Andreas Schaler

- Full-Stack AI Application Engineer -

andreas.schaler.cs@gmail.com
614 316 8982
https://www.linkedin.com/in/andreasschaler
https://github.com/Andreas3333

Technical Skills

Programming Languages - Python, Typescript, C++, Bash, Django REST Framework, FastAPI, Flask, React, Vue
Artificial Intelligence - Pytorch, Hugging Face, Tensorboard, Supervised Fine Tuning (SFT) for domain specific NLP, Prompt engineering, inference serving, vllm, llama.cpp, RAG systems, sentence-transformers, LangChain, LangGraph
Cloud and IaC - Terraform, AWS CDK, CloudFormation, SAM, boto3, EC2, ECS, EKS, Lambda, SNS, SQS, Event Bridge, S3
Etcetera - Linux, Systemd, D-Bus, Kubernetes, Docker, Podman, Kustomize, Helm, Skaffold, AWS Serverless, Ansible, Packer, GItlab CI/CD, GitHub Actions, Pydantic, Open API Spec., RabbitMQ, Postgres SQL, Vite, uv, Poetry, MkDocs

Employment - 01/2023 - Present

Nimbis Services — DevSecOps Engineer

- Implemented configurable entrypoint start up scripts and multi stage container builds to pre-provision dependency packages, LLM model weights, and set configurations resulting in improved efficiency for service and LLM model artifact packaging and deployments, reducing startup times in development and production by 15%.
- Enabled local and CI/CD persistent container build caches reducing CI/CD Job build times by 18%.
- Implemented IaC and Lambdas to automate network resource configuration reducing deployment times by 1.5 hours
- Extended IaC to support containerized deployments of LLM services and vector databases in AWS.
- Implemented DRF REST services in Domain Driven Design (DDD) patterns for concise internal and external interfaces to support code organization and modularity and Open API Spec.
- Co-Authored, built, packaged, released and deployed Vue 3 component library for traditional SPA web application and REST API deployed in AWS EKS Kubernetes cluster and deployments managed by Helm.
- Developed React based SPA utilizing the AWS Cloud Scape library and extended features.
- Implemented preemptive failure mechanism to fail jobs with missing prerequisite improving job submission by 15% in distributed synchronous multi AWS account system.
- Implemented max concurrency capability enabling full saturation of remote compute resources managed through an asynchronous service manager.
- Implemented Library Helm Charts for configuring and packaging Charts with configurable sub charts.
- Implemented Skaffold profiles and profile requirements to expand and support multiple configurations of services for local development and automated production deployment environments.
- Designed and implemented Event Driven POC solution architecture implementing virus scanner leveraging AWS Event Bridge, S3, ECS, and SQS for monitoring a quarantine S3 bucket containing potentially malicious files.

Projects - 01/2023 - Present

LLM NER REST API — Named Entity Recognition (NER) classifier API

Our small team of 3 completed data acquisition, feature engineering, model evaluation, model training, system design, system implementation, artifact release, and containerized system deployment of a RAG system in AWS. The custom REST NER API leveraged a DeBertav2 classification model which was selected based on its benchmark performance on the CoNLL03 NER dataset.

Personal Finances App — Web application for analyzing personal spending habits leveraging fine tuned base Bert

A SPA and REST API system that allows users to visualize and understand their monthly spending. Transaction data is classified by a domain specific SFT Bert model for multi class sequence classification.

Extension and Adaptation of AWS RES — Fork and extension project

An extension of the AWS Research and Engineering Studio (RES) system. This project involved reverse engineering of the solution, extension of the SPA web application, adding functionality via serverless implementations, altering and extending Cloudformation, building, packaging and publishing of deployment artifacts, extending compute cluster node bootstrapping mechanisms, and implementing hardened automated AMI build processes for cluster node images and VDIs.

Education

Bachelors, Computer Science (CS) Data Engineering Concentration — Kent State University – 01/2020 - 12/2022
Masters, Computer Science (CS) Data Mining and Intelligent Systems Concentration — University of Tennessee Knoxville – (Expected graduation 12/2027)

Certifications

CompTIA, Security + Certification — 03/2025