

Manager, Machine Learning Research Engineer, Generative AI

Apply

San Francisco, CA

Scale's Generative AI Data Engine powers the most advanced LLMs and generative models in the world through world-class RLHF/RLAIF, data generation, model evaluation, safety, and alignment.

As the Manager of the Generative AI team, you will be responsible for managing and leading a group of talented researchers and engineers. Your primary focus will be to leverage your expertise in LLMs, generative models, and other foundational models to create and execute an AI roadmap which will help Scale accelerate our customers' Generative AI initiatives forward. This is an exciting opportunity to work on cutting-edge technologies and collaborate with industry-leading professionals.

We are building a large hybrid human-machine system in service of ML pipelines for dozens of industry-leading customers. We currently complete millions of tasks a month and will grow to complete billions monthly.

You will:

- Manage a team of highly effective researchers and engineers. Provide guidance, mentorship, and technical leadership to a team of researchers and engineers working on Generative AI projects. Develop and evaluate methods for integrating machine learning into human-in-the-loop labeling systems to ensure high-quality and throughput labels for our customers.
- Implement and improve on state-of-the-art models developed internally and from the community and put them into production to solve problems for our customers and taskers.
- Work with product and research teams to identify opportunities for improvement in our current product line and for enabling upcoming product lines.
- Work with massive datasets to develop both generic models as well as fine-tune models for specific products.
- Work with customers and 3rd party research groups to understand their goals and define how we can enable them.
- Build a scalable ML platform to automate our ML services, including automated model retraining and evaluation.
- Be able and willing to multi-task and learn new technologies quickly.
- **Must be able to commute to the San Francisco Office 1-2x weekly.**

Ideally you'd have:

- 7+ years of full time work experience using LLM, deep learning, deep reinforcement learning, or natural language processing in a production environment. Especially training foundational AI models through pre-training, fine-tuning, and RLHF.

- A vision for where the field should go and what Scale should do to enable it.
- Strong programming skills in Python, experience in PyTorch or Tensorflow
- Experience with MLOps and the automation of model training & evaluation
- Experience working with cloud technology stack (eg. AWS or GCP) and developing machine learning models in a cloud environment
- Solid background in algorithms, data structures, and object-oriented programming
- Deep appreciation for building high-quality, robust, reusable machine-learning software
- Degree in computer science or related field

Nice to haves:

- Graduate degree in Computer Science, Machine Learning or Artificial Intelligence specialization
- Publication experience in the field or related topics.
- Experience with model optimization techniques for both training and inference

Compensation packages at Scale for eligible roles include base salary, equity, and benefits. The range displayed on each job posting reflects the minimum and maximum target for new hire salaries for the position, determined by work location and additional factors, including job-related skills, experience, interview performance, and relevant education or training. Scale employees in eligible roles are also granted equity based compensation, subject to Board of Director approval. Your recruiter can share more about the specific salary range for your preferred location during the hiring process, and confirm whether the hired role will be eligible for equity grant. You'll also receive benefits including, but not limited to: Comprehensive health, dental and vision coverage, retirement benefits, a learning and development stipend, and generous PTO. Additionally, this role may be eligible for additional benefits such as a commuter stipend.

Please reference the job posting's subtitle for where this position will be located. For pay transparency purposes, the base salary range for this full-time position in the locations of San Francisco, New York, Seattle is:

\$240,000 - \$300,000 USD

PLEASE NOTE: Our policy requires a 90-day waiting period before reconsidering candidates for the same role. This allows us to ensure a fair and thorough evaluation of all applicants.

About Us:

At Scale, we believe that the transition from traditional software to AI is one of the most important shifts of our time. Our mission is to make that happen faster across every industry, and our team is transforming how organizations build and deploy AI. Our products power the world's most advanced LLMs, generative models, and computer vision models. We are trusted by generative AI companies such as OpenAI, Meta, and Microsoft, government agencies like the U.S. Army and U.S. Air Force, and enterprises including GM and Accenture. We are expanding our team to accelerate the development of AI applications.

We believe that everyone should be able to bring their whole selves to work, which is why we are proud to be an affirmative action employer and inclusive and equal opportunity workplace. We are committed to equal employment opportunity regardless of race, color, ancestry, religion, sex, national origin, sexual orientation, age, citizenship, marital status, disability status, gender identity or Veteran status.

We are committed to working with and providing reasonable accommodations to applicants with physical and mental disabilities. If you need assistance and/or a reasonable accommodation in the application or recruiting process due to a disability, please contact us at accommodations@scale.com. Please see the United States Department of Labor's [Know Your Rights poster](#) for additional information.

We comply with the United States Department of Labor's [Pay Transparency provision](#).

PLEASE NOTE: *We collect, retain and use personal data for our professional business purposes, including notifying you of job opportunities that may be of interest and sharing with our affiliates. We limit the personal data we collect to that which we believe is appropriate and necessary to manage applicants' needs, provide our services, and comply with applicable laws. Any information we collect in connection with your application will be treated in accordance with our internal policies and programs designed to protect personal data. Please see our [privacy policy](#) for additional information.*