Senior Staff Machine Learning Engineer AI Foundations

Description

At Cruise, the AI Foundations organization (AIF) builds the essential AI infrastructure used by every ML engineer at Cruise. AIFâ??s goal is to support the AI lifecycle through the ML data platform, the ML training platform as well as deploying large complex ML models to the on-vehicle hardware under tight real-time constraints. AIFâ??s mission is to provide the highest levels of reliability, the fastest developer velocity at the lowest cost footprint through these platforms and services. Cruise is building large foundational models that are customized to solve the autonomous driving challenges. This naturally implies that the ML systems and training infrastructure must scale to the requirements of these large models. Deploying these large models to edge (on-vehicle) hardware is another critical challenge. In this role, as a Senior Staff engineer on this team, you will be leading the vision/strategy as well as tech-leading the execution to solve these hard problems. Cruise is a highly collaborative and dynamic work environment. In this role, as an uber TL, you will be responsible for collaboratively working with partners on the ML teams (Autonomy) to inspire a technical strategy / vision and drive execution through a cross-functional team. WHAT YOUâ??LL BE DOING: Be a strategic thought leader for AI Foundations as a whole Lead strategic execution around ML systems and training infrastructure and deploying large models Be as a technical leader across the broader Cruise AI and Engineering organizations to deliver production systems and mentor Engineers Define execution processes that strive for streamlined engineering development with robust quality and excellence Contribute broadly on the needs and strategy from data, to training, to deployment Utilize your deep understanding of data characteristics, model architectures, optimization techniques, or other ML domain-specific challenges to perform critical analysis of modeling results Propose ways to improve the quality of the data for the model training and evaluation, automate data collection for reproducibility. Closely follow industry and academic developments in the SOTA for the AI life-cycle and ML Systems domainâ??s and adopt technology that is the best fit for Cruiseâ??s needs. Own technical projects from start to finish, contribute to the teamâ??s product roadmap, and be responsible for major technical decisions and tradeoffs. Effectively participate in teamâ??s planning, code reviews and design discussions Consider the effects of projects across multiple teams and proactively manage risks/conflicts. Work together with partner teams and orgs to achieve cross-departmental goals and satisfy broad requirements Conduct technical interviews with well-calibrated standards and play an essential role in recruiting activities. Effectively onboard and mentor junior engineers and/or interns WHAT YOU MUST HAVE: 5+ years working on ML systems for training and deploying complex ML models by employing SOTA techniques in the area of distributed training and deployment Experience working on the scaling challenges associated with building LLMs and/or other large foundational models. Passionate about self-driving technology and its potential impact on the world Attention to detail and a passion for truth A track record of efficiently solving complex problems collaboratively on larger teams Startup mentality – openness to dealing with unknown unknowns and wearing many hats Strong expertise in writing production quality code and setting standards for code quality across engineering teams Experience in driving technical strategy and vision for engineering teams and organizations Leadership experience for planning and execution of cross-functional initiatives and projects In-depth understanding of software development lifecycle (SDLC) and best practices – CI/CD, coding, debugging, optimization, testing, integration and deployment PhD in CS/CE/EE, or equivalent, in industry experience Bonus points! Worked on distilling large foundational models to production on edge devices Strong experience in

Hiring organization
Cruise

Job Location
Remote

Base Salary
$ 60000 - $ 110000

Date posted
June 5, 2024

Apply Now

working with GPUs and GPU based datacenters Relevant publications   The salary range for this position is $217,600 – $320,000. Compensation will vary depending on location, job-related knowledge, skills, and experience. You may also be offered a bonus, long-term incentives, and benefits. These ranges are subject to change.Please mention the word **FLOURISH** and tag RNDQuMjM0LjE0NS4xMTU= when applying to show you read the job post completely (#RNDQuMjM0LjE0NS4xMTU=). This is a beta feature to avoid spam applicants. Companies can search these words to find applicants that read this and see they're human.

Contacts
Job listing via RemoteOK.com