Hackathon Technical Summary

What We Added

- Browser-side gesture control using **MediaPipe Hand Landmarker** (GPU with CPU fallback).
- Three pose slots (`0/1/2`) mapped to Harbor actions, plus secondary-hand modifier gestures for alternate routes.
- A Harbor bridge that converts pose events into `window.agent` / `window.ai` actions with scoped permission checks.
- Local model routing with `ollama/llama3.2` **default** and quick fallback to `ollama/phi3.5`.
- Voice integration through local ElevenLabs endpoints for **TTS**, **STT**, and **live conversation**.

Harbor Integration (Technical)

- New event pipeline: `pose match -> trigger engine -> harbor:gesture-trigger -> HarborSdkBridge`.
- Added action catalog + per-pose config for read/summarize, screenshot analysis, research workflows, MCP tools, and voice actions.
- Added runtime capability checks and safety gating (`confirm_each` or `cooldown + arm`).

Multimodal + Voice Path

- Screenshot actions capture active-tab images and emit screenshot events for downstream consumers.
- Summary path remains reliable on text-only models by combining screenshot metadata with page readability text.
- Local ElevenLabs proxy (`server.mjs`) exposes: `POST /api/elevenlabs/tts`, `POST /api/elevenlabs/stt`, `POST /api/elevenlabs/agent/create`, `GET /api/elevenlabs/agent/signed-url`, `GET /api/elevenlabs/health`.

Outcome

Gestures now act as a practical control surface for Harbor agents: fast input, explicit permissions, switchable local models, and integrated voice/multimodal execution.