

Pratham Chopra

+1 (781) 805-0647 | chopra.pr@northeastern.edu | [linkedin.com/in/pratham-chopra](https://www.linkedin.com/in/pratham-chopra) | github.com/prathamchopra001 | prathamchopra.page

EDUCATION

Northeastern University | Master of Applied Machine Intelligence, Boston, MA | April 2026 | GPA: 3.94/4.0

Jain (Deemed-to-be) University | B.Tech in Computer Science and Engineering (AI & ML), Bengaluru, IN | July 2024 | 8.76/10

EXPERIENCE

Futurense Technologies | Data Science Intern, Bengaluru, IN Jan 2024 - May 2024

- Built automated ETL pipelines unifying datasets from 8+ government sources, resolving schema inconsistencies across healthcare and housing records to produce a validated data warehouse for regional policy analysis.
- Developed Power BI dashboards backed by Python (Pandas/NumPy) processing to surface 5 KPIs across complex market data, enabling stakeholder decisions without manual reporting cycles.
- Designed a 12-table relational schema using SQL normalization, optimizing multi-dimensional query performance for insurance claims data across 3 downstream analytical modules.

TECHNICAL SKILLS

Languages & Frameworks: Python, SQL, FastAPI

Machine Learning & Deep Learning: PyTorch, TensorFlow, Keras, Scikit-learn, XGBoost, CNNs, LSTMs, MONAI, timm, Transformers, Ensemble Methods, Transfer Learning, Mixed-Precision Training, Stratified Cross-Validation, Test-Time Augmentation, Random Forest

Generative & Agentic AI: LangChain, ChromaDB, FAISS, Sentence-Transformers, RAG, BM25 Hybrid Retrieval, ReAct Agents, Tool-Calling, LLM-as-a-Judge, Zero-Shot and Few-Shot Prompting, Chain-of-Thought, QLoRA, PEFT

MLOps & Infrastructure: Docker, GitHub Actions, MLflow, Weights & Biases, Airflow, ONNX, TensorRT, PSI Drift Monitoring, Evidently, Grafana

Statistics & Evaluation: Cohen's Kappa, ROUGE, BLEU, Perplexity, Z-Test, VADER Sentiment Analysis, A/B Testing

Data & Visualization: Pandas, NumPy, Spark, NetworkX, Git, Matplotlib, Power BI

PROJECT EXPERIENCE

INQUIRO | Autonomous Scientific Research Agent | *Custom Python, ChromaDB, Sentence-Transformers, Docker, Semantic Scholar API, ArXiv API, PubMed API*

- Built a 3-agent orchestration framework using Adaptive Hierarchical Decomposition to decompose research objectives into DAGs, with a SQLite/NetworkX World Model enforcing deterministic plan-execute-evaluate state transitions.
- Developed a RAG pipeline using GROBID for TEI XML parsing and section-based PDF chunking, with hybrid BM25 + dense retrieval via ChromaDB, achieving a 44.8x reduction in token consumption over naive context injection.
- Implemented ScholarEval, an LLM-as-a-judge framework scoring findings across 8 quality dimensions, and added a VLM grounding layer (FigureVerifier) to cross-validate text claims against figures, improving discovery reliability by 30% across 154 tests.

Helix | End-to-End MLOps Platform | *Python, MLflow, ONNX, FastAPI, Airflow, Evidently, Grafana, Docker*

- Orchestrated a 4-stage MLOps pipeline (ingest, ETL, train, monitor) in Airflow with XCom-based conditional logic, enabling a fully automated retraining loop triggered by PSI drift thresholds across 11 equity and crypto tickers.
- Built a FastAPI inference engine with 3-tier fallback (TensorRT to ONNX to PyTorch), achieving sub-10ms p99 latency via TensorRT FP16 and cutting runtime memory by 40% through ONNX serialization.
- Deployed an Evidently + Grafana observability stack monitoring 29 features over 60-day sliding windows, with a custom PSI service that flagged drift before LSTM anomaly detection accuracy degraded.

RSNA Intracranial Aneurysm Detection | Kaggle Top 300 / 1,800+ Teams | *PyTorch, timm (EfficientNet), MONAI, Weights & Biases, Mixed-Precision Training*

- Cut DICOM preprocessing latency from 25 minutes to under 30 seconds per series using Python multiprocessing and vectorized HU windowing, with 2.5D slice-window extraction and NPZ serialization maintaining 98%+ GPU utilization.
- Trained a 2.5D dual-stream EfficientNet with SE attention across 4,000+ CTA scans, using Class-Weighted BCE Loss for severe class imbalance, securing a Top 300 global ranking out of 1,800+ teams.
- Managed 100+ hyperparameter sweeps via Weights & Biases with AMP and Gradient Accumulation to optimize VRAM, embedding 5-fold Stratified Cross-Validation and Test-Time Augmentation throughout.

Customer Journey & Conversion Analytics | *Python, Pandas, NumPy, Plotly, Seaborn, Streamlit, Scikit-learn*

- Built an e-commerce analytics pipeline processing millions of events across a 3-stage funnel (View, Cart, Purchase), computing conversion rates and pinpointing the 2 highest-impact drop-off points driving revenue loss.
- Applied cohort segmentation and A/B test simulation via proportions z-test across 50+ product segments, surfacing statistically validated brand-level purchase intent insights.
- Deployed a Streamlit dashboard with real-time funnel visualization, trend monitoring, and segment-level breakdowns, enabling stakeholder analysis without requiring code.

CERTIFICATIONS & PUBLICATIONS

Certifications: Microsoft AI-900 | PG Program in Data Science & AI, IIIT | Generative AI with LLMs, DeepLearning.AI | ML with Python, IBM | Architect Multi-Agent Systems with Agent Development Kit (ADK), Google Cloud

Publication: AR in Fashion Industry | 4th IEEE International Conference on Advances in Computing, Communication Control and Networking, 2022