Exploring and Participating in Decentralized AI Systems

Sungje Han

Supervised by Professor Jiasi SHEN

THE DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Introduction

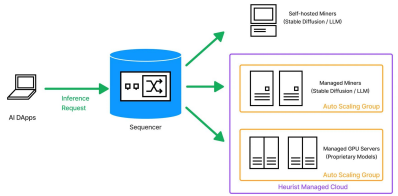
This project investigates the operational frameworks, scalability, and economic models of these decentralized systems, emphasizing their potential to democratize AI infrastructure and address critical issues of transparency, inclusivity, and monopolization in the AI landscape.

We participate and explore the following two protocols: 1. Decentralized inference protocol: Heurist protocol. 2. Collaborative learning protocol: Prime Intellect.

Objective

Answering these questions: What are some of the common frameworks/architectural designs of decentralized AI? And how can decentralization improve AI and what are the advantages and disadvantages of a decentralized AI network?

Analysis of Framework



- Model Creators: Upload model to the in-network Model Registry and receive a portion of the fees
- Application Integrators: Entities that develop interfaces that consumers can access leveraging its tech stack
- Miners/Model Hosts (us): individuals owning GPUs who host models and run inference tasks

Testing

OpenHermes-2 Pro Mistral 7B:
Average total latency: ~54 ms
Average compute latency: ~51.5 ms

BrainDance:
Average total latency: ~4.2 ms
Average compute latency: ~2.5 ms

Throughput Last 24 Hours:
Text tasks: 205,522
Image tasks: 30,920

Conclusion

Decentralized AI systems like Heurist and Prime Intellect demonstrate the feasibility of AI development that reduces reliance on monopolistic control by large tech companies. They offer opportunities for smaller players and individuals to participate and benefit from cutting-edge technologies while preserving transparency.