Institute of SCIENCE TOKYO

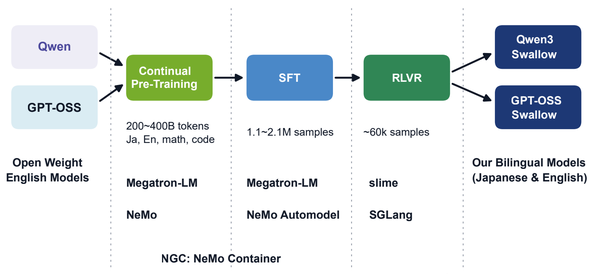
Supercomputing Research Center, IIR

Applications on TSUBAME4.0

Deep Learning and Large Scale CFD

Large Language Models and AI Hardware for Science

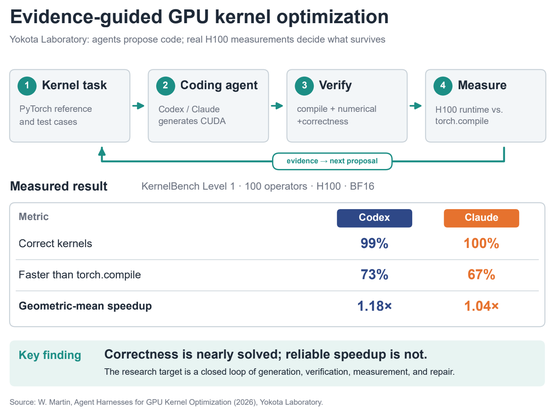
Swallow: Open Japanese LLMs with Reasoning



Swallow adapts open English-weight models to Japanese by continual pre-training on 200-400B tokens of Japanese, English, math and code, then supervised fine-tuning and RL with verifiable rewards.

In 2026 we released the reasoning models Qwen3 Swallow (8B, 30B-A3B, 32B) and GPT-OSS Swallow (20B, 120B), after Llama 3.3 Swallow 70B and Gemma-2-Llama Swallow (2B, 9B, 27B). Training uses Megatron-LM, NeMo, slime and SGLang on TSUBAME4.0 and other national GPU systems. The project has published 132 models and 19 datasets with more than 2.4 million model downloads.
https://huggingface.co/tokyotech-llm

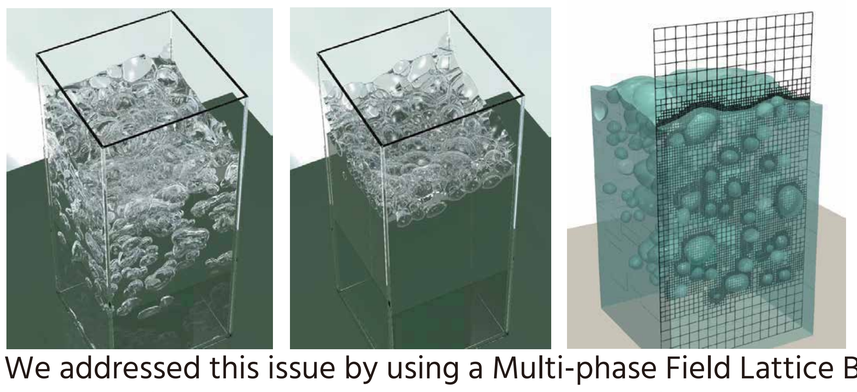
Can AI Agents Write Fast GPU Kernels?



Coding agents propose CUDA kernels; a dedicated H100 decides what survives by compiling, checking numerical results and timing against torch.compile. On the 100 single-operator problems of KernelBench Level 1 (BF16), two agent environments produced 99% and 100% correct kernels, but only 73% and 67% were faster than torch.compile, with geometric-mean speedups of 1.18x and 1.04x. For a single FP32 GEMM, 36 generate-verify-measure-repair attempts gave 25 correct kernels and a best of 574 TFLOPS (78% of cuBLAS). Correctness is nearly solved; reliable speedup is not, so the measurement loop itself is the research target.

