Cover letter: AI evaluation and test engineering

Zain Dana Harper | Kent, Washington | harperz9.github.io/cv.html | github.com/HarperZ9

Dear hiring team,

I measured which guards stop an AI agent from making a destructive tool call. On 28 held-out tasks, a check that compares the request with the action caught 64 of 66 harmful calls, and a schema check caught 1 of 66. The tools were mocked, and the test is merged, not yet released. That is the work I am applying to do: AI evaluation, security evaluation and test engineering.

I test whether a check can be fooled, and I publish the answer when it can. Flywheel, my evaluation harness, is at release 1.9.0 on PyPI. Its behavior check flagged 20 of 20 planted changes and 0 of 50 unchanged pairs. My findings page lists 30 results, and 12 of them are nulls.

Other maintainers accept my fixes. They have merged 41 of my pull requests in 34 projects, DeepEval, Pydantic AI and Datasette among them, counted November 14, 2026. Another 17 engineering pull requests were closed without merge, and the full list is public. I have programmed for about thirteen years. I have not yet held a paid software role or a paid red-team engagement. For eleven years I worked in commercial tree care, scoping jobs and watching rigging and clearances from the ground.

On your team I want to build the planted faults and controls that show whether an evaluation can pass a wrong result, and to write up each finding so someone else can rerun it. If the role has a hard requirement, ask me about it directly and I will answer with the evidence or with the gap.

Zain Dana Harper

Written November 14, 2026. Figures and their sources: harperz9.github.io/career/source-ledger.md