CHEN TIANQIAO & CHRISSY INSTITUTE

Email : talent@miromind.ai

Join us!

MiroMind 招聘手册

Recruitment Brochure

Foundational Model Researcher

Key Responsibilities

- ❑ **High-Impact Publications:** Publish research in top-tier conferences, file patents, and contribute to the open-source AI community through the release of datasets, models, and code.
- ❑ **Original Research Breakthroughs:** Explore cutting-edge AI research and industry trends to lead impactful, forward-looking research and achieve original breakthroughs.

Basic Requirements

- ❑ Solid theoretical foundation in machine learning, deep learning, natural language pro- cessing (NLP), computer vision (CV), and reinforcement learning (RL).
- ❑ Strong programming skills, proficient in Python and C/C++ under Linux, capable of independently implementing complex deep learning models and system modules, with expertise in debugging and performance optimization.
- ❑ Familiarity with mainstream architectures, including language models (Transformers and variants, Linear Attention), multimodal models (LLaVA-like, native MLLMs), generative models (Autoregressive, DiT), and reasoning models (o1 / PPO).
- ❑ Excellent problem-solving ability, strong teamwork mindset, and effective communi- cation skills.

Preferred Qualifications

- ❑ Ph.D. from a top university in Computer Science, Artificial Intelligence, or a related field.
- ❑ Publications in leading conferences/journals (e.g., NeurIPS, ICML, ICLR, ACL, EMNLP, CVPR, ICCV/ECCV).
- ❑ Outstanding performance in academic competitions (ACM/ICPC, NOI/IOI, CMO/IMO, CPhO/IPhO, etc.).
- ❑ Experience contributing to well-known large-model open-source projects or achiev- ing top results in related competitions.

Foundational Model Research Engineer

Key Responsibilities

- Support the scaling of cutting-edge research into industry-leading next-generation models by providing large-scale training data acquisition, reinforcement learning (RL) environment construction, and extreme training efficiency optimization.
- Build comprehensive and detailed automated evaluation systems for next-generation models to deepen understanding of capability boundaries and guide future research priorities.
- Apply theoretical breakthroughs to real-world product challenges, driving impactful AI applications.

Basic Requirements

- Strong programming skills, proficient in Python and C/C++ under Linux, familiar with PyTorch and mainstream large model training/finetuning frameworks; able to independently implement complex deep learning models and system modules with strong debugging and performance optimization abilities.
- Experience in large-scale data preprocessing, data generation, and data augmentation; understanding of data-driven model iteration workflows.
- Familiarity with large model training pipelines, including distributed training, model parallelism, and training efficiency optimization.
- Excellent problem-solving skills, collaborative mindset, and strong communication skills.

Preferred Qualifications

- Familiarity with high-performance operator frameworks such as CUDA/Triton/Cutlass.
- Experience with distributed RL frameworks such as veRL / OpenRLHF / Ray.
- Knowledge of large-scale RL environment construction for browser / computer use / code sandbox tasks.
- Experience with distributed training frameworks such as Megatron-Core / DeepSpeed, including multi-node training efficiency tuning and optimization of computation-communication overlap.
- Outstanding achievements in competitive programming (ACM/ICPC, NOI/IOI, Code- forces, TopCoder).
- Contributions to well-known open-source large model projects or winning results in related competitions.

Foundational Model System Researcher

Key Responsibilities

- **System Development & Optimization:** Lead the development and optimization of large-scale model training and inference systems. Leverage cutting-edge technologies such as hybrid parallelism, automatic parallelization, high-performance operator development, and communication optimization to significantly improve training speed and efficiency, accelerating model iteration.
- **Tackling Technical Challenges:** Focus on solving complex challenges in machine learning systems, including high concurrency, high reliability, and high scalability. Ensure stable and efficient system operation under diverse scenarios, providing strong technical support for continuous business growth.
- **Comprehensive Coverage Across Domains:** Take responsibility for multiple critical sub-domains of machine learning systems, including resource scheduling, model training, model inference, and reinforcement learning training. Drive overall system performance improvement and functional enhancement.
- **Performance Analysis & Innovation:** Conduct in-depth analysis of performance metrics during large-model training, accurately identify and resolve bottlenecks to maximize training efficiency. Stay at the forefront of emerging machine learning system technologies, actively research and adopt new methods, fully unlock hard- ware potential, and drive continuous innovation and upgrades.

