

ABHISHEK JAISWAL

AI Product Manager · LLM Evaluation & Quality · Agentic AI Systems

Plano, TX (open to relocation / remote) | +1 (551) 285-8560 | abhishek.jaiswal@rutgers.edu | github.com/abhishek2395 | linkedin.com/in/iamabhishekjaiswal

SUMMARY

Technical AI product manager with 5+ years across AI evaluation, LLM systems, and data products. I build the thing I would manage: an open-source LLM evaluation framework that benchmarks 300+ models with an LLM-as-judge, and a production multi-agent AI system shipped to 98% automation inside a regulated financial enterprise. I own product roadmaps and PRDs, run user discovery and experimentation, and define the quality metrics that decide whether an AI feature is ready. My edge is depth: I write the Python, design the evals, and sit with engineers on architecture, then translate it into product decisions a business and a model can both act on. I care about shipping AI that serves a wide audience and holds up when the model is wrong.

AI PROJECTS & OPEN SOURCE

LLM Evaluation & Benchmarking Framework

Open Source · github.com/abhishek2395/llm-eval

Python · Claude (LLM-as-judge) · OpenRouter (300+ models) · Streamlit · live eval dashboard

Open-source framework that benchmarks 300+ language models with an LLM-as-judge across six quality dimensions, rolled into a composite score and a value index that weighs quality against cost, with a live dashboard for quality, latency, and per-prompt drill-down. Built the product end to end: defined the quality dimensions, designed the scoring, shipped it publicly, and iterated on community feedback.

Production Multi-Agent AI System

Production · USAA (financial services)

Claude Sonnet · Snowflake MCP · Python · agent orchestration · retrieval · evaluation

Designed and shipped a 4-agent system end to end: document extraction, planning, query execution via MCP, and automated output generation. Defined the acceptance criteria and quality metrics, ran iterative evals, and aligned engineering and business stakeholders on what good looked like. Replaced a manual process and reached 98% automation across regulated cycles.

EXPERIENCE

Senior Analyst, Product & AI Engineering

Jun 2021 – Present

Tata Consultancy Services | Client: USAA (financial services) | Plano, TX

- Owned the product roadmap and PRDs for an internal analytics platform serving Finance, Operations, and Product teams: ran user discovery, tracked adoption, and prioritized from usage data. Cut executive time-to-insight by 30%.
- Defined the evaluation and quality bar for a production multi-agent AI system: set acceptance thresholds, built evaluation datasets, and used LLM-as-judge scoring to quantify accuracy, relevance, and hallucination before it shipped to 98% automation.
- Translated business workflows into agent capabilities and clear requirements, partnering directly with engineers on architecture (agent design, MCP integration, the data layer) rather than handing off specs.
- Drove delivery across engineering, data, and business stakeholders in Agile cycles, and led a 4-person data engineering team on day-to-day execution.
- Built governance and responsible-AI controls for a regulated environment: audit logging, access controls, and human-in-the-loop review on sensitive steps.

Data Analyst, Ad Intelligence Product

May 2020 – Dec 2020

Airfind Corp | Remote, US

- Took product ownership of an early-stage product: defined KPIs, designed A/B test methodology, and shipped dashboards tracking user behavior. Revenue grew 15% from the recommendations.

Research Analyst, Automation & Data Quality

Jun 2017 – May 2019

Media.Net Software Services | Mumbai, India

- Cut processing time 40% by building Python and SQL automation pipelines for data quality and analytics workflows.

SKILLS

AI Product: AI product management · Product roadmap · PRDs · Product strategy · User discovery · Experimentation & A/B testing · Adoption & quality metrics · Cross-functional delivery · Stakeholder alignment · Agile

LLM Evaluation & Quality: LLM-as-judge evaluation · Eval pipeline design · Accuracy / hallucination / reliability scoring · Acceptance thresholds · Responsible AI · Quality benchmarking

Agentic AI & LLMs: Agent design (planning · tool calling · memory) · Multi-step orchestration · RAG & retrieval · Prompt engineering · LLM APIs (Claude / OpenAI) · Snowflake MCP · LangChain (familiar)

Technical: Python (Pandas · NumPy) · SQL · AWS · Snowflake · dbt · Docker · Git · data pipelines · production shipping

EDUCATION

MS, Information Technology & Analytics | Rutgers University – Newark | GPA: 3.83 | 2020

BE, Electronics & Telecommunications | KJ Somaiya College of Engineering | 2017

CERTIFICATIONS

AI for Product Management | Pendo · Google Cloud · Mind the Product

AWS Certified AI Practitioner (AIF-C01) | Amazon Web Services | Active