GAUSS LABS

AI Engineer (US)

Palo Alto, CA

Tech – SWAP / Full-Time / Hybrid

At Gauss Labs we are looking for a passionate Machine Learning Engineer who will collaborate with scientists and engineers to improve manufacturing efficiency and reliability in IC manufacturing using data and artificial intelligence.

This role is responsible for implementing batch and online learning capabilities in our machine learning infrastructure. It is also responsible for capturing the lineage of our ML models as well as measuring and monitoring their performance.

Responsibilities

- Design, build, and maintain high availability, multi-tenant training, and inference system for many different content types, such as images and other sensor data

- Work closely with data scientists, micro-service developers, and data engineers in our Applied Research and Product Organizations to ensure efficient transfer of scalable solutions

- Implement seamless tracking of the lineages of data sets, transformations, feature sets, and hyper-parameters used during ML model training and inference

- Enable online learning and automated, continuous delivery of ML models

- Implement and deploy automated auditing systems to identify bias and other potential degradations of model performance

- Adapt and implement AI models to support real-world use-cases and satisfy production requirements as needed

Basic Qualifications

- MS or Ph.D. in quantitative fields (CS, Statistics, Math, or Engineering)

- 7+ years of work experience as MLE, Data Scientist, or related job function

- Strong understanding of ML fundamentals and scaling methodologies (e.g., ability to implement an ML algorithm mathematical formulations in an efficient manner w.r.t. scalability, optimization and so on; understanding of ML loss functions, evaluation/validation methods, statistical testing and so on)

- Strong communication skills both written and verbally

- You are excited to learn, explore new problem areas, and apply your creativity to some of the most challenging and rewarding problems

- Familiarity with all aspects of the model development lifecycle

- Experience working with cloud products (AWS, GCP, or Azure)

- Proficiency in Python, and experience in C++

- Prior experience with MLFlow

- Knowledge of at least one build tooling system

- Track record of writing robust, readable, well documented, well tested, high-performance code.

- Experience with containerization, automated eval/testing, CI/CD, MLOps, and cloud services is a plus