

Benjamin Kozel

Atlanta, Georgia — (252) 327-6521 — benjaminlkozel@gmail.com
github.com/Zidgel — linkedin.com/in/ben-kozel

Education

Georgia Institute of Technology

Master of Science in Computer Science – Specialization: Machine Learning
Atlanta, GA

August 2024 – Present

GPA: 4.00

Relevant Coursework: Machine Learning, Deep Learning, Reinforcement Learning, High-Performance Computing, GPU Software/Hardware

North Carolina State University

Bachelor of Science in Computer Engineering, Electrical Engineering
Raleigh, NC

August 2018 – May 2022

GPA: 3.85

Skills

Languages: Python, C, C++, Bash, TypeScript, JavaScript
Frameworks & Tools: PyTorch, Hugging Face, LlamaCPP, Svelte, FastAPI, FAISS, Redis, Docker, Git, CMake, Make
ML Techniques: RAG, Unsupervised Learning, Transformers, Reinforcement Learning
Infra & DevOps: Linux, CI/CD pipelines (Docker + systemd), SSH, YAML, Redis
Embedded Systems: BIOS/UEFI, memory training, SPI, I2C/I3C, UART

Experience

Montage Technology, Inc.

Software Engineer II

April 2023 – Present

Atlanta, GA

- Initiated and designed a RAG pipeline with 3B LLaMA Instruct, LlamaCPP, PyTorch, FAISS, and Redis, streamlining DRAM spec search.
- Built clustering-based anomaly detection for DDR telemetry, catching early signal errors and accelerating validation.
- Deployed Svelte/FastAPI platform for DDR margining and clustering results; adopted company-wide for memory debug.
- Engineered on-prem AI infra: GPU servers, CI/CD pipelines, and containerized model deployments.
- Optimized and debugged DDR training in BIOS/UEFI; supported DDR signals margining and 2D eye diagram tools.

Micron Technology, Inc.

Semiconductor Product Engineer

June 2022 – February 2023

Boise, ID

- Built validation routines for DRAM training, margining, and signal integrity on pre/post-silicon hardware.
- Automated test reports and memory data analysis with Python tools.
- Root-caused DDR training failures, supporting cross-functional debug.

Certificates

Certificates: Hugging Face Transformers Course Certificate, AWS Certified Cloud Practitioner, Edge Impulse Certified Edge AI Developer

Projects

Kozami

2025 – Present

Tools: AWS, Reinforcement Learning, RTOS, C/C++, Networking, TinyML, STM32, Raspberry Pi

- Built AI-powered media streamer on Raspberry Pi 5 linked to NAS.
- Trained RL agent for dynamic TV bumpers synced to music.
- Deployed on AWS; achieved sustained multi-household engagement.
- Integrated STM32 + TinyML voice control for hands-free use.
- Tuned RTOS scheduling for smooth real-time playback on constrained hardware.