Joshua Segal

segal.jo@northeastern.edu | linkedin.com/in/joshua-francis-segal | github.com/josh-segal | joshuasegal.dev

Education

Northeastern University Boston, MA

Bachelor of Science in Computer Science, AI concentration, Math minor ***Expected: May 2026***

Honors: GPA: 3.7 | Dean's List (all semesters)
Relevant Coursework: Algorithms & Data Structures | Object Oriented Programming | Machine Learning | Computer Systems | Database Design | Artificial Intelligence | Natural Language Processing

Technical Skills

Languages: Python (5yrs) | TypeScript/JavaScript (4yrs) | SQL (3yrs) | Java (2yrs) | Rust (1yrs) | C (1yrs)
Frameworks: FastAPI | Node.js | Express | React | SQLAlchemy | PyTorch | TensorFlow
Technologies: PostgreSQL/MySQL | AWS (ECS, RDS, Lambda, S3) | Docker | Terraform | CI/CD | Git

Experience

Co-Founder & CTO **Dec. 2024 - Present**

HouseFly - AI tools for the Real Estate Industry *Boston, MA*

- Launched 2 AI products, generating $10,000+ revenue from 5 real estate clients and securing $3,000 in funding
- Built property intelligence platform with AI agents that perform contextualized property and market analysis, eliminating weeks of manual research across 180,000+ Boston properties aggregated from 4 sources
- Built multi-agent AI system automating real estate lead qualification and reducing time to respond from 2 hours to 4 seconds through multi-step conversations using Python and LLMs, processing 1000+ messages/month

Software Engineer **July 2025 - Present**

NExT Consulting *Boston, MA*

- Led team of 6 engineers building production warehouse management system for hardtech manufacturer using Python/React, eliminating 50 hours/week manual tracking, improving accuracy from 50% to 95% for 10k+ items
- Designed PostgreSQL schema with temporal row versioning and genealogy tracking for complete traceability across 100k+ rows, and optimizing queries to achieve 2x system performance improvement
- Architected and deployed AWS infrastructure (ECS, RDS, S3) using Terraform with CI/CD pipeline, supporting 250+ users across 2+ warehouse locations with 99.9% uptime

Machine Learning Engineer **July 2024 - Dec. 2024**

Harvard Research Lab *Boston, MA*

- Built neuron segmentation desktop app reducing analysis time from 20 min to 0.5s/image for 6 neuroscientists across 5000+ images, enabling researchers to label data, fine-tune YOLOv8 models, and run inference
- Developed custom UNet denoising model (11M parameters) for preprocessing using PyTorch and integrated multi-annotator agreement weighting into YOLOv8 training, improving segmentation precision from 61% to 93%

Projects

SMS Assistant for Restaurant | *Enterprise Client* | *Agentic AI Project* | GitHub

- Built production RAG SMS agent to respond to employee texts, automating 75% of responses for 50+ staff
- Implemented self-learning system using Twilio webhooks and Node.js to capture manager corrections and continuously expand Supabase knowledge base, improving answer coverage from 55% to 75% in one week
- Deployed serverless architecture on AWS Lambda, reducing costs by 90% while handling 40+ queries/day

Calendar Automation System | *AI Project* | GitHub

- Developed calendar automation system using LLMs and Notion API, saving 5+ hours/week of planning
- Fine-tuned distilBERT classification model for calendar event types, reducing event misclassifications by 80% across 200+ events, using LLMs to generate 10,000+ labeled training examples to overcome data scarcity

Medical Claims Engine | GitHub

- Built parallel claims processing engine using Rust, achieving 100% reliability across 5,000+ claims/second
- Implemented real-time claims aging report with automated bucketing and statistical analysis