

ABHITAY SHINDE

(646) 225-0802 | abhitayshinde@gmail.com | New York, NY | [LinkedIn](#) | [Portfolio](#)

EDUCATION

Columbia University

M.S. Data Science (AI Specialization), GPA: 3.7/4.0

Sep 2024 - Dec 2025

New York, NY

- Coursework: Agentic AI, Fintech and Data Economy (PhD elective), Big Data Analytics, and Applied Machine Learning.

SELECTED PROJECTS

IBM Agentic Graph RAG for Document Q&A | *LangGraph, RAG, Neo4j, FAISS, ChromaDB, LLMs, Python, OpenAI*

- Enabled multi-step reasoning over enterprise documents by creating agentic LangGraph pipelines with CoT decomposition.
- Improved retrieval precision 30%+ over naive vector search, by modeling documents as entity-relationship graphs in Neo4j with FAISS hybrid retrieval.

Product Analytics Suite | *Python, Causal Inference, A/B Testing, Propensity Score Matching, DiD, Regression, MILP*

- Built a marketing mix model and paired it with mixed-integer optimization to reallocate spend under budget constraints, producing an incremental-LTV channel ranking that diverged from last-touch attribution.
- Attributed a 15.4% decline to mix shift, by decomposing across cohorts to show within-cohort performance was flat.

PROFESSIONAL EXPERIENCE

Ask2AI

Data Scientist (Customer Acquisition)

New York, NY

Feb 2026 - Present

- Corrected selection bias in the scoring model, by building a feature engineering and causal inference pipeline in Python and Polars that integrated 11 source systems into a unified 30 million customer dataset.
- Scored the rejected-applicant population at 0.86 held-out AUC, by training gradient-boosted models within a Double Machine Learning framework to correct the selection bias in approved-only training data.
- Expanded the approvable population by 15% at flat expected default, by handing corrected scores to credit policy.

Data Scientist Co-op (Customer Acquisition)

Jan 2025 - May 2025

- Improved cross-channel ROI attribution by 22% and identified optimal credit limit thresholds, by building uplift models and difference-in-differences designs that isolated incremental campaign impact from last-touch bias.
- Lifted customer LTV 12% and cut churn 25%, by pairing uplift modeling with mixed-integer optimization to reallocate acquisition budget across channels.

Julius Baer

Data Scientist (Generative AI and Personalization)

New York, NY

Jun 2025 - Sep 2025

- Replaced the analyst-and-copywriter handoff for client-facing investment documents, by engineering an agentic LangGraph system retrieving client holdings, stated preferences, and live market news into a single generated draft.
- Automated compliance review at 86% agreement with human reviewers, by building an LLM-as-a-judge critic that scored each draft against the firm's rubric and regenerated failures until they passed.
- Cut content selection time for relationship managers by 35%, by shipping the pipeline as a chat interface with human-in-the-loop revision, letting RMs iterate on a living client document.
- Surfaced a 25% engagement lift, using exploratory analysis on customer data in Python and SQL against matched controls.

TD Bank

Data Scientist Capstone (Customer Analytics and Segmentation)

New York, NY

Jan 2025 - May 2025

- Raised fraud model precision-recall AUC from 0.20 to 0.67, by training XGBoost with SHAP on 100 million transactions.
- Validated fraud-detection strategy for senior leadership, by quantifying incremental catch rate against a holdout.
- Shaped segmentation and anti-money-laundering strategy, by clustering customers on transaction behavior and delivering automated reporting on the resulting risk tiers.

AWARDS & PUBLICATIONS

- Winner of global AI hackathon, placing 1st out of 750 teams with international recognition for applied AI.
- Lead author on 4 peer-reviewed IEEE papers with 16 citations, spanning predictive maintenance and ML modeling.

TECHNICAL SKILLS

- **Generative AI & LLMs:** RAG, Agentic AI, LangGraph, NLP, Transformers, Vector DBs (FAISS, ChromaDB), Guardrails.
- **Machine Learning:** XGBoost, Random Forest, Regression, Clustering, Forecasting, PCA, SHAP, Scikit-learn, PyTorch.
- **Experimentation:** Causal Inference (PSM, DiD, Double ML), Uplift Modeling, MMM, A/B Testing, Metric Design.
- **Data Languages:** Python, SQL, PySpark, Pandas, NumPy, Polars, ETL, Data Modeling, Airflow, Databricks, Tableau.
- **Cloud & MLOps:** AWS, GCP, Vertex AI, BigQuery, Docker, Git, CI/CD, Model Monitoring, Data Pipelines.