Press release

Dallas, November 17th

Bloom : one of the largest open and multilingual NLP model awarded as the "Best HPC collaboration" by HPC Wire Readers' and Editors' Choice Awards

BigScience's global collaborative effort to develop Bloom, one of the largest, open and multilingual NLP (Natural Language Processing) model in the world, has been recognized in the annual *HPCwire* Readers' and Editors' Choice presented at the 2022 International Conference for High Performance Computing, Networking, Storage, and Analysis (SC22), in Dallas, Texas. This award of the HPC Best collaboration of the year was given to GENCI, IDRIS and HuggingFace teams together.

BigScience was a community adventure as well as a research and engineering challenge. It gathered more than 1200 researchers from academia and industry (startups, SMEs, large groups) from 38 countries with the goal to develop and train BLOOM using a public HPC infrastructure, the Jean Zay supercomputer of GENCI (Grand Equipement National de Calcul Intensif) hosted and operated at IDRIS (Institut du développement et des ressources en informatique scientifique, CNRS).

Orchestrated by Hugging Face, the open-source AI start-up, 30 working groups set to work between mid-2021 and mid-2022, addressing all the different steps of the building of a such large language model (LLM) such as data governance, choice of input data and sources, modeling, evaluation of the model, engineering including optimization and scaling of the model, generalization, ethical AI and legal frameworks, introducing the ROOTS open multilingual dataset and the RAIL open AI license.

The final and biggest version of BLOOM with 176 billion parameters over 70 layers learned from a total amount of 1.61 terabytes of text spanning 46 natural languages and 13 programming languages. The engineering working group attained state-of-the-art throughput with this Transformer-based model on the latest nVIDIA A100-80 partition of Jean Zay supercomputer (offering more than 400 A100 GPUs over the >3100 of the total configuration).

With the support of experts from IDRIS, Hugging Face, Microsoft and nVIDIA (using the DeepSpeed-Megatron framework), the model reached a sustained performance of 156 TFlops/GPU (50% of the FP32/BF16 peak performance).

The training of BLOOM-176B took 3.5 months, with 1,082,990 compute hours over 48 Jean Zay nodes, requiringa total power consumption of 433 MWh representing a carbon footprint of only 25 tons CO_2 eq emissions.

BLOOM is available openly upon a RAIL (Responsible AI Licenses) that limits potentially harmful use-case that BLOOM could enable. More information here:
https://huggingface.co/bigscience/bloom
https://arxiv.org/abs/2211.05100

From the right to the left : Pierre-François Lavallée (IDRIS), Tom Tabor (HPCWire), Stéphane Requena (GENCI)

About GENCI

Created by the public authorities in 2007, GENCI is a major research infrastructure. This public operator aims to democratize the use of digital simulation through high performance computing associated with the use of artificial intelligence, and now quantum computing to support French scientific and industrial competitiveness.

GENCI is in charge of three missions:
- to implement the national strategy for the provision of high-performance computing resources, storage and processing of massive data associated with AI technologies for the benefit of French open scientific research in conjunction with the three national computing centers
- support the creation of an integrated HPC ecosystem at the national and European levels
- promote digital simulation through HPC to academic research and industry scale

GENCI is a civil company, 49% of which is owned by the French government, represented by the Ministry of Higher Education and Research, 20% by the CEA, 20% by the CNRS, 10% by Universities represented by France Université and 1% by Inria.

About the CNRS

The French National Center for Scientific Research is one of the most recognised and renowned public research institutions in the world. For more than 80 years, it has continued to attract talent at the highest level and to nurture multi-disciplinary and interdisciplinary research projects at the national, European and international levels. Geared towards the public interest, it contributes to the scientific, economic, social and cultural progress of France. The CNRS is above all 33,000 women and men, more than 1,000 laboratories in partnership with universities and other higher education institutions bringing together more than 120,000 employees and 200 professions that advance knowledge by exploring the living world, matter, the Universe, and the functioning of human societies. The CNRS ensures that this mission is carried out in compliance with ethical rules and with a commitment to

professional equality. The close relationship it establishes between its research missions and the transfer of acquired knowledge to the public makes it today a key player in innovation in France and around the world. Partnerships with companies are at the heart of its technology transfer policy, and the start-ups that have emerged from CNRS laboratories bear witness to the economic potential of its research. The CNRS provides also access to research findings and data, and this sharing of knowledge targets many audiences: scientific communities, the media, decision-makers, economic players and the general public. For more information: www.cnrs.fr

About Hugging Face

Hugging Face are the creators of leading open-source ML libraries such as Transformers, Datasets Diffusers, and maintain the largest model sharing platform for practitioners and researchers alike. We're on a mission to democratize machine learning, one commit at a time. At Hugging Face, we build open source resources to empower people eager to easily integrate AI into their products and workflows. We are convinced AI itself can be accessible, optimized and responsible. Come discover our models, datasets, services and join our community!

Press and media contact
Nicolas Belot I Chief communication officer I GENCI I + 33 (7) 60 99 95 10 I nicolas.belot@genci.fr