Preferred Qualifications

- **Programming & Framework Skills:** Proficiency in at least one programming language (C, C++, Python) or experience in CUDA development. Familiarity with at least one distributed training framework such as PyTorch FSDP, DeepSpeed, or Megatron-LM. Candidates with awards in international programming competitions (e.g., ACM, ICPC, Codeforces) will be given priority.
- **Technical Solution Excellence:** Ability to design solutions with strict standards across multiple dimensions such as machine performance and system stability, ensuring scientific, rational, and efficient outcomes.
- **Domain Expertise & Passion:** Substantial practical experience and strong interest in one or more of the following areas:
- **Parallel Systems:** Deep research in distributed training of foundation models, efficient fine-tuning, reinforcement learning training, and inference engine optimization, including but not limited to parallel strategy design, quantization & compression techniques, and operator optimization.
- **High-Performance Operators:** Familiarity with parallel computing (e.g., Triton, CUDA), communication technologies (e.g., NCCL, NVSHMEM), and AI compilers (e.g., MLIR, TVM, Triton, LLVM), with relevant development and optimization experience.

Data Project Manager (Data Annotation / Data Operations)

Role Overview

- Responsible for planning and executing MiroMind's data production and annotation projects (text / image / audio / video / multimodal). This includes organizing and managing internal and external annotation teams and vendors, establishing standards and processes, and ensuring the balance of progress, quality, cost, and compliance.

Key Responsibilities

Project Planning & Delivery

- Break down data requirements based on model training goals (task types, scale, coverage, difficulty, and priority). Develop milestones, budgets, and resource plans.
- Manage schedules, risks, and dependencies across concurrent data projects to ensure on-time delivery aligned with training needs.

Team & Vendor Management

- Build and mentor annotation and quality inspection teams (full-time / part-time / crowdsourced / vendors), including scheduling, performance, and incentive management.
- Manage vendor onboarding and evaluation (bidding, SLA, pricing, delivery quality), as well as cost and contract management.

Standards & Processes (SOPs)

- Define and iterate annotation guidelines, label taxonomies, edge cases, and decision trees; maintain operation manuals and case libraries.
- Design layered quality control (self-check, peer review, expert sampling), gold standard sets, and re-review workflows to continuously reduce rework rates.

Quality & Data Governance

- **Establish quality metrics:** gold standard accuracy, IAA (e.g., Cohen's kappa / Krippendorff's alpha), coverage, noise rate, PII leakage rate, etc.
- Apply methods such as active learning, hard example mining, weak supervision, and LLM-as-judge to drive a "data flywheel" for continuous refinement and augmentation.

Data Engineer (AGI-Oriented R&D)

Key Responsibilities

- **Data System Development:** Build large-scale data processing systems to support the training and evaluation of trillion-parameter foundation models, ensuring efficiency, stability, and scalability across the entire data pipeline.
- **High-Quality Data Construction:** Lead the collection, cleaning, deduplication, annotation, and augmentation of data for training foundation models (language, multimodal, agent, etc.), continuously improving data quality and diversity.
- **Intelligent Data Tools:** Develop intelligent tools for data generation, synthesis, filtering, and automated evaluation to accelerate data iteration and closed-loop optimization, supporting model capability expansion and alignment training.

Basic Requirements

- Strong programming skills, proficient in Python/C++, with solid system design abilities and the capability to independently develop large-scale data processing modules.
- Familiarity with data processing and storage frameworks such as Spark, Flink, Ray, or Hadoop, with hands-on experience in building and optimizing data pipelines.
- Understanding of foundation model training workflows and data quality requirements, with awareness of data-driven model iteration and evaluation practices.
- Excellent problem-solving, engineering execution, and teamwork abilities.

Preferred Qualifications

- Experience in constructing training datasets for large models or leading the cleaning and management of million-scale high-quality data.
- Familiarity with data augmentation and synthesis techniques (e.g., Self-Instruct, RLAIF, synthetic QA generation, image-text alignment augmentation), or experience with agent-based data generation.
- Knowledge of web-scale data collection, crawler development, deduplication, information extraction, and web structure parsing.
- Familiarity with interactive log data construction and feedback data mining in reinforcement learning environments.
- Contributions to well-known open-source datasets (e.g., OpenWebMath, RefinedWeb, RedPajama, LAION, COYO) in the form of data processing tools or cleaning strategies.
- Strong performance in competitions such as ACM/ICPC, NOI/IOI, data mining, or data-centric AI.

