Gavin Daher

Redmond, WA · gavindaher@gmail.com
linkedin.com/in/gavin-daher-2aa32422b · github.com/largeorbitalobject · portfolio.orbital-network.com

I build oversight for AI agents that do real engineering work. At Denali I founded the Physical AI Team and built Summit, a harness where Claude runs my team's ML pipeline but can't mark its own results as verified. In my own lab I ran a 9-day agent research program of 564 experiments under pass rules written before any results came in. I want to do empirical research on the oversight and control of agentic systems.

Experience

Denali Advanced Integration **Redmond, WA, June 2021 to present**

Founder and Lead, Physical AI Team *July 2025 to present*

Started the team in July 2025 and hired it up to five people in eight months. We build industrial vision systems, from cameras and edge hardware to trained models, and run Denali's internal AI program.

- Built Summit, the harness Claude uses to run my team's ML pipeline. The server enforces the checks: Claude's results stay marked "claimed" until Summit re-measures the metric, and agents can propose labels, but training only uses labels a person has accepted. Every 2026 model line went through it (513 models, 571 evals and about 301,000 auto-labeled samples in five months). Its audits caught a training canary that never trained and VLMs that ignored the image.
- Took Denali from zero to 220+ weekly active Claude users in 2026, running licensing, the move to Claude Enterprise and workshops for 9 departments. My team built Denali's MCP connectors and a plugin marketplace with guardrails. In a 371-run ablation the guardrails cut harmful agent runs from 4 of 82 to 0 of 82 and raised task success from 0.73 to 0.93.
- Ran customer vision projects from pricing to delivered model, including a returns classifier that reached pilots at two logistics customers. On one customer's synthetic-data model I rebuilt the train/test splits to remove leakage and reported the real-only score of 0.942 instead of the leaked 0.979.
- Cisco Partner Innovation Challenge: 3rd worldwide in 2025 ($50K) with the returns-classifier demo I built, and 1st of 111 entries from 29 countries in 2026 ($200K plus a $25K sustainability prize) for a packing-station system I scoped into 15 build tasks.
- Co-author of Rainier-VL, Denali's first public model, a 2B vision-language model for industrial inspection on edge GPUs. I ran the pre-registered program, with blind regrades, that produced the base model the release was trained from: localization mIoU went from 0.07 to 0.48 and defect detection from 0.29 to 0.46.

Automation Architect *November 2023 to July 2025*

- Used GPT-4 at work from April 2023 and tried generative models on line clearance, a detection task the company had only handled with trained models. A Claude 3.5 Sonnet prototype got all 60 test images right at about 2 seconds each (GPT-4o took 9). I then moved the work to small local VLMs on edge hardware (Moondream, YOLO, ONNX), and that demo became the starting point for later customer projects.
- SolidWorks, FEA and NVIDIA Omniverse design of frames, fixtures and install drawings for warehouse automation at Fortune 500 customers.

Business Analyst II, Finance and MIS *June 2021 to November 2023*

- Ran vendor EDI integrations (7 shipped, including Dell US and HPE) and built finance reporting in Dynamics 365 and Power BI, including the CFO dashboard.

Independent research

Personal lab **November 2024 to present**

I use Minecraft world generation and mods as a low-stakes testbed for running AI agents on long research and engineering tasks. Write-ups are at portfolio.orbital-network.com.

- Ran a 9-day agent research program (August 9 to 17, 2026). A coding agent wrote the code, trained the models and ran and wrote up 564 experiments. I set the question and the limits, and committed the pass rules before any results. Blinded comparisons caught the agent choosing its own control arm as the example to present, checks of saved worlds found water inside a cave that the runtime counters had passed, and 74 runs stay filed as failures. The task was learning cave shapes from 12 real surveyed caves. Four neural approaches failed on held-out caves, an interpretable model won on all 12, and the result ships as an 83 KB Minecraft server mod.
- Built the agent platform my lab runs on: 27 scheduled agent jobs, a GPU lease broker and a hash-chained audit log past 17,000 events (2,900 runs, 92 recorded human decisions). Agents have to ask before any destructive, outward-facing or credential action, and I approve it. Permission tiers, tool deny-lists and sandboxing enforce that, and each overnight run has a written scope.
- Led a team of up to about 18 concurrent agents building a 3D map mod (Java, OpenGL) from first commit to closed beta in 9 weeks, with 172 hash-pinned release candidates, a 9,500-test release gate and a 4-machine test harness the agents drive over MCP. I tested every build and turned down an agent's 82% load-time gain because it made revisited areas look worse.
- Fine-tuned a terrain diffusion model and ran two pre-registered research programs on it (about 90 GPU-hours). A LoRA fixed the coarse terrain, turning a blurry, mostly-ocean sample into ridges and fjords. Deeper fixes failed. River networks stayed far from Earth's statistics, and lower training loss came with 5× worse realism. Both programs stopped on their pre-committed rules, and an audit withdrew 4 of 5 headline claims. What remains is a drainage-realism benchmark calibrated on 58,000 real-Earth patches.
- Fine-tuned Qwen3.5-9B and Gemma 4 with LoRA on conversations written by Claude and filtered automatically, to give three original characters their own voices.

Publications and patents

Inventor, US utility patent application "Machine Learning of Merchandise Condition" (pending; assigned to Denali Advanced Integration).

M. Sudharsanan, H. Puppala, G. Daher, J. Rosenbloom. "Rainier-VL: A Compact Vision-Language Model for Industrial Inspection Across Edge GPUs." *IEEE EdgeCom 2026*, regular paper (to appear).

Skills and education

AI agents: Claude Code (skills, plugins, MCP servers, hooks, subagents), Codex, multi-agent orchestration, agent permissions and sandboxing, guardrail ablations, LLM-as-reviewer pipelines.

Machine learning: VLM fine-tuning (SFT, GRPO, DPO, LoRA), diffusion models, evaluation design, pre-registration, paired and bootstrap statistics, serving with vLLM, llama.cpp, TensorRT and ONNX.

Systems: Python, FastAPI, React, Java, SQL, Linux, Proxmox and LXC, Docker, systemd, CUDA and ROCm.

Education: University of San Diego, 2018 to 2022. Business Analytics, Knauss School of Business, after two years in the Shiley-Marcos School of Engineering.