

AMIT THAKUR

AI PROGRAM MANAGEMENT & DELIVERY · RL ENVIRONMENTS · LLM EVALUATION
+91 83840 93413 | amitdexpert@gmail.com | [linkedin.com/in/amitdexpert](https://www.linkedin.com/in/amitdexpert) | Hyderabad, India

PROFESSIONAL SUMMARY

Program and delivery lead for the data, evaluation, and RL-environment systems that frontier AI models are trained and benchmarked on. At Deccan AI, set technical direction for agentic RL-Gym programs, own task-design and QC standards used across annotation and QC teams, and take new offerings from market research to client-ready pilots. Built the underlying systems hands-on (LLM-as-judge pipelines, task verifiers, synthetic enterprise environments), ran the human side, managing annotation and QC teams of 15–20 people day to day and up to 50 at peak (calibration, QC process design, resource planning and PnL), and worked directly with clients on solution design and pricing.

CORE COMPETENCIES

Program & Delivery: Technical direction and standards, roadmap and requirements, annotation strategy, QC and SQC process design, resource planning and PnL, stakeholder and client management, solution design and pricing, annotator calibration and mentoring, Jira

AI & Evaluation: LLM-as-judge and reward pipelines, agent evaluation frameworks, rubric and taxonomy design, golden diff and trajectory analysis, multi-turn agent workflows, adversarial prompting, RAG, OpenAI and Claude APIs, AWS Bedrock

Engineering: Python, SQL (MySQL, PostgreSQL, Snowflake), LangChain, LangGraph, Apache Airflow, Docker, AWS (EC2, S3, Lambda), FastAPI, ETL and data validation

EXPERIENCE

Deccan AI — Hyderabad, India

Oct 2025 – Present

Promoted from AI Operations Associate to Associate Product Manager within 5 months; scope expanded in May 2026 to lead the Agentic Pod's RL-Gym program. First engaged through Soul AI (Deccan's expert network) as a part-time AI model evaluator, Apr–Aug 2025, before joining full-time.

Associate Product Manager, Agentic Pod

May 2026 – Present

Program and technical lead for the agentic RL-Gym: task environments, evaluation standards, and new-offering pilots.

- Set technical direction for the non-coding agentic RL-Gym program: multi-turn, multi-server, tool-use task environments for training and evaluating frontier-lab agents, with task-design and QC standards adopted across annotation and QC teams.
- Led a Computer-Use Agent (CUA) pilot end to end — market research, ML/data pipeline design, annotation-platform technical requirements, and packaging of the sample deliverable — resulting in active client interest in a new engagement.
- Ran the annotation review process and reviewer management for a BrowserGym web-agent pilot (sample task creation), setting prompt-quality standards and screening tasks for reward hacking.
- Driving a camera-motion reasoning benchmark program for multimodal models (CameraBench-aligned), from rubric and taxonomy design through pilot data collection (in progress).
- Open every new program, and revisit running ones, with research into model behavior, benchmarks, and market requirements; findings become the new or revised task guidelines and rubrics the teams work to.
- Built LLM-as-Judge and reward/evaluation pipelines scoring agent reasoning, tool use, and trajectories at scale; designed task verifiers and difficulty rubrics defining reward signals and pass/fail criteria.
- Architected synthetic enterprise databases (multi-domain Snowflake schemas with referential integrity) that serve as realistic RL-Gym environments.
- Ran benchmarks across datasets and agent architectures to quantify quality and difficulty; findings drove roadmap and dataset-quality decisions.

Associate Product Manager, Multimodal Pod

Oct 2025 – May 2026

Joined as AI Operations Associate; promoted to APM in Mar 2026.

- Owned multimodal LLM/VLM dataset products end to end: roadmap, workflows, rubrics, and annotation design, from ideation to production.
- Defined annotation strategy, QC processes, and resource planning across freelancer, contractor, and internal teams, managing 15–20 people day to day and up to 50 at peak; owned execution, scaling decisions, and PnL.

- Partnered with clients on solution design and pricing, translating requirements into delivery plans.
- Built multi-turn validation frameworks and Jira-integrated QC question sets governing quality across annotation and QC/SQC stages.
- Benchmarked agent architectures (ReAct, planner-executor), improving task success by ~30%; built early LLM evaluation pipelines with golden datasets and ran pilot calibration and annotator training.

Turing — Remote

Jan 2025 – Oct 2025

Applied LLM Engineer • Contract • Google Gen AI project

- Analyzed 5,000+ LLM agent responses using rubric-based scoring and RLHF-style preference evaluation, improving prompt success rates by ~12%.
- Debugged agent failures by tracing reasoning chains and tool-call sequences, reducing tool-misuse errors by ~20%.
- Performed diff-based analysis of execution traces, rubrics, and golden trajectories to detect alignment gaps in agent workflows.

TELUS Digital AI — Remote

Sept 2024 – Feb 2025

Data Engineer

- Developed Airflow-based ETL pipelines processing 20k+ records/day with automated validation; cut processing time 30% through parallelization.

ArcadeDigitech — Noida, India

Oct 2023 – Aug 2024

Data Analyst

- Delivered forecasting and variance analysis improving budgeting accuracy by 15%; built Power BI dashboards and automated reporting, cutting prep time by 70%.

EDUCATION & CERTIFICATIONS

- PG Diploma, Data Science & Artificial Intelligence — IIT Guwahati | Aug 2023 – Dec 2024
- B.Tech, Computer Science & Engineering — Accurate Institute of Technology and Management | Aug 2019 – Jun 2023
- AWS Solutions Architect – Associate, training course (Edureka) | 2025