

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 pipeline using weak supervision and natural language processing.
- Trained time series and sentences neural network embeddings for our messy transactional data with Torch and FastText, achieving 15-point performance improvement over off-the-shelf embeddings. The classifications are supervised through weak labels generated by Snorkel, circumventing the need for massively expensive labeling and annotation.
- Implemented Ray for hyperparameter optimization and DVC as ML repository backbone. The solutions and project structure I shaped succeeded in greatly improving the models over the lifecycle while being robust, and are now being implemented in a product that surpasses the market-leading provider (Plaid API) in both accuracy and flexibility.
- 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 sanitize a multi-million dollar funnel, through significant rework of feature engineering, boosting and hyper-parameter tuning.
- Led research for a recommendation system platform leveraging contextual multi-armed bandits for Flowcast's app Tillful.

University of Nice Sophia-Antipolis - Research Institute

Nice, France

Applied Mathematics Researcher, Intern, 6 months

2019

- Conducted research on population dynamics modeling in social media networks and organic network growth, utilizing numerical simulation techniques and reaction-diffusion partial differential equations theory.
- Coded a reaction-diffusion solver with Julia, Python & sparse linear algebra min-residuals, that increased speeds by 50x.
- Solved theoretically and numerically validated a formula for network propagation speed using 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 on cloth/body simulation. Referenced in the following peer-reviewed [research article](#).

EDUCATION

ENS Lyon

Lyon, France

Masters degree in Applied 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, Advanced Linear Algebra, Algorithmics, PDEs, Programming, Stochastic Calculus 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, Semi supervised and unsupervised methods, Interpretability and Explainability, Uncertainty quantification, Multi Armed Bandits & Recommendation, Time Series, Classification, Clustering, Metrics, Data visualization, Large data sets, 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.