

Sahil Chordia

College Park, MD | (240) 610-1453 | chordiasahil24@gmail.com | [linkedin/in/sahil-chordia](https://www.linkedin.com/in/sahil-chordia) | [github/protocorn](https://github.com/protocorn)

PROFILE

Data Science M.S. candidate with experience building internal platforms to improve developer workflows, ensure data quality, and strengthen system reliability. Proficient in CI/CD automation, data validation, and system observability.

SKILLS

Python, JavaScript, TypeScript | PyTorch, TensorFlow, LangChain | React.js, Node.js, Express.js | MongoDB, PostgreSQL, Pinecone, FAISS, Qdrant | AWS, Docker, Git | *Telemetry* (Otel) | CI/CD | Ollama, Gemini API, CLIP

EDUCATION

University of Maryland | CGPA: 3.62

Master of Science in Data Science

College Park, MD

August 2024 - May 2026

A.P. Shah Institute of Technology | CGPA: 3.52

Bachelor of Engineering in Computer Science

Thane, MH

June 2020 - May 2024

WORK EXPERIENCE

Raising The Floor – *Nonprofit focused on accessibility technology*

AI Intern

Washington D.C (Remote)

June 2025 - August 2025

- Developed an automated pipeline that reduced manual verification time by 50%, improving developer productivity and data quality.
- Resolved **95%** of detected dataset inconsistencies (misabeled categories, duplicates, outdated entries) by automating enrichment workflows for a database of 300+ tools.
- Built an inclusive chatbot (Ollama Gemma2:9B), improving ambiguous query handling accuracy to **90%**
- Boosted search by integrating cosine similarity-based vector search (Qdrant) for AT tool recommendations.
- Significantly reduced QA cycles by developing reusable audit scripts, enhancing system reliability and *data validation*.

PROJECTS

Mimikree - Personalized AI Assistant | mimikree.com

February 2025 - Present

- Developed a platform that ingests diverse content to generate personalized LLMs, focusing on data quality and actionable insights.
- Engineered a hybrid retrieval pipeline (70% expanded query + 30% original), improving RAG accuracy.
- Integrated Gemini 2.0 Flash with Pinecone for dynamic vector search and real-time responses.

AI Document Generator – Capital Area Food Bank AI Challenge

February 2025 - May 2025

- Led a team of 3 to design a RAG engine generating grant proposals, PPTs, & posts from **1,500+ documents**.
- Embedded text and images (CLIP model + Cloudinary) into Pinecone with metadata for multimodal retrieval.
- Decreased revision time by **60%** through AI-assisted selective edits, full-document re-generation from a single prompt, and content suggestions with error detection.

Local Document Search Engine

March 2025 - May 2025

- Built privacy-first vector search (FAISS + Sentence Transformers) with intelligent code-aware chunking.
- Reduced re-indexing time by **~80%** for large repos through caching and parallel indexing.
- Integrated Ollama Llama3:8B for query expansion, improving semantic relevance by **~20%**.

PUBLICATIONS

Chordia, S. et al. Cross-Domain Sentiment Analysis: An Extensive Study of Machine Learning and Deep Learning Models, Datasets, and Preprocessing Techniques for Predictive Performance. CSCT 2023, [Link](#).

ACHIEVEMENTS

- **Finalist (Top 3%)** – i.mobilothon 3.0, Volkswagen Group Technology Solution India (Nov 2023)
- **3rd Place** – Firefly Hackathon, NIT Warangal (Apr 2022)