David William Erwin

daviderwin.me | Prague, Czech Republic | Native English | Fluent Czech

Education

Czech Technical University — Prague, CZ

Bachelor of Science in Electrical Engineering and Computer Science — *Aug. 2022 – June 2025*

- **Thesis: Analysis of State-of-the-Art Vision Transformers for Enhanced Performance (WIP).**
- Analyzed and optimized Vision Transformer architectures (ViT, Swin, DINO) by implementing pruning, attention adjustments, and architectural refinements, reducing computational complexity while maintaining accuracy. Utilized TIMM and +100 GPU clusters to retrain models, achieving improved efficiency across diverse datasets.
- **Relevant Coursework:** Statistical Machine Learning, Pattern Recognition, Computer Architectures, Cybernetics, C Programming, Logic Systems and Processors, Quantum Computing, Probability and Statistics.

Experience

Machine Learning Intern — June 2024 – August 2024

Private Team by Pablo Dylan — *Los Angeles, California, USA*

- Performed data acquisition and transformation using techniques like rolling averages and Z-Scores, improving dataset quality for optimal model training.
- Designed and implemented an efficient, scalable data pipeline enabling seamless data processing and model integration.
- Developed and deployed a custom neural network architecture to address specific financial forecasting challenges, leveraging state-of-the-art machine learning techniques.
- Collaborated with a high-performing dynamic team of engineers in a fast-paced environment, contributing to system design and delivering impactful solutions under pressure.
- Gained hands-on experience in new domains, showcasing adaptability to meaningfully contribute to topics initially new to me, and a commitment to producing high-quality results.

Projects

Inference Server with Infrastructure as Code — October 2024 – Present

Azure Cloud Platform, Terraform

- Designed and deployed an inference server on Azure Cloud Platform to serve custom deep learning models.
- Leveraged Terraform for Infrastructure as Code (IaC) to automate provisioning of cloud resources and ensure cost efficiency.
- Implemented scalable deployment practices to optimize performance and resource utilization.

Oxide Server — June 2023 – February 2024

Rust, Hyper, Docker, ORT

- Developed a lightweight HTTP/1.1 server using the Hyper HTTP and ORT inferencing libraries in Rust, currently used for testing model performance in the cloud on specific hardware.
- The server logic is custom-built without external frameworks, with inference handled manually using a tokenizer.

Technical Skills

Languages: Python, Rust, SQL, C, Kotlin, Go, Mojo, MATLAB
Frameworks: PyTorch, Google Jax, Databricks, PySpark, Kafka, Airflow, ORT.rs, Rocket.rs
Developer Tools: Git, GitHub Actions, Azure/Google Cloud, Relational Databases, Terraform, Docker/Compose, CUDA, Bash/Powershell Scripting, Linux/NixOS, SSH
Libraries: NumPy, Pandas, Seaborn, Hyper.rs, Pola.rs

Certificates

Databricks Certified Data Engineer Associate | *Databricks Cloud* — October 2024

- Validated expertise in building scalable ETL pipelines, managing large datasets with Apache Spark, and optimizing workflows on Databricks.

Machine Learning Specialization | *Stanford Online - Andrew Ng* — September 2024

- Mastered concepts like supervised/unsupervised learning, backpropagation, and gradient descent by implementing algorithms from scratch.