

Daniel Lin

Senior AI / Full-Stack Engineer

✉ grudters@gmail.com

📍 Chicago, IL

Profile

Strategic and hands-on Senior AI Engineer with 9+ years of experience building production-grade AI and full-stack systems. Proven track record in designing and deploying LLM-based architectures — including RAG pipelines, AI agents, and generative workflows — that reduce manual effort, enhance retrieval quality, and deliver low-latency performance at scale. Equally fluent in backend infrastructure, full-stack development, and modern AI tooling, with a passion for solving hard problems and shipping value end-to-end. Adept at thriving in both startup chaos and structured enterprise environments.

Skills

Languages

Python, JavaScript/TypeScript

Frontend

React, Next.js, Tailwind CSS, Responsive Web Apps

Databases

Pinecone, ChromaDB, PostgreSQL, MongoDB

DevOps & Cloud

AWS, GCP, Docker, GitHub Actions, Firebase/Supabase, CI/CD

Backend & APIs

FastAPI, Django, Express.js, REST, Microservices

AI/LLM Expertise

LangChain, LangGraph, RAG Systems, AI Agents, LLM Integration (OpenAI, Claude, Mistral), Prompt Engineering, Multi-step Reasoning, Vector Search, Text Classification, Summarization

Frameworks & Libraries

PyTorch, TensorFlow, Scikit-learn, Pandas, NumPy, Hugging Face, AutoGen

Professional Experience

VengoAI, Senior AI / Full-Stack Engineer

09/2023 – 08/2026

Chicago, IL

- *Tech: Python, FastAPI, LangChain, LangGraph, OpenAI, Claude, Pinecone, AWS, CI/CD*
- Developed LLM-powered automation that replaced manual review and decision steps, reducing operational workload by ~35%.
- Architected high-precision retrieval pipelines (embedding, filtering, reranking), significantly boosting answer accuracy in live deployments.
- Designed multi-step AI agents that coordinate tools, retrieval, and synthesis for complex user queries with zero human input.
- Refactored async backend systems, implemented caching, and optimized CI/CD pipelines — leading to faster, more reliable production releases.
- Operated as full lifecycle owner from experimentation to production rollout.

Sphere, Full-Stack AI Engineer

05/2022 – 07/2023

New York, NY

- *Tech: Python, FastAPI, LlamaIndex, RAG Systems, Vector Search, PostgreSQL, GCP*
- Delivered AI services for enterprise/government use, accelerating analysis of large regulatory datasets.
- Built ingestion + indexing pipelines with LlamaIndex + vector DBs to eliminate stale results and improve semantic recall.
- Developed scalable APIs with modular architecture to support evolving AI needs with minimal tech debt.
- Diagnosed & resolved real-world latency and stability issues across production workloads.

SimuHealth, AI Engineer

01/2021 – 04/2022

Chicago, IL

Tech: Python, React, Next.js, Tailwind CSS, GPT, MongoDB, REST APIs

- Developed and deployed AI-powered features for a healthcare platform, integrating GPT-based summarization and text classification into clinical and operational tools.
- Connected LLMs with backend services and intuitive frontend interfaces, enhancing data interpretation and decision support for healthcare professionals.
- Streamlined backend workflows and data preparation, accelerating iteration on patient-facing features.
- Collaborated with product and design teams to ensure AI outputs were transparent, user-friendly, and aligned with healthcare compliance standards.

Disqus, Software Engineer (Backend / Data)

06/2018 – 12/2020

San Francisco, CA

Tech: Python, TypeScript, Django, Express.js, SQL, Pandas, NumPy, REST APIs

- Delivered backend infrastructure and data workflows supporting analytics and automation.
- Built REST APIs and data integrations for large-scale comment and engagement systems.
- Leveraged Pandas/NumPy for transforming structured data into actionable insights.
- Gained strong production-level experience in debugging, scaling, and backend ownership.

Education**Vanderbilt university, Bachelor of Science in Computer Science (GPA: 3.8/4.0)**

05/2014 – 05/2018

Nashville, TN

Certificates

- LangChain for LLM Application Development – DeepLearning.AI
- Building Systems with the ChatGPT API – DeepLearning.AI