MPAS) verwendet, das für hochauflösende, globale Vorhersagen über Zeiträume von zwei bis drei Monaten eingesetzt wird. Die Simulationen erzeugen Daten im Umfang von mehr als 100 Terabyte, deren Speicherung und Schreibzeit erheblichen Ressourcenaufwand erfordern. TOPIO hat das Ziel, diese Datenmenge bei einem maximalen Informationsverlust von einem Prozent deutlich zu reduzieren und die Schreibleistung zu optimieren. Die Kooperation zwischen dem HLRS und der Universität Hohenheim wird innerhalb von SCALEXA vom BMBF gefördert. Weitere Informationen: hlrs.de/de/projekte/detail/topio. (Kontakt: [Sophia Honisch](mailto:Sophia.Honisch@hlrs.de), HLRS@GCS)

10-Qubit-Prototyp im Projekt QSolid entwickelt

Die Entwicklung von Quantenbits (Qubits) mit minimalen Fehlerraten gehört aktuell zu den größten Herausforderungen des Quantencomputings. Fokus des Projekts QSolid unter der Leitung von Professor Frank Wilhelm-Mauch am Forschungszentrum Jülich, ist die Entwicklung eines hochwertigen Quantencomputers „Made in Germany“ mit supraleitenden Quantenchips. In der ersten Projektphase von Januar 2022 bis jetzt haben mehr als 160 Expert:innen aus 25 Partnereinrichtungen Schlüsseltechnologien entwickelt, die zu einem funktionsfähigen 10-Qubit-Prototyp führen. Dieser wird in Kürze in Betrieb

gehen und niedrige Fehlerraten, einen integrierten Software-Stack sowie einen Cloud-Zugang über die JUNIQ-Plattform des JSC bieten. Bis Dezember 2026 soll das System so weit ausgebaut werden, dass es bis zu 30 Qubits mit größtmöglicher Fehlerkorrektur kontrollieren kann. Das Projekt wird vom BMBF im Rahmenprogramm „Quantentechnologien – von der Grundlagenforschung zum Markt“ gefördert. Weitere Informationen: go.fzj.de/jsc-qsolid-projekt-erfolgreiche-halbzeit. (Kontakt: [Kristel Michielsen](mailto:Kristel.Michielsen@GCS), JSC@GCS)

Offene Ausschreibungen von FFplus

Das vom HLRS koordinierte Projekt “Fortissimo Plus“ (FFplus) fördert die Nutzung von HPC und KI bei kleinen und mittelständischen Unternehmen (KMUs) sowie Start-ups. Um die geschäftlichen Vorteile dieser Technologien aufzuzeigen, werden europäische KMUs und Start-ups dazu eingeladen, Anträge für 2 Arten von Projekten – Geschäftsexperimente und Innovationstudien – einzureichen, die die Vorteile von HPC und generativer KI demonstrieren. Die zur Umsetzung von unabhängigen externen Expert:innen ausgewählten Anträge erhalten eine Anschubfinanzierung und fachliche Unterstützung. Für die erste von 6 geplanten Ausschreibungen können noch bis 4. September Anträge eingereicht werden. FFplus wird von der EuroHPC Joint Undertaking innerhalb des Programms „Digitales Europa“ seit Mai für einen Zeitraum von 48 Monaten mit einem Gesamtvolumen von 30 Mio. Euro gefördert. Weitere Informationen: ffplus-project.eu/. (Kontakt: [Sophia Honisch](mailto: Sophia.Honisch@HLRS), HLRS@GCS)

