Yashvi Vora

LLM Calibrator | LLM Annotation | Prompt Engineering | Team Leadership | RLHF
yashvivora33@gmail.com | linkedin.com/in/yashvi-vora-427871179 | +91 9137186647
yashvivora.pages.dev/?role=ai

LLM Calibrator with experience in LLM evaluation, RLHF, prompt engineering, and function calling assessment, focused on improving reasoning accuracy and reducing hallucinations.

EDUCATION

- **Deakin University** - Melbourne, AU
  Master's in Data Science | Nov. 2024 - Feb. 2026
- **Duke University** - Durham, US
  PG Certificate in Product Management | Jan. 2024 - Jul. 2024
- **The University of Texas at Austin** - Austin, US
  PG Program in Artificial Intelligence & Machine Learning | Feb. 2023 - Mar. 2024
- **University of Mumbai** - Mumbai, IN
  B.Tech in Information Technology | Aug. 2018 - Jun. 2022

EXPERIENCE

Calibrator, Turing | May 2026 - Present | Remote

- The final quality check for a 68-person team (8 squad leaders and 60 AI trainers) before their evaluation data goes to model training.
- Write model-breaking tasks (a problem statement, test cases and verifiers) built so the model fails on its own (under 50% of tests pass) but gets through once given the right guidance document; scenarios span business, technology, web & UI, and software engineering.
- Review the leaders' work for accuracy and consistency, standardizing the rubrics and scoring so everyone rates the same way.
- Spot recurring quality gaps across teams and push corrective feedback back down at scale.
- Keep team output lined up with the model's performance benchmarks.
- Set the standard for scoring function-calling work, so tool use, API-call accuracy and output format stay consistent across leaders.

Squad Lead, Turing | May 2025 - May 2026 | Remote

- Led a team of 10 AI trainers across agentic code-generation and function-calling evaluation, keeping quality consistent.
- Reviewed model outputs for logical consistency and hallucination-free reasoning, giving calibration-aligned ratings and feedback.
- Coached the team on prompting and on how to check a model's reasoning.
- Wrote the evaluation guidelines the team followed.
- Checked model outputs on function-calling tasks: tool selection, parameter extraction and response format.

AI Trainer / LLM Evaluator, Turing | Sep. 2024 - May 2025 | Remote

- Trained an agentic model on real GitHub PRs: gave it access to the codebase, wrote step-by-step reasoning to steer it toward the fix described in the PR, and marked down wrong attempts with written critiques (RLHF).
- Checked the model's tool use across APIs like Slack, Teams and Jira; when it called the wrong API or passed bad parameters, flagged it and rewrote the call correctly.
- Ran RLHF-style reviews on model outputs to improve accuracy, coherence and reasoning, and annotated LLM trajectories linking GitHub issues to their code patches.
- Caught hallucinations and gaps in reasoning and gave detailed feedback to raise response quality.
- Used prompting to steer outputs toward the right answer, and Python and SQL to validate tasks; got through 15-20 evaluation tasks a week on internal tools.

Data Science Intern, Liases Foras | Oct. 2023 - Mar. 2024 | Mumbai, India

- Used the LLaMA model to pull address parts (city, street, locality, building name) out of raw address text.
- Translated Marathi addresses to English and ran NER to tag the entities in them.
- Scored the model on IOB2 tags with precision, recall, accuracy and F-score, and checked translation quality with the BLEU measure.
- Deployed the model behind a FastAPI service in Python and packaged it with Docker.
- Sent the extracted entities to a geolocation API for their coordinates and saved everything to a database.

Data Science Intern, FlixDrop | Jul. 2023 - Oct. 2023 | Cattle Health Monitoring · Mumbai, India

- Predicted cattle rumination periods for an Animal Behavior Prediction project, so farmers could look after their livestock better.
- Built the target labels with KMeans clustering, then trained and compared several supervised models: Logistic Regression, KNN, SVM, Decision Tree, Bagging, Random Forest, AdaBoost and Gradient Boosting.

