

# surge-ai/dayjob evaluation harness

## PRIMARY

|                  |                                                                                     |
|------------------|-------------------------------------------------------------------------------------|
| <b>Publisher</b> | GitHub                                                                              |
| <b>Published</b> | 2026-10-01                                                                          |
| <b>Retrieved</b> | 2026-10-05                                                                          |
| <b>Cite key</b>  | github2026surgeaida                                                                 |
| <b>URL</b>       | <a href="https://github.com/surge-ai/dayjob">https://github.com/surge-ai/dayjob</a> |

**Assessment.** The harness. The one place that states what the grader actually is.

Evaluation harness for DAYJOB. Tasks distributed on Hugging Face under the Surge AI account.

Grading uses "Claude Opus 4.8 as an agentic LLM judge (OpenCode) running inside the task container". Tasks follow Harbor format and run through the OpenHands SDK. Licence: Apache-2.0.

The README carries no task counts, scores or developer counts — those come from the paper and the boards.