Physical AI Model Training & Harness Research Engineer

Alps Lab | Research Engineer | Physical AI Models and Harness Optimization

Hiring in China, Singapore, North America and Europe · Remote

About the team

Alps Lab is building recursively self-improving Cognizers with scientific judgment, advancing model architectures native to Physical AI and hardware-software co-design. We aim to help models perform more complex tasks on robots and other physical platforms, and improve continuously through experiments, evaluation and feedback from real-world operation.

The role

You will own model training, data and evaluation for embodied tasks, while building a Harness that continually improves the research process. We are looking for someone who can turn research ideas into reproducible experiments, understand why results occur, and use tools, skills and workflows to make the next round of experiments better.

For this role, the Harness is the research runtime that coordinates data processing, training experiments, tool use, context and evaluation. You will improve both the task capabilities of Physical AI models and the efficiency of discovering and validating those improvements.

Responsibilities

- **Train models for embodied tasks.** Design model architectures, training objectives and data mixtures for perception, physical prediction, planning and action generation. Explore World Models, World Action Models (WAMs) and Vision-Language-Action (VLA) models, taking experiments from training through closed-loop task evaluation.
- **Build reusable data pipelines.** Work with images, video, language, robot states and action trajectories. Address temporal alignment, coordinate systems, action representations, data quality and differences across embodiments. Understand the composition, usage boundaries and training splits of datasets such as Open X-Embodiment, DROID and BridgeData V2.
- **Establish trustworthy task evaluations.** Use appropriate benchmarks such as LIBERO, CALVIN and RoboCasa alongside real-robot tasks to measure task success, long-horizon completion, generalization and robustness. Maintain held-out tasks, regression tests and failure cases, distinguishing model improvements from evaluation bias.
- **Enable recursive improvement of the research Harness.** Use experiment traces to propose changes to skills, context, tool interfaces, task decomposition and workflows. Validate gains with fixed-model comparisons and held-out evaluations. Keep effective versions and records of unsuccessful changes, allowing the improved Harness to propose and test the next round of improvements.
- **Bring on-device constraints into model design.** Understand limits on GPU memory, compute, bandwidth, power and real-time performance. Work with deployment engineers to validate architectural choices on Jetson Orin, Jetson Thor or other target platforms, and analyze trade-offs between task quality and operating cost.
- **Improve experiment quality and research velocity.** Maintain training code, experiment tracking and reproducible configurations. Conduct ablations, error attribution and cost analysis. Write clear technical records of what works, why it works and what is most valuable to test next.

Qualifications & application

What you bring

- Hands-on experience with multimodal language models (MLLMs/VLMs), with direct responsibility for training, data processing, debugging and evaluation. You can investigate issues across models, data and systems in depth.
- An understanding of how World Models, WAMs and VLAs are modeled, their applicable tasks and their limitations, with demonstrable experimental or engineering results in at least one of these areas.
- Proficiency in Python, PyTorch and Linux, and the ability to run and diagnose GPU training jobs independently. Familiarity with distributed training, performance profiling and experiment versioning.
- Familiarity with robot training data and Physical AI benchmarks. You can explain how data distributions, task protocols, cross-scene splits and simulation-to-reality gaps affect conclusions.
- An understanding of in-context learning, context construction, memory and tool use, and the ability to design controlled experiments that establish their practical value in embodied tasks or a research Harness.
- Familiarity with feedback-driven skill and Harness optimization, including the ideas behind Microsoft SkillOpt and Sakana AI/UC Berkeley RHI. You can implement the full iteration cycle: propose a change, run evaluations, compare results, and retain or roll back the change.

Nice to have

- Experience with real-robot data collection, modeling across robot embodiments, long-horizon tasks or sim-to-real transfer; experience with ROS 2, Isaac Lab, MuJoCo or LeRobot.
- Projects involving automated experiments, agent workflows, reusable skill libraries or Harness optimization, with sustained gains demonstrated on held-out tasks.
- Experience in hardware-aware model design, training-efficiency optimization, open-source contributions or high-quality research.

How we measure impact

Reproducible capability gains on held-out tasks and in real environments, while reducing the experiment time and compute cost required to achieve a useful improvement. We value clear research judgment, reliable implementation and the ability to identify the next step from failures.

How to apply

Send your CV and 1-2 projects, code repositories, papers or experiment reports that best represent your work. Describe the task, data and evaluation protocol, your specific contribution, and an instance where failure analysis led you to improve a model or Harness.