AISSyS 2024 Special Track über HPC-KI-Konvergenz

Das HLRS ruft Forschende dazu auf, Vorträge, Poster und Demonstrationen für die Konferenz AISSyS 2024 in Venedig (29. September bis 4. Oktober) einzureichen. Ein Special Track widmet sich dem Zusammenspiel von Höchstleistungsrechnen und künstlicher Intelligenz (KI). In diesem Format können aktuelle Forschungsergebnisse sowie -ansätze eingereicht werden, die den Einsatz von KI für die Verbesserung der Systemleistung bzw. für komplexe Modellierungs- und Simulationsaufgaben aufzeigen. Organisiert wird der Track von Dennis Hoppe und Professor Dr. Michael Resch (HLRS). Themenschwerpunkte sind unter anderem KI-integrierte Simulationen, verteilte Datenverarbeitung, KI-gesteuerte Optimierungstechniken sowie die Cybersicherheit in HPC-Systemen. Die Einreichungsfrist endet am 14. August. Weitere Informationen: hlrs.de/de/news/detail/aufruf-zur-einreichung-von-beitraegen-aisys-2024-special-track-ueber-hpc-ki-konvergenz. (Kontakt: [Sophia Honisch](mailto: Sophia.Honisch@HLRS), HLRS@GCS)

Women in HPC: Neues Chapter NHR|WHPC

Auf Initiative von Vertreterinnen der 9 NHR-Zentren wurde im Mai das Chapter “National High Performance Computing Alliance (NHR) (Germany)” in die Organi-

sation “Women in HPC“ (WHPC) aufgenommen. Das NHR|WHPC-Netzwerk hat als Ziel, eine breite und starke Gemeinschaft im NHR-Verbund sowie im wissenschaftlichen Umfeld in Deutschland aufzubauen. Dies umfasst insbesondere die Unterstützung von Frauen und Minderheiten sowie die Förderung von Vielfalt, Gleichberechtigung, Inklusion und einer ermutigenden Kultur. NHR|WHPC richtet sich in erster Linie an die Beschäftigten und Nutzenden der NHR-Zentren, steht jedoch allen offen, die im wissenschaftlichen HPC-Bereich in Deutschland tätig sind und die NHR|WHPC unterstützen möchten. Weitere Informationen und Registrierung: nhr-verein.de/whpc. (Kontakt: [Ayesha Afzal](mailto: Ayesha.Afzal@FAU), NHR@FAU)

Bewerbungsfrist für das NHR-Stipendienprogramm

Die gemeinsame Graduiertenschule der 9 Zentren des NHR-Verbundes – die NHR Graduate School – vergibt Promotionsstipendien an junge Wissenschaftler:innen im Bereich HPC. Die aufgenommenen Promovierenden werden im Rahmen des Stipendienprogramms ab dem 1. April 2025 für bis zu 3 Jahre mit einer monatlichen Förderung in Höhe von 2.220 Euro unterstützt. Zusätzlich zu der Betreuung an einem NHR-Zentrum vermittelt ein Gastaufenthalt an einem weiteren NHR-Zentrum Einblicke in die Arbeitsweise des NHR-Verbundes. Die jährlich stattfindende Summer School sowie die NHR-Konferenz ermöglichen den Promovierenden zudem Anschluss an eine bundesweite, exzellent vernetzte und erfolgreiche Wissenschaftsgemeinschaft. Der NHR-Verbund setzt sich für Chancengleichheit und Diversität ein und ermutigt alle qualifizierten Masterabsolvent:innen, sich bis zum 15. September für ein Stipendium zu bewerben. Weitere Informationen: nhr-verein.de/die-graduiertenschule. (Kontakt: [scholarship@nhr-verein.de](mailto: scholarship@nhr-verein.de), NHR-Verein)

Veranstaltungen

- 27.08.2024: [R on HPC Systems](#), HKHLR, online
- 02.–04.09.2024: [Node-Level Performance Engineering](#), NHR@FAU
- 03.09.2024: [Proficiency Training High Performance Computing: Batch Job Scheduling](#), HKHLR, online
- 04.–06.09.2024: [Introduction to Parallel Programming with OpenMP](#), NHR@FAU, online
- 05.09.2024: [Data Analysis and Plotting in Python with Pandas](#), JSC@GCS, online
- 09.–12.09.2024: [Deep Learning and GPU Programming Workshop](#), LRZ@GCS, online

HPC-Kalender der Gauß-Allianz:
hpc-calendar.gauss-allianz.de