

AI Community Meeting

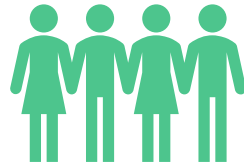
April 2026



Today's Topics



Recent updates and
changes



Transparency and
Communication



Decisions & Future
Goals

Recent updates

- Released Chat AI v0.9.3 – mostly QoL improvements
 - Fixes issues #136, #122, #120, #117 on Github
 - Preparing new features for speech mode (demo)
- New models
 - Qwen 3.5 Series
 - Anthropic Claude Opus and Sonnet
- Removed older models as planned
 - Qwen 3 Series
 - Llama 3.1 SauerkrautLM

Transparency

We have competing demands:

- **Offering the best SOTA models**
 - GLM 5.1
 - Qwen 3.6
 - Anthropic Claude Opus
- **Performance & Reliability**
 - More model instances, scalability
- **Long-term Support & Stability**
 - Keep older models for longer
 - Change on regular intervals, not immediate
 - Support researchers, specific models

We also have limitations:

- **Costs for external models (Claude Opus)**
- **Latest Hardware**
 - We have many GPUs, but some models requires high-end hardware
 - E.g. GLM-5.1 requires 8xH200 GPUs
- **Personnel** – Rotating models involves testing

We cannot have everything at once

Moving Forward

- Our priorities are:
 - Long-term support, stability for researchers and users
 - Performance and reliability, increasing scalability of infrastructure and models
- This means, we cannot quickly provide the best and newest models!

Upcoming Changes

- Overnight benchmark on specific node
 - This may cause one model to temporarily be unavailable overnight
 - Will not affect our service during the day
 - More details in upcoming announcement (Matrix channel, mailing lists)
- Upgrading to our new infrastructure
 - Improve scalability and performance
 - Token-based accounting and institution-wide limiting
- Next Chat AI version
 - Introduce speech options, more customization
- New models! **Qwen 3.6 30B A3B, Gemma 4, GPT-5.5 ...**

Summary and Overview

- Monthly community meetings, usually last Wednesday at 14:00
 - Event page: <https://gwdg.de/hpc/events/chatai/>
- Model voting takes place in the weeks before community meetings
- Results, decisions, and goals are announced in community meetings
- Join our channels and mailing lists:
 - <https://matrix.to/#/#AIComm:uni-goettingen.de>
 - <https://listserv.gwdg.de/mailman/listinfo/ai-saia-users>
 - <https://listserv.gwdg.de/mailman/listinfo/ai-chat-users>
- GöAID - <https://gwdg.de/hpc/events/goeaid/>

Thank you for your attention