

Sahil Chordia

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

EDUCATION

University of Maryland | CGPA: 3.77/4.0

Master of Science in Data Science

Relevant Coursework: Machine Learning, Probability and Statistics, NLP, Computer Vision, Data Modelling, Robotics

College Park, MD

Aug 2024 - May 2026

A.P Shah Institute of Technology

Bachelor of Engineering in Computer Engineering

Thane, India

Jun 2020 - May 2024

SKILLS

Languages: Python, SQL

Machine Learning & AI: PyTorch, Scikit-learn, Hugging Face, CLIP, Fine-Tuning, QLoRA, DPO

LLM & Agent Systems: RAG, LangChain, LangGraph, Ollama, Agentic Workflows, Prompt Engineering

Data & Vector Systems: Pandas, NumPy, PostgreSQL, Redis, Pinecone, Qdrant, FAISS

Backend & Infrastructure: FastAPI, Docker, AWS

WORK EXPERIENCE

RAISING THE FLOOR (non-profit organization focussed on assistive technology)

Washington D.C (Remote)

AI Intern

Jun 2025 – May 2026

- Reduced audit time by **96% (7 days → under 2 hours)** by building an LLM pipeline that cross-validates assistive technology tools metadata against source docs and auto-flags inconsistencies across **360+ records**.
- Eliminated manual tool discovery by building a scrape-deduplicate-validate pipeline using LLM self-critique, enabling continuous ingestion of clean assistive technology entries into the production database.
- Improved assistive technology tool recommendation accuracy by developing a **Gemma2 (9B)** chatbot with Qdrant vector search that resolves ambiguous queries through structured clarification.

UNIVERSITY OF MARYLAND

College Park, MD

Teaching Assistant for Big Data Systems

Jan 2026 – May 2026

- Designed and graded assignments for 60 graduate students in distributed systems, Spark, and large-scale data processing.

PROJECTS

Clippy Vision – Local-First AI Memory Assistant | [GitHub](#)

May 2026 - Present

- Developed a continuous-memory system that captures user activity, generates session summaries, distills long-term semantic memories, and enables context-aware retrieval using local LLMs, embeddings, and vector search.
- Designed a 3-stage classification pipeline combining rule-based filtering, statistical heuristics, and multimodal LLMs with vision fallback, reducing inference calls by **37%** across 1,000+ real events.
- Engineered screenshot deduplication using perceptual hashing and Union-Find clustering, lowering vision-processing workload by **23%** across 900+ screenshots while preserving classification accuracy.

AI-Assisted Job Application Automation Platform | [Demo](#)

Sep 2025 - Present

- Cut average job application time from ~14 minutes to under **2 minutes** by building an autonomous platform combining AI resume tailoring, relevance-based job ranking, and Playwright-driven ATS form-filling.
- Achieved **86% field-population success** across 650+ applications for 12 beta users by architecting a 3-layer form-filling fallback (deterministic mapping, learned patterns, AI inference) with telemetry across 41 run-level metrics.

Domain-Aligned Speculative Decoding

Mar 2026

- Fine-tuned Phi-3 Mini (3.8B) via QLoRA + DPO on 985 synthetic preference pairs and built a speculative decoding engine from scratch using Phi-3 Medium as the verifier to evaluate alignment's effect on throughput.
- Discovered that DPO alignment does not generalize uniformly across query types: performance improved for tool-suggestion queries (**+6.7%**) but degraded for instructional queries (**-5.2%**), suggesting alignment is context-sensitive.

PUBLICATION

Cross-Domain Sentiment Analysis | CSCT 2023 (Springer) | [Link](#)

Oct 2024

- Trained models on a generic Twitter dataset to predict sentiment on Amazon and IMDb reviews, with **BiLSTM** achieving the best cross-domain accuracy (**82.56%, 80.43%, 79.48%** across 3 datasets).

ACHIEVEMENTS

- Finalist (**Top 3%**) – i.mobilathon 3.0, Volkswagen Group Technology Solution India (Nov 2024)
- 3rd Place** – Firefly Hackathon, NIT Warangal (Apr 2023)