

# Yashowardhan Singh Tomar

## AI EVALUATION ENGINEER | RLHF SPECIALIST | LLM QUALITY ANALYST

Indore, India | Open to remote and relocation | yashtomar10122@gmail.com | +91-97549-50809

[linkedin.com/in/yashowardhansinghtomar](https://www.linkedin.com/in/yashowardhansinghtomar) | [github.com/yashowardhansinghtomar](https://github.com/yashowardhansinghtomar)

---

### PROFILE

AI engineer focused on LLM evaluation, RLHF workflows, rubric design, and model-quality analysis. Hands-on experience contributing to RLHF pipelines for xAI and Meta Llama 4 through Turing-sourced work, including single-turn and multi-turn prompt design, side-by-side response evaluation, preference judgments, and ground-truth corrections. Public portfolio now includes config-driven LLM evaluation and RAG evaluation workbenches with validation, scoring, failure tags, and generated reports.

[LLM Evaluation](#) | [RLHF](#) | [Prompt Evaluation](#) | [Rubric Design](#) | [Preference Data](#) | [Failure Analysis](#) | [Python](#) | [RAG Evaluation](#)

### EXPERIENCE

#### Jr. AI / Data Science Engineer | Engineer Master

Sep 2024 - Jan 2025

- Contributed to RLHF training pipelines for xAI and Meta Llama 4 through Turing-sourced work, evaluating model behavior across Python and data-science tasks.
- Designed single-turn and multi-turn prompts to stress-test correctness, instruction following, reasoning quality, and response clarity.
- Reviewed side-by-side model responses using structured 5-point rubrics and submitted preference judgments for alignment workflows.
- Created ground-truth corrections and failure notes to improve prompt-response quality and reduce recurring model errors.
- Integrated TTS and STT models into PreCall AI, a production AI calling workflow, adding applied systems experience beyond evaluation-only work.

#### Business Development Associate | Byju's

2020 - 2022

- Managed high-volume B2C sales of online education products with weekly revenue targets.
- Developed consultative discovery, stakeholder communication, and objection-handling skills across a large customer base.

### SELECTED PROJECTS

#### llm-evaluation-lab

<https://github.com/yashowardhansinghtomar/llm-evaluation-lab>

Python, JSONL, CLI, Markdown/JSON reports

- Built a config-driven LLM evaluation workbench with rubric validation, pairwise scoring, agreement checks, and failure-mode reporting.
- Modeled RLHF-style review data with prompt, candidate responses, weighted dimensions, human preference, notes, and failure tags.

#### rag-evaluation-workbench

<https://github.com/yashowardhansinghtomar/rag-evaluation-workbench>

Python, BM25 retrieval, citation checks, reports

- Evaluates RAG answers for retrieval recall, citation precision/recall, required-fact coverage, grounded answer rate, and failure tags.

## SKILLS

---

- **Evaluation:** LLM evaluation, RLHF, pairwise comparison, rubric design, prompt regression, failure tagging
- **RAG Quality:** Retrieval recall, citation precision/recall, groundedness checks, required-fact coverage
- **Tools:** Python, JSONL, Markdown/JSON reporting, LangChain, OpenAI API, Groq, Ollama, Hugging Face
- **Interfaces:** CLI tools, Streamlit, Gradio

## EDUCATION

---

### **Prestige Institute of Management & Research**

Bachelor of Business Administration (BBA), Global Marketing & Brand Management | 2017 - 2020

## CERTIFICATIONS

---

Data Analyst - Edubridge | Generative AI - Debugshala