Archana Yadav

+91-7021750741 | archanaqre@gmail.com | https://archanaqre.github.io/

Education

Birla Institute of Technology & Science, Pilani — Hyderabad, India
Master of Engineering - Computer Science (CGPA: 8.74/10) — *2021 – 2023*

University of Mumbai — Mumbai, India
Bachelor of Engineering - Computer Engineering (CGPA: 7.37/10) — *2016 – 2020*

Experience

Research Intern — Aug. 2025 – Present
Microsoft Research, Akshay Nambi

- Built **Magentic Marketplace**, an open-source environment for studying two-sided agentic markets, enabling controlled evaluation of search, communication, negotiation, welfare, manipulation resistance, and behavioral biases in multi-agent economic systems.
- Developed a failure-centric evaluation framework for agentic systems by constructing a unified execution-level failure taxonomy and trajectory diagnostics that localize execution failures beyond final-task success.
- Developed synthetic web environments and curated task distributions for **Fara 1.5**, enabling controlled generation of computer-use-agent trajectories that expanded coverage of interaction skills underrepresented in real-world web data.

Project Research Associate — Dec. 2023 – Aug. 2025
CFILT Lab, Indian Institute of Technology, Bombay, Prof. Pushpak Bhattacharyya

- Published a first-author **ACL 2025 Industry Track Oral** paper by developing a Bloom's Taxonomy-based framework that systematically identifies cognitive failure points in conversational AI systems.
- Developed **GriceWise**, a reference-free evaluation framework achieving Spearman correlations above 0.6 with human judgments, and constructed the **Blooms-FQ** dataset containing 3,750 follow-up questions for conversational AI evaluation.
- Built a multilingual automated closed-captioning system for Indian TV content using Whisper ASR, audio-event detection, and a curated 150-hour Hindi, Marathi, and Tamil television corpus.

NLP Research Intern — Jan. 2023 – July 2023
Speech Lab/@SCSE, Nanyang Technological University, Prof. Chng Eng Siong

- Conducted comprehensive evaluation of the implemented statistical LMs using SRILM toolkit for Text Generation of Conversational Singaporean English and Code-switching Malay-English to augment the IMDA corpus by 60%.
- Implemented novel techniques using BERTScore, BLEU score, and token length for filtering out higher-quality code-switched sentences, achieving a 6.5% decrease in the perplexity score of the Language Model of augmented code-switched sentences.

NLP Intern — July 2022 – March 2023
Textify AI

- Developed and deployed a passive-to-active voice conversion service using the PassivePy and spaCy framework, resulting in a 90% reduction of passive voice in university essays.
- Researched and analysed email quality grading metrics, including redundancy check, tone conversion, and grammar check, with enhanced email clarity and professionalism.
- Developed and fine-tuned several models for multi-label classification, summarisation, question-answering systems and formality style transfer.

Teaching Assistant — Sept. 2021 – Dec. 2022
Birla Institute of Technology & Science, Pilani

- Courses: Approximation Algorithms, Deep Learning, Advanced Algorithms and Complexity.

Publications

- ***From Recall to Creation: Generating Follow-Up Questions using Bloom's Taxonomy and Grice's Maxims***: Archana Yadav, Harshvivek Kashid, Pushpak Bhattacharyya. **ACL 2025 (Industry Track) - Oral Presentation** [Paper Link]
- ***Magentic Marketplace: An Open-Source Environment for Studying Agentic Markets***. Submitted to ICLR 2026 *(Under Review)* [Paper Link]
- ***Whisper-ing with Subtitles: Making TV Shows Chatty for All***: Archana Yadav, Pushpak Bhattacharyya. Submitted to ARR August 2024 *[Pre-print]*

Projects

Document Classification on FinCausal and BBC News Classification Dataset *[GitHub]*

- Tested Hierarchical Attention Network architecture for document classification of BBC News dataset, with dynamic batching and achieving ∼90% accuracy.
- Implemented BERT model for the FinCausal dataset classification task with an F1-score of 0.97.

Conversational Dataset Fine-Tuning of LLaMA-3 using Unsloth and LoRA *[GitHub]*

- Fine-tuned LLaMA-3 (1B and 3B) models using Unsloth and LoRA on multi-turn conversational data subset (FineTome-100k), achieving improved coherence and response formatting in under 5 minutes on a single GPU.

Simulation of PoET - Proof of Elapsed Time - Blockchain Consensus Mechanisms

- Developed a blockchain-based land management platform using Proof of Elapsed Time (PoET) consensus, optimising for secure, energy-efficient property transactions and transparent ownership tracking.
- Implemented cryptographic tokenisation for property assets, leveraging Merkle Root to ensure tamper-proof transaction records.

Technical Skills

Programming Languages: Python, C/C++, Shell Script
Frameworks and Libraries: vLLM, PyTorch, Huggingface, FairSeq, SciPy, fast.ai, Unsloth, LangChain, Tensorflow
Developer Tools: Git, Docker, Google Cloud Platform
Research Areas: Agentic Systems, AI Evaluation, Conversational AI, Large Language Models, Synthetic Environments, Speech Processing