Dominick DeVincenzo

(717) 343-2932 | hello@dominick.dev | [Portfolio Website](#) | [GitHub](#) | [LinkedIn](#)

Professional Summary

Aspiring Software Engineer with over 3 years of experience as a Clinical Research Data Coordinator, demonstrating strong skills in data management, communication, and analytical problem-solving. Currently balancing a full-time role while pursuing a Master of Computer and Information Technology at the University of Pennsylvania. Actively engaged in meaningful side projects in full-stack software development and data analysis, showcasing dedication to expanding technical expertise beyond academic coursework. Eager to leverage a unique blend of practical experience and academic knowledge to transition into a software engineering role.

Education

University of Pennsylvania   Philadelphia, PA
Master of Computer and Information Technology   *Dec 2025*
- Computer Systems Programming
- Database & Information Systems
- Networked Systems
- Data Structures & Software Design
- Algorithms & Computation
- Intro to Software Development

University of Pittsburgh   Pittsburgh, PA
B.S. in Chemistry | Minor in Italian Language and Culture

Technical Skills

Languages: Java, Python, C/C++, SQL (PostgreSQL), JavaScript, TypeScript, HTML/CSS
Frameworks: React, Next.js, Streamlit, Node.js, Express, JUnit
Developer Tools: Linux/Unix, Git, Docker, MongoDB, GCP, AWS, VS Code, Postman, Makefile, DataGrip
Libraries: LangChain, Mongoose, JWT, Axios, pandas, NumPy

Projects

Clinical Trial AI Chatbot | *Python, LangChain, Streamlit, ChromaDB, Docker, Llama 3*   2025
- Built a Retrieval-Augmented Generation (RAG) chatbot to screen patients for clinical trial eligibility by extracting inclusion/exclusion criteria from research protocol PDFs
- Implemented section-based chunking, embedding with all-MiniLM-L6-v2, and metadata tagging to enable accurate retrieval
- Used ChromaDB for vector search and LangChain to orchestrate LLM interaction with local inference via Llama 3
- Developed a lightweight Streamlit frontend with a conversational UI for guided patient qualification without storing responses
- Optimized PDF ingestion and embedding pipeline for GPU-accelerated local development using Docker across Mac and Windows
- Enabled study updates by allowing document replacement and metadata refresh, supporting evolving clinical trial requirements

Distributed Search Engine over Chord DHT | *C++, ns-3, Distributed Systems*   2025
- Implemented a distributed keyword-based search engine over a Chord Distributed Hash Table (DHT) with $\mathcal{O}(\log N)$ routing and lookup efficiency
- Designed support for dynamic node joins and departures with consistent data redistribution and finger table updates for fault-tolerant operation
- Built and tested the system using the ns-3 network simulator, modeling realistic network delays, failures, and message passing
- Implemented ring stabilization protocols and periodic finger table maintenance to ensure consistent routing state across a dynamic peer network
- Evaluated system performance through logging and measurement of hop counts, query latency, and recovery time under churn

BookBase | *JavaScript, React, Node.js, Express.js, MongoDB, PostgreSQL, AWS RDS*   2024
- Developed a full-stack web application for book discovery, providing advanced search and unique insights powered by optimized SQL queries and map-based exploration
- Implemented user authentication with JWT and Google/FB OAuth, enabling secure login and personalized virtual bookshelves
- Integrated large datasets from Amazon and Goodreads to provide detailed book metadata, user reviews, and geographic insights
- Designed optimized SQL queries to uncover insights such as hidden gems, polarizing books, and age-based popular books
- Developed an interactive map using React-Leaflet, allowing users to explore reviews geographically
- Preprocessed and normalized over 7 million reviews and 6 million books by standardizing ISBNs, geocoding locations, and resolving schema inconsistencies for seamless PostgreSQL integration
- Reduced query execution times by 80% through restructuring, precomputing, and indexing for high-efficiency database operations

Experience

Data Coordinator | National Capital Area Minority/Underserved NCORP   Feb 2022 – Present
Georgetown University Medical Center   *Washington D.C.*
- Designed and built a proof-of-concept Retrieval-Augmented Generation (RAG) AI chatbot to explore automating patient screening using local LLMs and research documentation; demoed the tool to leadership as a potential innovation for clinical workflows
- Effectively coordinate patients and manage data for six oncology studies totaling over 100 subjects; ensuring timely and accurate data collection, entry, and verification in accordance with protocol requirement and regulatory standards
- Developed data quality assurance protocols, playing a key role in maintaining the credibility and reliability of research outcomes
- Optimize data and patient management processes by implementing best practices and emerging technologies, enhancing efficiency and data integrity