DevOps Engineer

Responsibilities:

- Participate in the architecture design and core component development of AI training clusters, building high-performance and highly available computing platforms.
- Develop observability systems for training and inference tasks and resources, enhancing cluster monitoring, alerting, and log analysis capabilities.
- Optimize key components such as compute scheduling, RDMA, and container runtimes to ensure efficient and stable execution of training and inference workloads.
- Support large-scale cluster automation for deployment, operations, and troubleshooting, improving system maintainability and availability.

Requirements:

- Bachelor's degree or above from a 211 (or higher) university, in Computer Science, Software Engineering, Electronic Information, or related fields.
- Solid foundation in operating systems; familiarity with Linux kernel, networking, storage, and performance tuning.
- Proficiency in Golang with strong coding ability; Kubernetes development experience is a plus.
- Familiar with cloud-native technologies such as Kubernetes, Docker, Prometheus, and Grafana.
- Experience in operations or development of large-scale distributed systems, with the ability to quickly identify and resolve complex issues.

Full-Stack Engineer (AI Core Team)

Key Responsibilities

- **Full-Stack Development for AI Foundation Model Services**
  - Build highly available and scalable inference services for LLMs/LMMs, including front-end interfaces, back-end APIs, task orchestration, and microservice governance.
- **Data Annotation and Management Platform Development**
  - Design and maintain multimodal annotation platforms (text / image / audio / video), supporting task distribution, quality assurance feedback, dynamic priority scheduling, and visualized monitoring.
- **Model Deployment and Continuous Delivery**
  - Develop DevOps/MLOps pipelines covering containerization, automated testing, canary release, A/B testing, and version rollback, ensuring rapid model iteration and reliable deployment.

Basic Requirements

- Proficient in JavaScript/TypeScript with frameworks such as React/Vue/Svelte; familiar with at least one backend language (Node.js / Python / Go / Java) and common web frameworks.
- Familiar with Docker, Kubernetes, and Helm; understanding of service mesh, auto-scaling, and monitoring/alerting systems (Prometheus / Grafana / Loki).
- Skilled in modeling and optimization with MySQL/PostgreSQL and NoSQL databases (Redis/MongoDB/ClickHouse, etc.); knowledge of vector databases (Milvus / PGVector/ Faiss).
- Proficient with Git/GitHub/GitLab; experienced in unit, integration, and end-to-end testing; capable of producing clear technical documentation and collaborating across teams.

Community Operations Specialist (Community Operations / Growth)

Role Positioning

- Drive continuous improvement of community user **Growth (G) – Activation (A) – Contribution (C) – Retention (R)** through content, events, community mechanisms, and data-driven operations, fostering a positive cycle of research influence and open-source ecosystem growth.

Key Responsibilities

- **Community Growth & Activation**
  - Manage and grow community platforms including Discord, Telegram, Feishu, GitHub, Twitter/X, Zhihu, Jike, Xiaohongshu, Bilibili, and WeChat official accounts. Handle daily operations such as onboarding/welcome, tiered engagement, FAQs, and community guidelines.
  - Design onboarding flows and incentive systems (badges, contributor levels, task boards, event points) to improve 7-day activation and 30-day retention rates.
- **Content & Branding**
  - **Plan technical content:** research insights, model release notes, project roadmaps, case studies, tutorials, weekly reports/newsletters.
  - Coordinate with researchers and engineers to produce articles, short videos, and livestreams; maintain a content calendar and asset library.
- **Events & Ecosystem**
  - **Organize online and offline activities:** paper reading groups, open-source sprints, hackathons, developer salons, and joint events with university communities.
  - Build volunteer/ambassador programs, connect with student clubs, open-source communities, and key opinion leaders (KOLs).
- **Contribution & Conversion**
  - Optimize GitHub contribution pathways (Good First Issues, contribution guides, Code of Conduct) to increase both the quantity and quality of PRs/Issues.
  - Guide active members into becoming core contributors, speakers, partners, or candidates.
- **Data & Tools**
  - Build dashboards (activity, retention, conversion, reputation), conduct A/B tests and retrospectives; create playbooks and SOPs.