

Liam Toran

liam.toran@gmail.com • (415) 410-9920 • San Francisco, CA • liamtoran.com • US Citizen • [LinkedIn](#)

WORK EXPERIENCE

Flowcast

San Francisco, CA

Senior Data Scientist

March 2022 - Present

- Led the development of a bank transaction classification system using weak supervision, natural language processing and deep neural networks, accuracies currently surpassing the market-leading provider (Plaid API) by 21% [\[arXiv paper\]](#).
- Trained time series and text RNN embeddings for unstructured transactional data utilizing PyTorch and FastText, resulted in 13% F1 improvement for 9 classification tasks compared to off-the-shelf pre-trained embeddings.
- Generated weak-labels to supervise DNNs, bypassing the bottleneck of massively expensive labeling and annotation.
- Designed the pipeline architecture, implemented correct tools to ensure reproducibility and robustness, all while accomplishing >10x speedups in training and feature engineering steps over the lifetime of the project.
- Interviewed and selected top DS, data engineer, and ML engineer talent who contributed to Flowcast's success.

Data Scientist

January 2020 - March 2022

- Designed and launched a large-scale machine learning model for a top-10 European bank, currently scoring three million clients/month and providing fast and accurate results via PySpark and XGBoost for the last two years uninterrupted.
- Achieved 47% error reduction of an interpretable model currently used by the leading global fashion brand to control a multi-million dollar funnel, through significant rework of feature engineering, boosting and hyper-parameter tuning.

University of Nice Sophia-Antipolis - Research Institute

Nice, France

Applied Mathematics Researcher, Intern, 6 months

2019

- Conducted research on population dynamics modeling for social media networks and organic network growth, utilizing numerical simulation techniques and reaction-diffusion partial differential equations theory.
- Implemented a reaction-diffusion solver with Julia, Python & sparse linear algebra residuals, that increased speeds by 50x.
- Solved and validated through simulations formula for deriving network propagation speed from local variables.

UCSD - Biomedical Research Institute - Knight Lab

San Diego, CA

Machine Learning Researcher, Intern, 7 months

2017

- Applied UMAP, PCA, UniFrac distance and phylogenetic trees to analyze three real world microbiological datasets.
- Conducted research to solve long-time bias in dimensionality reduction, resulting in co-authoring a peer-reviewed [research article](#) (63 citations) and presenting findings to an audience of 55 research scientists at a conference hosted by UCSD.

Inria - National Institute for Research in Technology Sciences

Grenoble, France

Research Engineer, Intern, 3 months

2016

- Developed a new cloth physics numerical solver in Python and C++ resulting in 10x speed over baseline and enabling unprecedented realism for cloth simulation. Referenced in the following peer-reviewed [research article](#).

EDUCATION

École Normale Supérieure de Lyon

Lyon, France

Masters degree in Mathematics, BS Math with Computer Science specialization

2019

- Succeeded a top 0.5% ranking in nationwide scientific exams to enter ENSL (top 3 school in France) with a scholarship.
- Studied ML, Statistics, Linear Algebra, Computer Science, Stochastic Calculus, PDEs, Numerical Simulation, and more.
- Taught undergraduate math and physics lessons to help 6 students succeed in french ivy-league entrance exams.

SKILLS

Machine Learning: Deep Learning, Natural Language Processing, NLP, Weak Supervision, Unsupervised learning, Interpretability and Explainability, Uncertainty, Multi Armed Bandits & Recommendation, Time Series, Classification, Clustering, Metrics, Visualization, Transformers, Large data sets, Imbalance, Neural Networks, Gradient Boosting.

Software Development: Python ecosystem (pandas, numpy, scikit-learn, matplotlib), PyTorch, PySpark, Ray, Git, GitHub, DVC, SQL, Distributed & Parallel computing, Linux, AWS, Azure, TensorFlow, Spark, LaTeX, Excel, R, C++, C.

Hobbies: Technology, DIY, Swimming, Finding innovative solutions, Learning new things!

Current Projects: [flippers](#), an open source library for weak supervision. Learning graph neural network techniques.