

Vedant Somani

vedants0@uw.edu | +1 (740) 760 0011 | Seattle, WA | [LinkedIn: vedant-somani17](#) | [GitHub: vedant1711](#)

WORK EXPERIENCE

Text Mining Research Analyst, UW Libraries & eScience Institute- (NLP, BERT, Haystack, RAG) **Mar 2026 – Present**

- Engineered scalable **NLP** ingestion pipelines using **spaCy** and **pandas** to clean, normalize, and deduplicate **100K+ scholarly documents**, reducing downstream preprocessing time **by 40%**.
- Applied **BERT** models to extract structured entities and semantic relationships from unstructured research corpora, enabling **semantic search** across **50K+ records**.
- Integrated **Haystack-based retrieval-augmented generation (RAG)** to surface contextually relevant passages from large document archives, cutting manual research query time by **~30%** for library staff.

Machine Learning Intern, Xenon Nation – (PyTorch, FastAPI, Docker) **Mar 2024 – Sept 2024**

- Improved sentiment analysis F1-score by **18%** through fine-tuning **BERT** models with **Hugging Face Transformers** using **PyTorch** on **AWS EC2**.
- Reduced processing latency by **35% (-250 ms)** by designing a transformer-based summarization pipeline and deploying scalable **FastAPI** microservices with **Docker** on **AWS ECS**.

Machine Learning Engineer Intern, [HemansAI](#) – (NumPy, Pandas, Scikit-learn, REST APIs) **Sept 2023 – Feb 2024**

- Enhanced model accuracy by **15%** through feature engineering, cross-validation, and hyperparameter tuning for predictive analytics workflows.
- Improved backend inference performance by **20%** via optimized Django REST APIs, PostgreSQL indexing, and Redis caching for high-throughput ML serving.

PROJECTS

ResolveIQ – AI-Powered Incident Knowledge Management Platform, (LangChain, FastAPI, Pinecone)

- Architected an event-driven microservices platform to auto-generate structured **Confluence KB articles** from resolved **Jira tickets** by aggregating **Slack & Zoom context** eliminating 100% of manual documentation overhead.
- Built a semantic retrieval engine (**Voyage AI embeddings + Pinecone vector search + GPT-4o**) to surface relevant KB articles inline in Jira **cutting MTTR** for recurring issues by an estimated **30%**.

Vegapunk – Autonomous AI Code Agent, (LangGraph, FastAPI, NVIDIA NIM, Next.js, Docker)

- Built a fully autonomous GitHub issue-to-PR pipeline using a 7-node **LangGraph** multi-agent system backed by **NVIDIA NIM LLMs**, resolving GitHub issues end-to-end with self-correcting retry loops on test failure.
- Engineered a real-time **SSE event bus** streaming 19+ live agent events to a **Next.js** dashboard with **Docker-sandboxed code** execution and **baseline-aware pytest comparison**, ensuring pre-existing test failures never block new fixes.

FarmFit – Smart Precision Farming Solution, (TensorFlow, Django, React Native, AWS, IoT)

- Optimized irrigation efficiency by **26%** and boosted crop yield by **11%** by developing an **IoT-ML** platform that dynamically adjusted watering and fertilization using **real-time sensor data**.
- Engineered predictive model for **fertilizer recommendation** and deployed via **Django + AWS**, integrating with **React Native dashboard** for real-time visualization and remote control.

SKILLS

Languages and Databases: C++, Python, Java, HTML, CSS, JavaScript, TypeScript, MySQL, PostgreSQL, MongoDB, Pinecone

Machine Learning, Frameworks: NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, Snowflake, TensorFlow, RNN, LSTM, OpenCV, LLM, PyTorch, Hadoop, Scala, Apache Spark, Apache Kafka, MATLAB, FastAPI, Django, React, Next.js, Node.js

AI & Automation: Claude API, OpenAI GPT-4o, Langchain, LangGraph, LangSmith, LlamaIndex, RAG pipelines, BERT, IBM Watson API, Haystack, AutoGen, CrewAI, AI Tool calling, AI Skills, Multi Agent systems, MCP, LLM prompt engineering

Platforms: AWS, Jira, Tableau, Figma, Git & GitHub, Postman, Azure, Docker.

ACHIEVEMENTS & PUBLICATIONS

- Dempsey Startup Competition 2026 **Investment Round Finalist**
- iEngage – Salesforce – Lovable **Hackathon Runner-up**
- Association of Information Management Students at UW (UW AIMS) **Product Competition Runner-up**
- IEEE ACROSET 2025: PyFao-Driven Precision Irrigation and Fertilizer Optimization in Agriculture**
- IEEE ICTBIG 2024: Elderly Well-Being Through Volunteer Engagement**
- Fundamentals of Deep Learning (NVIDIA) Certification**
- AWS Academy Graduate – AWS Academy Cloud Architecting**

EDUCATION

University of Washington, Seattle
Master of Science – Information Management

June 2027
GPA: 4.0

Acropolis Institute of Technology and Research, India
B.Tech. – Computer Science Engineering

June 2025