Software Developer, Arcon TechSolutions | Dec. 2022 - Aug. 2024 | Mumbai, India

- Built RESTful APIs and .NET desktop apps, working over HTTP and integrating third-party APIs.
- Rolled out the company's Identity Access Management (IAM) product for four clients (NGCP, Orient, Deutsche Telekom and Unity Small Finance Bank); Orient was the first to go live on it.
- Now building the provisioning engine for a new IAM-Converged Identity (CI) product that handles the full user lifecycle, from provisioning through to deprovisioning.

DevOps Engineer, Arcon TechSolutions | Jun. 2022 - Dec. 2022 | Mumbai, India

- Managed code repositories and worked with the dev teams using Git.
- Set up and maintained CI/CD pipelines on Jenkins and GitLab.

TECHNICAL SKILLS

- **LLM & AI:** LLM Evaluation, RLHF, Prompt Engineering, Model Alignment, Hallucination Detection, Reasoning Validation, Function-Calling Assessment, LLM Inferencing, Gemini Vision API
- **NLP & Deep Learning:** NLP (Lexical, Syntactic, Semantic), Named Entity Recognition (NER), Deep Learning (CNN, RNN, Transfer Learning), Computer Vision
- **Machine Learning:** Model Selection & Evaluation, Clustering, SVM, Decision Trees, Boosting, PCA, Statistical Analysis, Hypothesis Testing
- **Programming:** Python, SQL, TypeScript, JavaScript, C#, Data Structures & Algorithms
- **Data Handling:** Data Annotation, Data Validation, Structured Data Analysis, Entity Resolution & Deduplication
- **Frameworks & APIs:** PyTorch, TensorFlow, Keras, FastAPI, Flask, React, Next.js, Node.js / Express, REST APIs
- **Databases:** PostgreSQL (Prisma), SQLite, Redis (Upstash), Supabase, IndexedDB
- **MLOps & Cloud:** MLflow, Docker, Kubernetes, KubeFlow, AWS, Azure, GCP, Vercel, Cloudflare Pages, Render
- **Tools:** Git, GitLab, Jenkins, GitHub Actions, CI/CD, Vitest

PROJECTS

Library Vault | library-vault.pages.dev

- Offline book and manga tracking with ISBN scanning, book search, duplicate checks, virtual shelves, missing-volume tracking, Google and email sign-in, owner cloud backup, and a public library showcase.
- Reading sessions, streaks, goals, progress and pace dashboards, a reading calendar, shareable yearly recaps, Goodreads and StoryGraph imports, Kindle clippings, and exports.

Core stack: React / TypeScript | IndexedDB | Supabase

Rooster Daily | roosterdaily.com

- Sourced news briefs with topic desks, duplicate-story grouping, search, entity timelines, related coverage, a Coverage Lens, and country and city editions.
- 26 interface languages, read-aloud, swipe reading, an official-statements world map, follow and breaking alerts, personal newsletters, membership workflows, and an admin dashboard for feed and delivery health.

Core stack: React / TypeScript | Node / Express | PostgreSQL / Prisma

Diet Accountability | diet-accountability-app-one.vercel.app

- Meal-photo nutrition estimates with food-level calorie ranges, meal and weight logging, craving and trigger insights, resistance streaks, consistency scores, and daily to yearly summaries.
- Timezone-aware reminders with snooze, Google sign-in, a read-only guest demo, offline use, cloud backups, and CSV and ZIP exports.

Core stack: Next.js / TypeScript | Gemini Vision | SQLite / Redis

Xteink Library | xteink-library.onrender.com

- Free ebook discovery across eight sources with rights filtering, duplicate and edition checks, personalized ranking, an approval queue, a searchable and filterable library, scan history, and scheduled discovery.
- EPUB and PDF workflows, pooled Google Drive storage, local Wi-Fi transfer, QR and phone sharing, and an OPDS catalogue for e-reader access.

Core stack: React / TypeScript | Node / Express | SQLite

MeepleMind | Personal Project

- Board-game discovery with capability labels, player and match setup, saved sessions and resume, rule guides, interactive demos, session history, and rule questions.
- Photo capture and AI-provider integration, confidence indicators, manual corrections, and dedicated scoring, setup and move checks for Catan, Azul, and Wingspan.

Core stack: React / TypeScript | Node / Express | SQLite / Prisma