

# CARLOS GUSTAVO SALAS FLORES

Data Scientist

☎ +1 984-837-2465 ✉ [yuseicarlos2560@gmail.com](mailto:yuseicarlos2560@gmail.com) 🔗 [linkedin.com/in/carlogustavosalas](https://www.linkedin.com/in/carlogustavosalas) 🐙 [github.com/cs582](https://github.com/cs582)

## Education

### Duke Kunshan University & Duke University

May 2023

Dual Degree: *B.S. in Data Science & B.S. in Interdisciplinary Studies*

Kunshan, China & Durham, NC

- **GPA:** 3.6/4.0.
- **Awards:** Full Scholarship and Dean's List (2019 and 2020).
- **Leadership:** Pre-Law Society Treasurer (Spring 2023).
- **Relevant Coursework:** Statistical Machine Learning, Advanced Linear Algebra, Databases, and Cloud Computing.

## Experience

### Amazon Web Services

May 2022 – Aug 2022

Software Development Engineer/Data Science Intern

Seattle, WA

- Deployed a customer behavior analysis tool that maintained AWS Lambda availability at 99.99999% in > 20 regions.
- Handled 10PB+ of data, designed a ML pipeline, and identified low availability customers with 97% accuracy.
- Built a data analysis package that saved engineers +100 hours/week (EC2, Docker, S3, and Lambda).
- Wrote research and technical papers for engineers, scientists, and stakeholders.

## Projects

### A Low-Resource Implementation of CLIP

Apr 2023 - May 2023

- Ran extensive tests and achieved >70% zero-shoot classification accuracy on ImageNet and CIFAR-10.
- Trained a 211M-parameter CLIP model that used ViT and GPT as backbones and monitored training using S3.
- Re-designed the model to use 1/3rd of its parameters and still achieved >60% zero-shoot accuracy.
- Designed a data pipeline to create the 10M pairs training dataset by scrapping through 14M Wikipedia articles.
- Implemented unit tests to accelerate development and facilitate error tracing, saved over 60 minutes of work every day.
- Used less than a 100th of the resources used by OpenAI (256 GPUs vs 1 GPU).

### GANs and Diffusion Models Implentation

Mar 2023

- Independently implemented the first papers on GANs and Diffusion models from scratch within five days.
- Trained the models on the MNIST and CIFAR-10 datasets and evaluated their performance matching original results.
- Integrated posterior advances in beta scheduling for diffusion models and backbone architecture for better results.

### Cloud-Native Playlist Recommendation System

Feb 2023 - Mar 2023

- Developed a Cloud-Native Playlist Recommendation System based on user's play history taking ~5s per prediction.
- Trained a Machine Learning algorithm on-cloud and automatically deployed it using Kubernetes and ArgoCD.

### Real-Time Semantic Segmentation for Autonomous Vehicles

Jan 2022 - Mar 2022

- Led a team of four people to implement state-of-the-art Semantic Segmentation algorithms.
- Achieved outstanding accuracy results > 65% on all algorithms.
- Optimized the performance of the models for real-time inference on a low-power embedded device, achieving an average frame rate of 30 fps.

## Research

### Duke Kunshan University

May 2021 – Mar 2023

Research Assistant with Prof. Luyao Zhang

Kunshan, CN

- Prepared financial data from the S&P 500 for algorithmic trading and achieved 122% return on investment.
- Evaluated Deep Learning approaches to statistical arbitrage and optimized running time by more than 50%.
- Designed a pipeline to retrieve hourly data from an API to a Reinforcement Learning algorithm.
- Trained and evaluated ViT and CNN agents to manage 121 different digital assets.
- Demonstrated relative superior performance of ViTs by a 0.4 Sharpe ratio and 133% daily return of investment.

## Duke University

Jan 2022 – May 2022

Research Assistant with Prof. Lisa Gennetian

Durham, NC

- Gathered and cleaned US Census and survey data for policy-making used to improve accessibility to social programs.
- Compiled more than 15,000,000 data points in a database.
- Produced info-graphics and dashboards to convey information to the general public.

## Duke University

Dec 2020 – Mar 2021

Research Assistant with Prof. Ashutosh Kotwal

Durham, NC

- Built an image reconstruction and pattern recognition algorithm to detect Dark Matter particles 1000 times faster.
- Implemented ML algorithms into integrated circuits (FPGAs) to improve the performance of Large Hadron Collider.

## Publications

---

**Salas-Flores, C. G.** (2023). A Low-Resource Formal Algorithm for Text-Image Connections (Under Review). *ICML 2023*.

Zhang, L., Wu, T., Lahrichi, S., **Salas-Flores, C. G.**, & Li, J. (2022). A Data Science Pipeline for Algorithmic Trading: A Comparative Study of Applications for Finance and Cryptoeconomics. *2022 IEEE Blockchain*. DOI: [10.1109/Blockchain55522.2022.00048](https://doi.org/10.1109/Blockchain55522.2022.00048).

## Technical Skills

---

**Languages:** Python, Java, C/C++, R, and SQL.

**AI Research:** Reinforcement Learning, Diffusion Models, and Generative Neural Networks.

**Data Analysis:** Data Visualization, Clustering, Regression, and Time Series.

**Frameworks:** PyTorch, NumPy, SciPy, Pandas, Scikit-learn, Matplotlib, Seaborn, Ggplot, and D3.js.

**Tools:** ArgoCD, Kubernetes, Docker, EC2, ECR, Lambda, S3, Git, and LaTeX.

## Achievements

---

- 2nd Place at Hack Duke 2022, Oct 2022.
- National Finalist in the Alibaba GET Challenge (top 12 out of 250+ teams), May 2020.
- Professional Certificate on AI Engineering by IBM, Aug 2020.
- Fundamentals of Deep Learning for Computer Vision by NVIDIA, Feb 2020.