Towards Democratized Academic LLM Hosting

A 4TU Joint Workshop on Local Inference Infrastructure (hosted at TU Delft)

Final Report

1 Project Team

Name	Position	Affiliation
Azza Ahmed	Senior HPC Engineer; lead contact	REIT, TU Delft
Thomas Abeel	Associate Professor	REIT, TU Delft
Avishek Anand	Associate Professor	REIT, TU Delft
Sven Warris	WR Researcher	Applied Bioinformatics, Bioscience, WUR
Nick Brummans	ML Application Support Engineer	Supercomputing Center, TU/e
Paul Jansen	Product Manager, AI Hub	SURF

The workshop was supported by additional scientific, technical and operational staff across the participating institutions. Only the main named contributions are listed here.

2 Short introduction

Academic LLM hosting is an increasingly important topic for institutions seeking greater control over data, infrastructure and AI services. 4TU NIRICT funding was used to support a full-day workshop at TU Delft (Mondai), where focus was on aligning and strengthening the Dutch community working on Large Language Model (LLM) inference infrastructure.

This was the first time Dutch practitioners and researchers could align on the practical aspects of self-hosting at scale. Discussions throughout the day were lively, as siloed hosting initiatives across Dutch institutions (eg, Delft, TU/e, VU, WUR, SURF) were connected, with ample opportunities for inspiration, knowledge exchange, and collaborations. A key focus was linking infrastructure choices (eg, utilization, deployment models) to electrical engineering priorities in hardware efficiency, power-aware design, and sustainable operations. It laid the groundwork for a more coordinated, sustainable Dutch ecosystem around responsible open LLM hosting.

3 Activities

- Workshop was on 16 March 2026, at Mondai House of AI (Delft)
- 100 registrations, 70 on-site attendees:
    - 81 from Academia & Research (including HBO and Specialized research institutes), 15 from Industry & Infrastructure partners, and 4 from Public sector & Applied research
- 2 Keynote speakers
    - Alexandre Strube (Helmholtz AI) – Lessons from the Blablador project (a mature LLM hosting platform at JSC, Germany)
    - Julio Alexandrino de Oliveira Filho (TNO) – National perspectives and GPT-NL
- 2 breakout rounds (6 parallel sessions in total)
    - Energy consumption: Benchmarking approaches; power-aware optimization strategies
    - Governance & Operational Sustainability: Service creation pathways; Financing and maintenance models; Security, privacy, and access control; Architectural quality assurance
    - Tooling around model hosting: Monitoring and observability; guardrails and safety layers; RAG and MCPs; Service abstractions to end-users
    - Accelerate production-ready LLM deployment using commercial solutions: Why is this relevant: Possible use cases; NVIDIA AI Enterprise; Microsoft AI Foundry

    - Education & Research Use Cases: Practical implementations; researcher-education feedback loops; implications for curricula; Policy & ethics considerations
    - Infrastructure & Deployment: Utilization strategies; deployment models: on-prem, shared, and hybrid
- Panel discussion: Future of AI & LLM/model hosting in the Netherlands and Europe
    - A moderated discussion with academic leaders and national stakeholders exploring how academic institutions should position themselves in a rapidly evolving landscape shaped by national and European initiatives, open and proprietary models, emerging AI hubs, and increasing policy attention. Particular emphasis will be placed on the quality, governance, and sustainability of LLM services in an academic context.
    - Panelists included: Corry Wouters (TU/e), Arie van Deursen (TU Delft), Damian Podareanu (SURF), Niels Taatgen (RUG), Thijs van der Plas (WUR); and moderated by Avishek Anand (TU Delft)
- Networking reception

4 Achieved results and impact

- The workshop brought together a strong multi-institutional mix of participants from Dutch universities, research support teams, national initiatives, and a few industry/ infrastructure partners
- The workshop enabled concrete & practical exchange on academic LLM hosting across infrastructure, energy, governance and sustainable service design, tooling and research/education use cases
- The discussions were consistently described as technical, practical, and well-run, with strong engagement across the keynotes, breakouts, and panel.
- A clear takeaway was that there is substantial demand for continued exchange and coordination on institutionally governed LLM services. Several participants explicitly said that one day was not enough and expressed interest in continuing this conversation in a future edition, possibly rotating host institute.
- The workshop also laid the groundwork for a post-event synthesis/ white paper, drawing on structured breakouts outputs and cross-institutional discussions.
- More broadly, the event showed that academic LLM hosting is no longer a niche topic, but an emerging area of shared infrastructure practice and coordination across institutions.

5 Public communication & visibility

- Public event page:
    - TU Delft listing: Towards Democratized Academic LLM Hosting - Mondai
    - TU/e listing: 4TU Workshop: Towards Democratized Academic LLM Hosting
    - 4TU listing: Workshop Towards Democratized Academic LLM Hosting
- LinkedIn posts (after the even):
    - TU Delft: https://www.linkedin.com/posts/mondai-house-of-ai_llmhosting-ai-llm-activity-7443227887694143488-4NyB?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAdmXw0BLoI9gHSig8Q4SzJ_mOozvGAzdWw
    - TU/e https://www.linkedin.com/posts/nick-brummans-5a308830a_the-tue-supercomputing-center-for-ai-hpc-activity-7444325697122254848-0S6O/
- Workshop photos: https://surfdrive.surf.nl/s/8qjAPwAzco59QfX