Large-scale Mesh-based and Particle-based Simulations

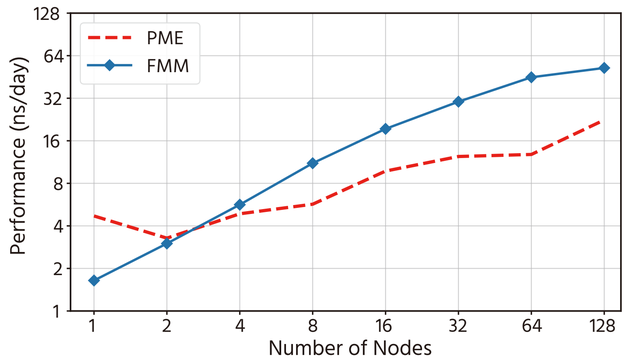
A Large-scale Foam Simulation using a Multi-phase Field LBM and AMR



Foam formation with a stable thin liquid film is very difficult to simulate using conventional methods due to the limitation of mesh resolution that can be used.

We addressed this issue by using a Multi-phase Field Lattice Boltzmann Method and Adaptive Mesh Refinement. Using the Multi-Phase Field LBM, we can prevent a “numerical coalesce” phenomenon that leads to bubble break-up. Adaptive Mesh Refinement has been introduced so that the thin liquid film can be simulated efficiently. Herein, we demonstrate a simulation of foam formed from 200 air bubbles using our proposed method.

From PME to FMM: Long-range Forces in GROMACS



Particle Mesh Ewald needs all-to-all FFT communication, which limits strong scaling of molecular dynamics. The Fast Multipole Method exchanges less data with farther nodes, and its most expensive operator, M2L, maps to dense matrix products on Tensor Cores.

We integrated a PU-accelerated FMM into GROMACS, reusing its tuned short-range kernels and synchronizing tree rebuilds with pair-list updates. On 1-128 Grace Hopper nodes of Alps (4 GH200 each), FMM overtakes GROMACS' PME from 4 nodes and reaches up to 3.52x speedup at scale for a 44-million-atom water box, within GROMACS' energy-conservation tolerances (Huang et al., SC26).

TSUBAME Grand Challenge

History of TSUBAME Grand Challenge

The TSUBAME Grand Challenge solicits proposals for grand challenge problems that can utilize all nodes of TSUBAME4.0 and has two categories.

Category A: Exclusive use of all nodes for 24 hours
Category B: Exclusive use of 1/3 of the nodes for up to 1 week.

Number of Accepted Proposals

	2026		2025		2024		2023		2022		2021		2020		2019		2018		2017		2016		2015		2014		2013		2012		2011		Total
	S	F	S	F	S	F	S	F	S	F	S	F	S	F	S	F	S	F	S	F	S	F	S	F	S	F	S	F	S	F	S	F	
Category A	0	0	0	0	1	0	0	0	0	0																							
Category B	2	0	1	0	2	0	1	0	1	0	0	0	0	0	0	0	1	0	0	2	1	1	2	1	2	1	1	0	2	2	4	3	24
Total	2	0	1	0	3	0	1	0	1	0	2	0	1	0	2	1	2	0	1	1	1	2	3	1	2	2	1	1	0	0	0	2	32
											2	0	1	0	2	1	3	0	1	3	2	3	5	2	4	3	2	1	2	2	4	5	56

We started this program in 2011 and have run the Grand Challenge every year since. In total 56 proposals have been accepted (24 in Category A, 32 in Category B); some of them were awarded the Gordon Bell Prize.

Grand Challenge Projects Accepted in FY2026

Two Category B projects were accepted. Each Category B project receives exclusive use of about one-third of TSUBAME4.0 for one week.

May 25 - June 1

7B Open Uniform Diffusion Language Model Large-scale Pretraining
Mengyu Ye, Tohoku University

A seven-billion-parameter open diffusion language model trained at scale as the May Category B project.

June 22 - 29

Sampling for Markov-state Models of Non-standard Cyclic Peptides
Yutaka Akiyama, Science Tokyo

The June Category B project reserved 82 nodes for seven days to establish a sampling protocol for cyclic peptides.

Source: FY2026 Grand Challenge selection and operation records (Science Tokyo).

https://www.isct.ac.jp/