Assignment 5:
Mads-Stack: Optimize the Full Stack

Overview

- Generalize the harness and optimize for the full stack
- Share your learning from assignments

- Leaderboard
  - Terminal Bench (all tasks)
  - 26 hidden tasks picked from SWE-bench Pro
- Blog post
- Video

Generalize and Improve Your Harness

- Generalize your Assignment 4 agent for Terminal Bench
  - The harness should be runnable on Terminal-Bench and SWE-bench Pro
- Improve the full stack
  - Continue on improving the harness (We will have a sharing session next Wednesday)
  - (Optional) Change the model
    - Different models, quantization, etc.

Leaderboard

- 50% of assignment score
  - You will **get full score** as long as you meet the baseline
    - Baseline: 20 tasks within 2 hours, or 30 tasks within 6 hours
- 5 point on the course bonus
  - Only submissions that meet the baseline are qualified
  - Score depending on the round (number of unique ranks)

Your round	The 26 tasks	Terminal-Bench 2.0
1	3	2
2	2	1
3	1	0
4 or later	0	0

Blog Post and Video

- Show your understanding
- Write a blog post on the course blog
  - Covers Assignment 2–5
  - Includes: each kind of techniques you implemented, your lessons, mistakes and etc.
- Video
  - 3-minute video with your face and real human interaction

Notices

- The leaderboard will stop receiving submissions right after the deadline
  - Submitted requests will be graded