

# DENIS SIMINIUC

Cambridge, MA · simdenis@mit.edu · +1 (857) 706-9124 · [github.com/simdenis](https://github.com/simdenis) · [linkedin.com/in/simdenis](https://linkedin.com/in/simdenis)

## EDUCATION

### Massachusetts Institute of Technology

Cambridge, MA

Bachelor of Science in Computer Science and Engineering

Bachelor of Science in Mathematics | Minor: Linguistics

Sep 2022 – May 2026

**Relevant Coursework:** Machine Learning, Design and Analysis of Algorithms, Software Construction, Computer Systems, Probability, Statistics, NLP, Cryptography, Computation Theory

## WORK EXPERIENCE

### Research Assistant | MIT CSAIL

Sep 2025 – Present

- Building an end-to-end pipeline over job advertisements to identify science research roles, classify postings into OpenAlex subfields / OECD research levels, and extract task verb phrases via LLM prompting.
- Cleaning and deduplicating extracted tasks using sentence embeddings and FAISS nearest-neighbor search with similarity thresholds/guardrails to merge near-duplicates and filter composite tasks.
- Clustering tasks with UMAP/HDBSCAN and inducing a task hierarchy by scoring subtask relations with MNLI (RoBERTa-large-mnli); validating extraction quality on labeled samples.

### Software Developer Intern (Risk Systems) | ExodusPoint Capital Management

May 2025 – Aug 2025

- Built a configurable monitoring + alerting web app (Flask + JS/HTML) for business-critical financial pipelines, allowing Risk users to define datasets/files to watch, SQL checks to run, recipients, and expected delivery windows.
- Integrated checks across internal MySQL and Snowflake sources; replaced hardcoded 5-minute polling with schedule-aware execution and stateful alerting via the internal email system, reducing database calls.
- Containerized the service with Docker and validated behavior on a Linux VM to match production constraints (dependencies, networking, runtime configuration).

### Research Assistant | FutureTech Initiative, MIT CSAIL

Feb 2024 – Feb 2025

- Curated and standardized metadata for AI foundation models (e.g., parameter counts, provenance/origin, release/access details) to support analysis of model usage and adoption in scientific literature.
- Read primary sources (papers/model cards/docs), resolved naming/versioning ambiguity across model variants, and QA'd entries so models could be reliably joined to papers and citation/usage signals.

### Research Assistant | MIT Lab for Nuclear Science & CERN

Jan 2023 – Dec 2023

- Applied machine learning methods to improve Higgs boson detection workflows for a Future Circular Collider feasibility study.
- Processed large particle-physics datasets in Python and C++ and contributed to analysis scripts and experiment reporting for research conclusions.

## PROJECTS

### MIT Dining Alert Web Service | Python, Flask, PostgreSQL, HTML/CSS

Nov 2025 – Present

- Built a full-stack web app that scrapes MIT dining menus and emails subscribers when chosen items appear, supporting keyword and dining hall filters.
- Implemented a PostgreSQL-backed subscription model, tokenized unsubscribe flows, and an idempotent scheduled notifier (GitHub Actions cron) for reliable recurring alerts.

### Semantic Embedding Optimization for NYT *Connections* | Python, PyTorch, SentenceTransformers

Sep 2025 –

Dec 2025

- Built an end-to-end pipeline for learning task-specific text embeddings for the NYT *Connections* word-grouping game (data prep, training, evaluation, and error analysis).
- Fine-tuned a MiniLM-based embedding model and added lightweight post-processing; improved grouping performance on held-out puzzles versus a frozen embedding baseline.

## SKILLS

**Languages:** Python, C/C++, JavaScript/TypeScript, Java, Bash

**Web/DB:** Flask, REST APIs, SQLAlchemy, PostgreSQL, SQL Server, MySQL, Snowflake, Node, Express, React

**ML/Data:** PyTorch, SentenceTransformers, FAISS, UMAP, HDBSCAN

**Tools:** Git, Linux, Docker/Containers, CI/CD

**Interests:** Linguistics, teaching, literature, cinema, running