Big Brain Bigger Appetite

Sentient, Datathon 2025

OpenDeepSearch
– but (try to) make it better

(Left-to-right) David Hofer (MSc Cyber Security), Yi-Yi Ly (MSc Data Science), Arvid Ban (MSc Computer Science), Frederieke Lohmann (MSc Data Science)

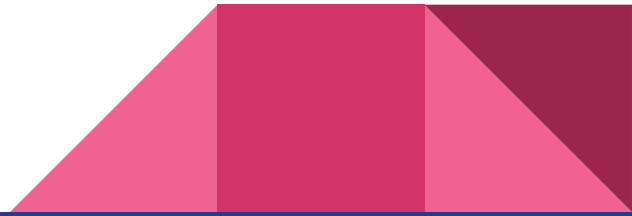
Exploration Steps

General attack plan

Our approach consisted of

1. extensive brainstorming and research of possible methods
2. implementation/hacking
3. evaluating on a small subset of the FRAMES dataset
4. reiterate at step (1)

Comparison to our own baseline (FRAMES subset):
llama-v3p1-70b-instruct to grade our results -> 52.7% accuracy



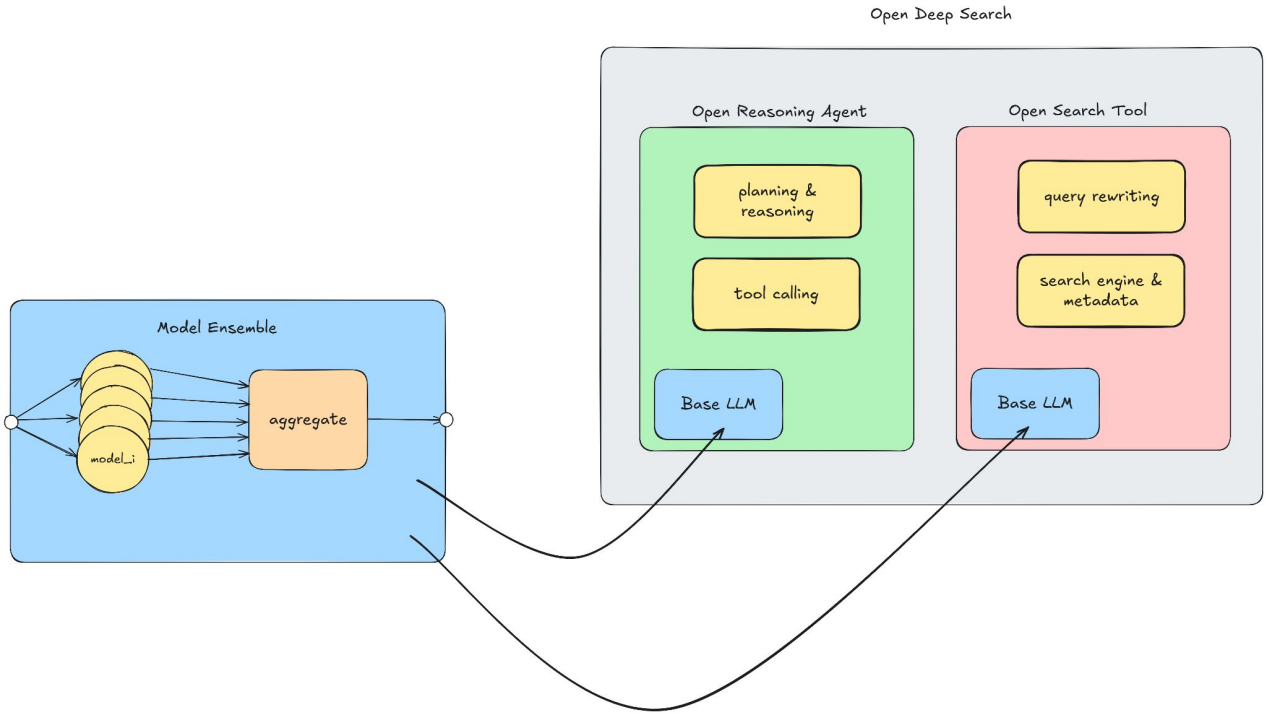
Accuracy metric

- acc = no. of grades “A” / total no. of samples
- for own validation, we used fixed smaller subset of FRAMES
  - size 243
  - size 88
- final FRAMES dataset size: 824 queries

What did we explore?

1. Ensemble Method and Aggregation
2. Query Rephrasing/Augmentation
3. Planning Strategies

Agent Architecture



Submission Results

Final Take-Aways

THANK YOU VERY MUCH!

(Left-to-right) David Hofer (MSc Cyber Security), Yi-Yi Ly (MSc Data Science), Arvid Ban (MSc Computer Science), Frederieke Lohmann (MSc Data Science)