

Arie Selinger

arie.selinger@gmail.com (857) 266-2786

EXPERIENCE

Quilt Research, Inc.

Staff AI Engineer (Founding Engineer)

San Francisco, CA

Mar 2024 - Jan 2025

- Owned the design and development of Quilt's flagship Live Assistant – an event-driven, LLM-powered system that listens to live customer calls (Zoom or browser) and delivers contextual insights and real-time reasoning support
- Led the end-to-end architecture and backend implementation using asynchronous systems (asyncio, websockets) for real-time, event-driven interactions
- Built an internal simulation and replay framework to reproduce multi-actor sessions from recorded events, enabling fast iteration and debugging
- Designed LLM evaluation pipelines and data logging systems to measure performance and prevent regressions
- Integrated Retrieval-Augmented Generation (RAG) into live product flows, to ensure low-latency contextual retrieval
- Collaborated cross-functionally with leadership to align AI research, system design, and product strategy
- Ensured reproducibility and robustness in experiments and production pipelines through strong typing and scalable cloud infrastructure (GCP, Cloud Run)

Tesla Motors

Senior Data Scientist & Software Engineer

Palo Alto, CA

Dec 2021 - Mar 2024

- Designed and implemented a petabyte-scale distributed compute system in Scala with JVM profiling and optimization, reliable Parquet I/O, and idempotent, fault-tolerant task orchestration (Docker, Kubernetes, AWS)
- Ensured reliability and performance of this mission-critical system supporting 1000s of downstream applications, enabling faster, more efficient fleet-wide data processing and delivering multi-million-dollar annual cost savings
- Developed large-scale ETL and analytics pipelines (Airflow, Spark) on fleet-wide telemetry data to improve data quality and analytical accuracy
- Conducted safety and reliability investigations for U.S. governmental agencies
- Led the ML-driven prognostics initiative, designing scalable data and modeling pipelines to predict component failures and improve fleet reliability
- Owned the AI component of a cross-team initiative integrating large language models (LLMs) and delivered live demos to VP-level executives and Elon Musk's direct reports

Juice – Audio companion app providing a personal, dynamic & interactive audio stream (visit: juice.tech)

Paris, France

Co-Founder & CTO

2020 - 2022

- Speech Recognition** – Created a dataset of 3000+ labelled French voice samples | Implemented from scratch and fitted a small-footprint state-of-the-art speech keyword spotting model (KWS) in TensorFlow2 | Deployed the corresponding end-to-end mobile-embedded pipeline in C (using TFLite)
- Text-to-speech** – Trained an artificial human-like French speaking voice using Tacotron2 & WaveGlow
- Mobile App** – Designed, developed and deployed the app architecture and APIs with React Native, Node.js and MongoDB

MIT Sloan / Neon Pagamentos – Brazilian Fintech offering retail banking services

US Remote

Data Scientist Intern – MIT Capstone Project

Summer 2021

- Improving credit default prediction models using customers' relationship data by training the GraphSAGE graph neural-network

MIT Operations Research Center

Research Assistant to Prof. Dimitris Bertsimas

Cambridge, MA

Aug 2020 - Apr 2021

- Developed a robust strategy to generate Congressional Districts while preventing Partisan Gerrymandering
- Designed IP formulations and implemented tailor-made cutting-planes methods with Lazy Constraints in Julia/JuMP
- Computed efficient warm-start solutions by implementing a Local Search heuristic algorithm

xBrain – Startup specialized in Natural Language Processing which builds conversational agents

Menlo Park, CA

Machine Learning Intern, Applied AI Team

Aug 2019 - Feb 2020

- Trained a Bidirectional LSTM with attention for joint intent detection and slot filling in conversational AI, improving chatbot accuracy and user experience
- Collaborated with software engineers to deploy the trained model as a production API
- Used a *topic modeling* approach to segment 100,000+ customer emails for a financial services client

Paris Digital Lab – Selective specialization program in AI and Full-Stack development

Paris, France

Machine Learning & Software Engineering (2-month projects for 3 companies)

Feb 2019 - Aug 2019

- AirLiquide** – Automated detection of gas cylinders illegally sold on a French marketplace with CNNs finetuning
- Randstad** – Trained language models such as ELMo in TensorFlow using 30,000+ resumes provided by the HR company
- Melun Hospital Center** – Provided a web application tool helping to optimize the flow of patients in the Emergency Room

EDUCATION

Massachusetts Institute of Technology

Cambridge, MA

Master of Business Analytics, Operations Research Center & MIT Sloan

2020 – 2021

- Optimization Methods, Machine Learning, Advanced Natural Language Processing, Integer Programming
- Machine Learning project – *Scaling up Holistic Regression computational time by using a Dual Approach*
- NLP project – *Semantic Textual Similarity for News Aggregation using Siamese DistilBERT Networks*

Ecole Centrale Paris (CentraleSupélec)

Paris, France

Master of Science in Applied Mathematics, top-tier French Engineering School

2017 – 2020

- GPA: 4.07/4.33 | Top 5%
- Machine Learning, Deep Learning, Advanced Statistics, Advanced Algorithms, Object-oriented Programming
- Teaching Assistant in Mathematics (Linear Algebra, Analysis, Probability & Statistics)

ADDITIONAL

- **Music** – Passionate violinist since the age of 5, taught by Maestro Igor Volochine at Ecole Normale de Musique de Paris
- **Volunteering** – Took part in a 7-week volunteering retreat in Madagascar, renovating homes and tutoring children
- **Hackathon**
 - Won 1st Prize (\$2000) of a Sony Hackathon involving NFC technologies
 - Invited as a jury member for an AI hackathon at AGI House, alongside Adam D'Angelo (CEO, Quora)
- **Languages** – French (native), Hebrew (basic proficiency)