Member of Technical Staff
Robotics

Location: Bay Area, CA · Type: Full-time · Team: Nvex R&D (AlcheEval · AlcheSims)

About Alchedata

Alchedata is building the self-improving data layer for Physical AI. Robots fail in the real world in ways no lab benchmark predicts. Our platform, Nvex, turns those failures into evidence: it diagnoses why a model failed, decides what data or environment would fix it, and proves whether the next checkpoint is actually better before it is promoted.

Frontier models can now label video, build simulation twins and generate synthetic data. We use them as labor inside the loop. Our job is the part they cannot do for a customer on their own: governed, auditable evidence that a fix worked on that customer's robots, tasks and real-world data. In short, frontier models can generate the fix; Nvex proves it worked.

About The Role

We are looking for a Member of Technical Staff to build the research and engineering core of how robots get better after deployment. You will own problems end to end across benchmarks, simulation environments and data, Real2Sim and Sim2Real, and self-evolving Physical AI models, and your work will ship directly into Nvex's AlcheEval and AlcheSims modules.

This is a high-ownership, early-team role. You will define problems as much as solve them, work directly with the CTO and research team, and see your results used in real customer model decisions and published at top robotics and ML venues.

What You Will Work On

Benchmark development

- Design benchmarks and evaluation suites built from real deployment failures, not only curated lab tasks
- Build failure taxonomies, success and progress metrics, and regression gates that decide whether a candidate checkpoint is promoted
- Evaluate a range of policies, including VLAs, frontier-model-driven policies and hybrids, under one consistent interface
- Develop self-evolving benchmarks that grow as new failures are observed in the field

Simulation environments and data

- Build scalable pipelines that generate replayable simulation environments from text, images, video and verified assets
- Author tasks, scenes and sim-ready assets in Isaac Sim / Isaac Lab, MuJoCo, Genesis, SAPIEN or similar
- Generate and curate synthetic data, and measure which synthetic data actually improves a policy and which does not
- Support cross-simulator transfer and cross-embodiment replay

Real2Sim

- Reconstruct real-world scenes and failures into simulation, using 3D reconstruction, Gaussian splatting, physical parameter identification and model-assisted generation
- Verify fidelity against real robot logs, so a failure reproduced in sim behaves like the one seen on hardware

Sim2Real

- Close the gap between simulation results and real-robot performance through domain randomization, system identification and calibration
- Validate fixes on real robots, and feed what you learn back into environments and benchmarks

Self-evolving Physical AI models

- Develop methods that let robot policies and world models improve from their own evaluation signal: test-time adaptation, online policy distillation, imagined experience and reward modeling
- Train and post-train VLM-based judges, reward models and world models that score rollouts and locate failures
- Turn research results into reliable components of the Nvex improvement loop

Minimum Qualifications

- MS or PhD in Robotics, Computer Science, Machine Learning or a related field, or equivalent industry experience
- 3+ years of hands-on work in robot learning, robotics simulation or embodied AI
- Strong Python and PyTorch; comfortable with large-scale training and evaluation pipelines
- Depth in at least one simulator (Isaac Sim / Isaac Lab, MuJoCo, Genesis, SAPIEN), including environment and asset authoring (URDF / USD)
- Experience with at least one of: imitation learning, reinforcement learning, VLA models or world models
- Ability to design rigorous experiments and explain clearly why a result does or does not hold

Preferred Qualifications

- First-author publications at CoRL, RSS, ICRA, CVPR, ICCV, NeurIPS, ICLR or similar
- Hands-on Real2Sim experience: 3D reconstruction, Gaussian splatting, digital twins or physical parameter estimation
- Sim2Real transfer of manipulation policies to real robots such as Franka, UR or ALOHA
- Experience building robotics benchmarks, or with LeRobot, LIBERO, RoboCasa or similar
- Experience with reward modeling, VLM-as-judge, test-time training or world-model post-training
- Familiarity with ROS / ROS 2 and real robot data collection
- Experience at an early-stage startup or a research lab shipping to real users

What Success Looks Like In Your First Six Months

- A benchmark built from real failures is running inside AlcheEval and used in at least one real checkpoint decision
- At least one new environment-generation or Real2Sim route is validated against real robot logs
- A self-evolving method you developed measurably improves a policy on our internal benchmark

Why Join

- Work on the core open problem of Physical AI: how robots keep getting better after they meet the real world
- Real robots, real deployment failures and real customer decisions, not only lab benchmarks
- Publish: we encourage papers and open releases under our PEARL research program
- Early-team ownership and meaningful equity