

JEBISH PURBEY

📞 (+977) 9842043988 ✉ jebishpurbey@gmail.com 🔗 [linkedin.com/in/jebishpurbey](https://www.linkedin.com/in/jebishpurbey) 🎓 Scholar

EDUCATION

Pulchowk Campus, Tribhuvan University, Nepal

Dec 2023

Bachelor of Engineering - Electronics and Communication Engineering

Valedictorian, 83.75%

Honors: Ncell Excellence Award for Rank holder, Ncell Scholarship Award for third year, Eight Distinction Scholarship (Merit list), Entrance Rank Holder Scholarship

RESEARCH EXPERIENCE

ML Agents Community Lead

June 2025 – Present

Cohere Labs

- Leading **20+ researchers** on a large-scale project benchmarking LLM reasoning across 50 task categories, including Mathematical Reasoning, Coding, and Instruction Following (*Manuscript in preparation*).
- Designed the evaluation (LLM-as-Judge) and inference pipelines for 6 reasoning approaches (Chain-of-Thought (CoT), Skeleton-of-Thought (SoT), Buffer of Thoughts (BoT), Least-to-Most (LtM), Chain-of-Verification (CoVE), and Graph of Thoughts (GoT)) as part of the community project for 8 different Language Models.

Researcher

Sep 2024 – Present

M2ai

- Leading research projects on multilingual, multimodal, and low-resource NLP, secured by **\$20,000+ in grants** from OpenAI, Anthropic, and Cohere.
- Developed **Mantra-14B**, a state-of-the-art Hindi-English bilingual model, by fine-tuning Phi and Qwen on a self-curated, empirically balanced instruction dataset (**IJCNLP-AAACL 2025**).
- Built a robust AI-generated text detection model by training Longformer on novel multilingual multi-generator dataset (23 languages, 12 generators) to detect attributions via token-boundary analysis
- Implemented Lexical re-ranking for Semantic Retrieval of regulatory documents for Regulatory Question Answering system, achieving Recall@10 of 0.82 in ObliQA dataset (**RegNLP COLING 2025**).
- Building a unified LLM for 22 South Asian languages by merging fine-tuned language-specific models trained on custom datasets.

Research Fellow

Dec 2024 – May 2025

Traversaal.ai

- Architected **AgentPro**, a REACT-based agentic framework optimized for complex data science workflows and automated reasoning as the back-bone for Olive OS.
- Constructed a comprehensive benchmark across 8 task categories to evaluate agentic architecture performance and analyze hyperparameter sensitivity in data-science contexts (**IJCNLP-AAACL 2025**)

Community Researcher

Sep 2023 – May 2025

Cohere Labs

- Led a study on **cultural knowledge disparities** in VLMs by evaluating performance on determining the origin of culturally significant images for over 200 countries (*Advisors: Prof. Desmond Elliott & Alejandro R. Salamanca, IJCNLP-AAACL 2025*).
- Co-developed **INCLUDE**, a multilingual benchmark dataset specifically designed to assess localized and regional knowledge in language understanding (**ICLR 2025**).
- Contributed to **Kaleidoscope**, a multimodal evaluation benchmark utilizing global exam questions to assess the regional knowledge of VLMs.
- Executed inference testing and linguistic evaluation for the AYA-expanse model, serving as the primary Language Ambassador for Hindi and English.

Research Assistant & Undergraduate Researcher

Apr 2021 – Apr 2025

Pulchowk Campus, IoE, Tribhuvan University

- **Hate Speech Detection (Aug 2024 – Apr 2025):** Compiled a Nepali-Hindi hate-speech dataset with fine-grained socio-cultural annotations, and developed multitask baselines with interpretability and cross-lingual analysis (*Advisors: Assoc. Prof. Sanjeeb P. Panday & Asst. Prof. Bikash Balami, Manuscript Under Revision*).
- **NLP Summarization (Jan 2023 – Jul 2023):** Fine-tuned BART architectures on a self-collected corpus of Nepali news and folk tales to establish abstractive summarization baselines for low-resource languages. (*Advisor: Asst. Prof. Bikash Balami*).
- **Undergraduate Thesis (Apr 2021 – Apr 2022):** Implemented a Deep Learning-based Virtual Try-On system (CP-VTON+ architecture) and integrated it into a React/Flask web application for real-time human parsing and pose estimation. (*Advisor: Assoc. Prof. Sanjeeb P. Panday*).

Research Assistant

May 2024 – Nov 2024

Pulchowk Campus, IoE, Tribhuvan University

Advisor: Asst. Prof. Santosh Giri

- Designed a custom MobileNetV2-based CNN for cultural heritage detection in Patan Durbar Square, achieving an F1 score of 0.98 (**CANDAR 2025**).
- Conducted ablation studies to analyze the impact of dropout rates and classification head depth on transfer learning efficiency.

PUBLICATIONS

[12] “DSBC: Data Science task Benchmarking with Context engineering.” Ram Mohan Rao Kadiyala, **Jebish Purbey**, Siddhant Gupta, Giulio Martini, Ali Shafique, Suman Debnath, Hamza Farooq. *IJCNLP-AAACL 2025*. arXiv:2507.23336.