We welcome candidates who demonstrate their capabilities through research, industry or open-source projects. You do not need to cover every nice-to-have.

contact@alps-lab.com

Related approaches: Microsoft SkillOpt · Recursive Harness Self-Improvement

Physical AI
模型训练与 Harness 研究工程师

Alps Lab | Research Engineer | Physical AI Models and Harness Optimization

招聘地区：中国、新加坡、北美、欧洲 工作方式：远程

关于团队

Alps Lab 正在构建具有科研判断能力、可递归自改进的认知体，推动 Physical AI 原生模型架构与软硬件协同设计。我们希望让模型在机器人等物理载体上完成更复杂的任务，并通过实验、评测和真实运行反馈持续改进。

关于岗位

你将负责面向 physical embodiment 任务的模型训练、数据与评测，并构建能够持续改进研究流程的 Harness。我们希望你能把研究想法变成可复现的实验，理解结果背后的原因，再让工具、技能与工作流帮助下一轮实验做得更好。

本岗位的 Harness 是组织数据处理、训练实验、工具调用、上下文和评测的研究运行框架。你需要同时提高 Physical AI 模型的任务能力，以及发现和验证这些改进的效率。

你将负责

- **训练面向具身任务的模型**。围绕感知、物理预测、规划与动作生成，设计模型架构、训练目标与数据配比；研究 World Model、World Action Model（WAM）与 Vision-Language-Action（VLA）模型，并完成从训练到闭环任务评测的实验。
- **建设可复用的数据管线**。处理视觉、视频、语言、机器人状态与动作轨迹，解决时序对齐、坐标系和动作表征、数据质量及跨形态差异；熟悉 Open X-Embodiment、DROID、BridgeData V2 等数据集的组成、使用边界和训练划分。
- **建立可信的任务评测**。使用 LIBERO、CALVIN、RoboCasa 等适用 benchmark，并结合真机任务，评估任务成功率、长时序完成度、泛化和鲁棒性；维护留出任务、回归测试和失败案例，区分模型收益与评测偏差。
- **让研究 Harness 递归改进**。从实验轨迹提出对技能、上下文、工具接口、任务分解和工作流的修改，用固定模型对照与留出评测验证收益；保留有效版本与失败修改记录，让改进后的 Harness 继续提出和检验下一轮改进。
- **把端侧约束带入模型设计**。理解显存、算力、带宽、功耗和实时性限制；与部署工程师协作，在 Jetson Orin、Jetson Thor 或其他目标平台上验证架构选择，并分析任务质量与运行成本的取舍。
- **提高实验质量与研究速度**。维护训练代码、实验追踪和可复现配置，进行消融、错误归因与成本分析；用清晰的技术记录说明什么有效、为什么有效，以及下一步最值得验证什么。

任职要求与申请方式

我们希望你具备

- 有多模态语言模型（MLLM／VLM）的 hands-on 经验，实际负责过训练、数据处理、调试和评测；能深入定位模型、数据与系统之间的问题。
- 理解 World Model、WAM 与 VLA 的建模方式、适用任务和局限，并在其中至少一个方向有可展示的实验或工程成果。
- 熟练使用 Python、PyTorch 和 Linux，能独立运行和诊断 GPU 训练任务；熟悉分布式训练、性能分析与实验版本管理。
- 熟悉机器人训练数据和 Physical AI benchmark，能说明数据分布、任务协议、跨场景划分以及仿真到真机差异如何影响结论。
- 理解 in-context learning、上下文构建、记忆与工具调用，能设计对照实验验证它们在具身任务或研究 Harness 中的实际价值。
- 熟悉基于反馈的技能与 Harness 优化，理解 Microsoft SkillOpt 和 Sakana AI／UC Berkeley RHI 的思路；能够实现“提出修改、运行评测、比较结果、保留或回退”的完整迭代流程。

加分经历

- 有真机数据采集、跨机器人形态建模、长时序任务或 sim-to-real 经验；使用过 ROS 2、Isaac Lab、MuJoCo 或 LeRobot。
- 有自动实验、agent 工作流、可复用技能库或 Harness 优化项目，能展示留出任务上的持续收益。
- 有硬件感知模型设计、训练效率优化、开源贡献或高质量研究成果。

我们如何衡量贡献

让模型在留出任务与真实环境中获得可复现的能力提升，同时降低获得一次有效改进所需的实验时间和计算成本。我们重视清楚的研究判断、可靠的实现，以及从失败中找到下一步的能力。

申请方式

请提供简历，以及 1-2 个最能代表你能力的项目、代码、论文或实验报告。请说明任务、数据与评测协议、你的具体贡献，以及一次基于失败分析改进模型或 Harness 的经历。

我们欢迎通过研究、产业或开源项目证明能力的候选人，不要求你覆盖所有加分项。

contact@alps-lab.com

相关方法：Microsoft SkillOpt · Recursive Harness Self-Improvement