

# Matthew Francis-Landau

Baltimore, Maryland

4246463355 | matthew@matthewfl.com | matthewfl.com | matthewfl | matthew-francis-landau | Matthew Francis-Landau

## Summary

Looking for Staff Machine Learning (ML) Compiler Engineer that is remote or near Baltimore, Maryland. I have several years of experience building compilers and domain specific languages (DSLs) for ML/AI and have a Ph.D. on compilers and machine learning.

## Education

### Johns Hopkins University, Center for Language and Speech Processing

Baltimore, Maryland

Doctor of Philosophy in Computer Science (Ph.D.)

Sept 2016 - Feb 2024

- Dissertation: Declarative Programming Via Term Rewriting — A Weighted, Declarative, Logic Programming Language for Machine Learning & NLP
- **5 Papers**, Logic Programming for Machine Learning and Natural Language Processing, Approximate Nearest Neighbors, and Libraries for NLP
- **Research/Studied**: Natural Language Processing, Declarative Methods & Programming, Representation Learning, Semantics, Bayesian Statistics, Matrix Analysis, Convex Optimization, Discrete Optimization

### University of California Berkeley

Berkeley, California

Bachelors of Art in Computer Science and Applied Mathematics

Sept 2012 - May 2016

- **Research**: Neural Entity Linking, Automatic recompilation of Java applications for Distributed Computation, Predictable Performance by Gang Scheduling in Xen and the Linux Kernel
- **Classes**: Natural Language Processing, Machine Learning, Deep Reinforcement Learning, Advanced Computer Systems, Advanced Operating Systems, Virtual Machines and Managed Runtimes, Combinatorial Optimization, Data Structures and Randomized Algorithms, Parallel Communication Avoiding Algorithms

## Work Experience

### Cruise / General Motors

Remote

Senior 2 Deep Learning Compiler Engineer / Machine Learning Engineer

May 2024 -

- Contribute to the design & development of a SotA MLIR Deep Learning Compiler for deploying PyTorch Models to Self-Driving Cars
- Extend and Develop functionality of Open Source ML & Compiler frameworks, including PyTorch, Torch-MLIR, MLIR-Tcp, MLIR-TensorRT
- Designed automatic debugging tools that identify bugs in the compiler and runtime stack, saving thousands of engineering hours
- Created MLIR optimization pass that optimizes tensor reshapes and transposes enabling more efficient kernel fusion
- Manage and update critical software infrastructure and manage relationships with third-party vendors
- Create documentation for onboarding models and debugging issues
- **Technical Skills**: C++, MLIR, Python, PyTorch, Dynamo, Bazel, LLVM, TensorRT, Torch-MLIR, MLIR-Tcp, MLIR-TensorRT, Torch-TensorRT, Debugging
- **Soft Skills**: Teamwork, Communication, Problem-Solving, Project Planning, Time Management, Collaboration, Customer Support

### Johns Hopkins University, Center for Language and Speech Processing

Baltimore, Maryland

PhD Researcher

Sep 2016 - Feb 2024

- **5 Papers** on Logic Programming for Machine Learning and Natural Language Processing, Approximate Nearest Neighbors, and Libraries for NLP
- Self-guided research, created research proposals, wrote detailed project specifications, developed implementations, created experiential benchmarks, published research papers
- Created novel term rewriting framework, and formalized it mathematically, and created an efficient implementation
- **Technical Skills**: Clojure, Python, PyTorch, C++, Compiler Implementation, Relational Algebras, Term Rewriting, Domain Specific Languages, Bayesian Modeling, Java, Antlr4, Tensorflow, Eigen, NumPy, FSMs, PyBind11, Slurm, CUDA, Matplotlib, SMT/SAT, MC-Stan, Scala, Truffle/Graal
- **Soft Skills**: Planning, Problem-Solving, Time Management, Teamwork, Communication, Writing, Co-authoring, Presentations, Public Speaking

### Johns Hopkins University, Department of Computer Science

Baltimore, Maryland

Lead Teaching Assistant

Jan 2018 - Dec 2023

- Lead Teaching Assistant for Advanced Natural Language Processing (Deep Learning, Structural Prediction, Sequence Modeling, Language Modeling), Cloud Computing, Object-Oriented Software Engineering, Mathematical Foundations for Computer Science, and Algorithms
- Awarded the Excellence in Teaching, Advising, and Mentoring Award
- Wrote 8 homework assignments on structural prediction with Neural Networks using PyTorch for NLP and Graph/Dynamic Programming Algorithms
- Supervised and trained teams of 20 course assistants
- **Soft Skills**: Leadership, Management, Teaching, Lecturing, Presentations, Conflict Resolution (Customer Relations), Flexibility