[11] “Uncovering Cultural Representation Disparities in Vision-Language Models.” **Jebish Purbey***, Ram Mohan Rao Kadiyala*, Siddhant Gupta*, Srishti Yadav, Suman Debnath, Alejandro Salamanca, Desmond Elliott. *IJCNLP-AAACL 2025 (Findings)*. arXiv:2505.14729

[10] “Improving Multilingual Capabilities with Cultural and Local Knowledge in LLMs.” Ram Mohan Rao Kadiyala, Siddhartha Pullakhandam, Siddhant Gupta, **Jebish Purbey**, ..., Hamza Farooq. *IJCNLP-AAACL 2025*. arXiv:2504.09753

[9] “Evaluating Generalization Capabilities of LLM-Based Agents in Mixed-Motive Scenarios Using Concordia.” Chandler Smith, Marwa Abdulhai, ..., **Jebish Purbey**, ..., Zixia Jia, Ziyang Wang. *NeurIPS 2025*. [NeurIPS Poster]

[8] “Global PIQA: Evaluating Physical Commonsense Reasoning Across 100+ Languages and Cultures.” Tyler A. Chang, Catherine Arnett, ..., **Jebish Purbey**, ..., Jiayi Wang, David Ifeoluwa Adelani. *Preprint*. arXiv:2510.24081

[7] “Robust and Fine-Grained Detection of AI Generated Texts.” Ram Mohan Rao Kadiyala, Siddhartha Pullakhandam, Kanwal Mehreen, Siddhant Gupta, **Jebish Purbey**, ..., Hamza Farooq. *Under Review*. arXiv:2504.11952

[6] “Kaleidoscope: In-language Exams for Massively Multilingual Vision Evaluation.” Israfel Salazar, Manuel Fernández Burda, ..., **Jebish Purbey**, ..., Sara Hooker, Marzieh Fadaee. *Under Review*. arXiv:2504.07072

[5] “INCLUDE: Evaluating Multilingual Language Understanding with Regional Knowledge.” Angelika Romanou, Negar Foroutan, ..., **Jebish Purbey**, ..., Antoine Bosselut. *ICLR 2025 (Spotlight)*. arXiv:2411.19799

[4] “Lexical Reranking of Semantic Retrieval (LeSeR) in Regulatory Domains.” **Jebish Purbey**, Drishti Sharma, Khawaja Murrad, Siddhant Gupta, Nikhil Manali, Ram Mohan Rao Kadiyala. *RegNLP at COLING 2025*. [ACL]

[3] “SeQwen at the Financial Misinformation Detection Challenge Task.” **Jebish Purbey**, Siddhant Gupta, Nikhil Manali, Siddhartha Pullakhandam, Drishti Sharma, Ashay Srivastava, Ram Mohan Rao Kadiyala. *FinNLP at COLING 2025*. [ACL]

[2] “1-800-SHARED-TASKS @ NLU of Devanagari Script Languages.” **Jebish Purbey**, Siddartha Pullkhandam, Kanwal Mehreen, Muhammad Arham, Drishti Sharma, Ashay Srivastava, Ram Mohan Rao Kadiyala. *CHiPSAL at COLING 2025*. [ACL]

[1] “A transfer learning-based approach for automatic monument detection.” Santosh Giri, **Jebish Purbey**, Sunil Adhikari, Atsushi Ito, Baburam Dawadi, Sushant Chalise, Sanjivan Satyal. *GCA CANDAR 2025*.

TEACHING EXPERIENCE

Instructor

Jan 2024 – Aug 2025

National College of Engineering, Tribhuvan University

Lalitpur, Nepal

- Instructed undergraduate engineering students in *Artificial Intelligence (CT710)*, *Embedded Systems (CT655)*, *Numerical Methods (SH603)*, and *Computer Programming (CT101)*.
- Managed curriculum delivery, laboratory sessions, and student evaluations for multiple semesters.

Instructor

May 2024 – Apr 2025

Pulchowk Campus, IoE, Tribhuvan University

Lalitpur, Nepal

- Delivered lectures and supervised lab sessions for *Artificial Intelligence (CT710)*, *Digital Signal Analysis and Processing (EX710)*, and *Computer Programming (CT101)*.
- Supervised undergraduate seniors in their AI coursework projects, and oversaw the presentations.

INDUSTRY EXPERIENCE

Associate AI Engineer

June 2025 – Present

Dogma International (UK-Based)

Kathmandu, Nepal

- Co-developed the **Scottish Charity Agent** for the **Scottish Government**, implementing a RAG-based pipeline to automate large-scale compliance document reviews for OSCR regulations.
- Deployed a full-stack **Feasibility Analysis Agent** on Azure (React/FastAPI) that utilizes LLMs and semantic search to automatically estimate feature development requirements from user prompts.
- Engineered a production-grade OCR and LLM pipeline on Azure to extract Penalty Charge Numbers (PCN) for a UK transport client, driving automation in fleet management workflows.

FELLOWSHIP

AI Microdegree Fellow

Apr 2024 – Oct 2024

Fusemachines

Remote

- **Program Overview:** Selected for a competitive 6-month intensive AI specialization covering Deep Learning, Natural Language Processing, and Computer Vision under industry mentorship.
- **Capstone Project:** Engineered a mental health support chatbot by fine-tuning the **Llama-3.2** large language model on conversational health datasets to enhance empathetic response generation.
- **Technical Implementation:** Implemented Retrieval-Augmented Generation (RAG) to ground model responses in verified psychological resources, reducing hallucinations and ensuring domain safety.

LEADERSHIP & SERVICE

LOCUS Organization Committee

Nov 2019 – Feb 2022

Exhibition & Event Manager

Lalitpur, Nepal

- Orchestrated logistics and technical safety protocols for **RoboWarz** (2020–2021), managing brackets for 20+ combat robots in Nepal’s largest student technological festival.
- Instructed high school students in software and electronics workshops (Arduino/Python) to foster STEM interest in the local community.
- Managed exhibition stalls and led a volunteer team for crowd mobilization during the 3-day annual tech fest.

Hack-A-Week

Jan 2020

Finalist (Health Category)

Lalitpur, Nepal

- Prototyped a smart-vaccination tracking system including ambulance routing optimization, reaching the final round of the competition (Runner-up).