

Kai Kleinbard

917 693 9960 | bakukai@gmail.com | Brooklyn, NY | Portfolio: <https://bit.ly/Kai-Kleinbard> | <https://github.com/kai4avaya>

FULL STACK SOFTWARE ENGINEER / PRODUCT ENGINEER | AI Engineer

- Innovative and results-oriented Software Engineer and data scientist with a proven track record
- Strong background in product / UI / UX design, building customer facing products
- Adept at designing, developing, and training AI models from scratch, demonstrating a strong commitment to cutting-edge technology and continuous learning

KEY SKILLS

- Full Stack Development
- Frontend Specialist
- Deep Learning Model Design
- LLM prompt engineering
- AWS / Azure Cloud Infrastructure
- LangChain, LlamaIndex
- Cloud optimization
- Multi-Agent Systems, MCP
- Serverless Microservices
- Data pipelines
- Tech lead / project management

EDUCATION

May, 2021

Masters of Computer and Information Technology
University of Pennsylvania

Dec, 2023

Masters of Analytics
Georgia Tech

Dec, 2024

Masters of Digital Media / Design
Harvard Extension School

LANGUAGES

Python, JS/TS, Go, Rust

AWARDS

[Excellence in Teaching / UPenn](#)

RESEARCH

[Generative Learning](#), 2025

[TensorFlow Sys Engineering](#), 2024

[ADHD and EEG Data](#), 2024

[Can AI Predict Emotions](#), 2024

PROFESSIONAL EXPERIENCE

AI / ML Principal Engineer | Tech Lead | Mission Data Grid

Feb, 2026- Present | Talon Lab (Tyto Athene)

- Network AI Agentic systems for mission critical situations (search and rescue, intelligence, medical).

Harvard Edge Computer Lab Researcher | Full Stack | Data Engineering

Jan, 2024 - Present | Harvard

- Parallel multi-agent assessment system for 100k dataset, [Quarch](#)
- Supported by [Google](#) and [Qualcomm](#) to build edtech AI on edge devices

Avaya Software Engineer | Generative AI | Full Stack | Frontend

Jan, 2024 - 2026 | Avaya

- Design and build Generative AI chatbots frontend orchestration. [Demo](#)
- Digital Human Avatars, 3D and generative AI multi-agent orchestrations. Training and fine-tuning.
- Supporting 400+ Sales enablement/partners - AI enabled tools for building customized demos for customers, ex. [Website generation app](#)
- Data orchestration and visualization. [Demo](#)

MLSysBook and TinyTorch | AI and ML Systems | Generative AI / Full Stack

Jan, 2023 - 2025 | Harvard

- Generative Learning research, [MLSysBook](#) textbook 11000+ GitHub stars
- Building generative AI agents into the browser to support learning
- Open source - [TinyTorch](#) building PyTorch from scratch, 1000s users

SocratiQ | App Development AI Specialist | Full Stack

March, 2023 - 2025 | Harvard

- Co-founder Generative AI startup - deploy custom AI to any website
- Collaborated with Google Gemini team on Generative AI Learning, adaptive textbook: Harvard, TinyML integration. [Demo](#)
- Partnered with Qualcomm on a global learning initiative

Freelance Software Engineer | AI Specialist

Jan, 2019 - 2024 | Self

- Fine-tuned GPT model to improve LLM prompts for Stability AI increased accuracy by 80%
- Built end to end products using LangChain, LlamaIndex and NextJS
- Built custom LLM evals from scratch - increasing eval scores 20%
- Custom AWS inference for HuggingFace QA model - achieved a 5x improvement in processing speed and accuracy
- Big Lift - AI system - cut latency in half, increased accuracy 2X

Lead Software Engineer | AI

Jan, 2019 - Dec 2022 | StudyDuck LLC

- Built AI model on Azure to process and clean PDF text and summarize it within seconds
- Trained deep learning model for 3X increase in speed and accuracy

Graduate Algorithms and Discrete Mathematics Teaching Assistant

May, 2020 - Dec 2021 | UPenn

- Course design and teaching for over 2000 graduate students
- Graduate level algorithms and discrete math to beginner students