### Balanced Payments

San Francisco, California

Software Engineer

May 2013 - Sep 2014

- Designed, Developed, and Deployed Machine Learning credit card fraud model
- Resigned REST API: Created design specification, implementation, and API clients · Developed framework for reconciling bank statements
- **Technical Skills**: Python, PostgreSQL, Scikit-learn, Flask, SQLAlchemy, Git, Github, Docker, Nginx, Node.js, REST, JSON API, Software Development
- **Soft Skills**: Teamwork, Time Management, Problem Solving, Documentation, Customer Support

## Skills

---

**Highly Proficient** C++, Python, C, Clojure, Scala, Java

**Skills** Compilers, Natural Language Processing, Deep Learning, Structural Prediction, Bayesian Statistics, REST API Design, Integer Linear Programming, High-Performance Data Structures, Relational Algebras, Term Rewriting, Optimization

**Technologies** Linux, PyTorch, MLIR, LLVM, Bazel, TensorRT,  $\LaTeX$ , Antlr4, Prolog, Datalog, PostgreSQL, JavaScript, Mathematica, PyBind11, OpenFst, Tensorflow, Node.js, Docker, Eigen, NumPy, SciPy, Scikit-Learn, Git, AWS, Truffle/Graal, Slurm, CUDA, Matplotlib, Lua, SQLite, SMT/SAT/z3, MC-Stan, Nginx, Flask, SQLAlchemy, Theano, Simplex, SQL, OpenMP, Continuous Integration

**Soft Skills** Problem-Solving, Time Management, Teamwork, Communication, Coauthoring, Presentations, Public Speaking, Management, Teaching, Documentation, Conflict Resolution (Customer Relations), Flexibility

## Research Projects

---

### Dyna: Logic Programming for Machine Learning

Baltimore, Maryland

Johns Hopkins University

Sept 2016 – Feb 2024

- Long-running research project to Design and Develop Programming Languages for Machine Learning Applications
- Created design specifications and implementations. Experimented with different designs to achieve the desired result
- Developed JIT compiled prototype using Truffle and Graal Framework
- Created Relational Algebra capable of representing general purpose programs. Created execution engine using Term Rewriting
- Created tracing JIT compiler that merges sequences of rewrites to achieve a 5x performance boost
- Created flexible memoization technique that builds on the relational algebra that can memoize *any* expression

### Exact and/or Fast Nearest Neighbors

Baltimore, Maryland

Johns Hopkins University

Nov 2017 – Oct 2019

- Approximate Nearest Neighbor (ANN) algorithm that generates a *certificate* that *guarantees* the answer is correct
- A certificate is non-probabilistic and generated by intersecting half-spaces in a high-dim vector space using Simplex and Convex Set intersection
- Achieved runtime performance comparable to state-of-the-art probabilistic systems despite the additional certification mechanism

### Redmagic: JIT Compiler for x86 Assembly

Independent Research Project

- Developed Tracing JIT that generates execution traces directly from x86 assembly to create inputs a JIT compiler
- Integrated with cPython using a small 200 line patch and achieved a 2% performance boost over baseline

### Capturing Semantic Similarity for Entity Linking with Convolutional Neural Networks

Berkeley, California

University of California Berkeley

- Neural Net Entity Linker that achieved stated-of-the-art performance by combining neuralized and hand-crafted features

### DJ: Distributed JIT

Berkeley, California

University of California Berkeley

- Developed new techniques for writing distributed programs as if running on a single machine
- Utilizes automatic Bytecode rewriting of Java programs to convert programs from single-machine to multi-machine execution

## Publications

---

Declarative Programming Via Term Rewriting (PhD Thesis)

Matthew Francis-Landau

Johns Hopkins University, Ph.D. Thesis, 2024

Evaluation of Logic Programs with Built-Ins and Aggregation: A Calculus for Bag Relations

Matthew Francis-Landau, Tim Vieira, Jason Eisner

13th International Workshop on Rewriting Logic and Its Applications, 2020

Fast and/or Exact Nearest Neighbors

Matthew Francis-Landau, Benjamin Van Durme

2019

MFST: A Python OpenFst Wrapper With Support for Custom Semirings and Jupyter Notebooks

Matthew Francis-Landau

2018

Dyna: Toward a Self-Optimizing Declarative Language for Machine Learning Applications

Tim Vieira, Matthew Francis-Landau, Nathaniel Wesley Filardo, Farzad Khorasani, Jason Eisner

Proceedings of the First ACM SIGPLAN Workshop on Machine Learning and Programming Languages (MAPL), 2017, Barcelona

Capturing Semantic Similarity for Entity Linking with Convolutional Neural Networks

Matthew Francis-Landau, Greg Durrett, Dan Klein

Proceedings of the North American Association for Computational Linguistics, 2016, San Diego, California, USA

Fine-grained parallelism in probabilistic parsing with Habanero Java

Matthew Francis-Landau, Bing Xue, Jason Eisner, Vivek Sarkar

Proceedings of the Sixth Workshop on Irregular Applications: Architectures and Algorithms (IA<sup>3</sup>), 2016, Salt Lake City