

# PUJAN PANT

Kathmandu, Nepal ♦ +977-9869671443 ♦ pantpujan017@gmail.com  
[linkedin.com/in/pujan-pant](https://www.linkedin.com/in/pujan-pant) ♦ [github.com/pantpujan017](https://github.com/pantpujan017) ♦ [pantpujan017.github.io](https://pantpujan017.github.io)

## EDUCATION

### BSc (Hons), Computer Science

2022 – 2025

*Herald College Kathmandu · Affiliated to University of Wolverhampton, UK — 76.44%, First Class Honours*

- Final Year Project: Self-Driving Car — embedded object detection and autonomous navigation on Raspberry Pi
- Relevant Coursework: Machine Learning & AI, Big Data Analytics, Distributed & Cloud Systems, Algorithms & Data Structures, Computational Mathematics

## SKILLS

**LANGUAGES** Python, C/C++, SQL

**ML / DEEP LEARNING** PyTorch, TensorFlow, Scikit-learn, OpenCV, HuggingFace Transformers, NetworkX

**LLM / GENAI** QLoRA & PEFT, LangChain, RAG, Pinecone, prompt engineering, strict-JSON structured extraction, OpenAI Batch API

**MLOPS & CLOUD** Docker, Azure, GitHub Actions, MLflow, DVC

## RESEARCH EXPERIENCE

### Independent Researcher — LLM-SC Framework Study

April 2026 – June 2026

*Under voluntary mentorship of Prof. Hari Sitaula, Montana Technological University*

- Extended the LLM–Social Computing (LLM-SC) framework from museum settings to living sacred heritage sites — "Extending the LLM-SC Framework to Living Sacred Heritage Sites: A Case Study of Pashupatinath Temple, Kathmandu" (manuscript in preparation, 2026)
- Assembled an 18-year TripAdvisor corpus of 3,939 reviews (Oct 2008 – Apr 2026), the largest computational dataset yet compiled for a South Asian sacred heritage site; 2,485 of 2,500 sampled reviews (99.4%) returned schema-valid output
- Designed a 12-type domain-adaptive entity taxonomy through grounded-theory analysis of 300 reviews, adding 8 sacred-heritage classes absent from CoNLL-style schemas; these account for 64.8% of the 11,314 entities extracted
- Built a GPT-4o extraction pipeline over five prompt-refinement cycles — OpenAI Batch API, checkpointed and resumable, strict-JSON schema validation,  $T = 0.1$  — selecting GPT-4o against four alternatives (Llama-3.1-8B, Llama-3.3-70B, GLiNER + SetFitABSA, GPT-4.1-mini) on the two hardest extraction targets
- Constructed a 5,110-node typed semantic network with five directed relation types (1,124 causal and 1,068 contrast edges) in NetworkX, using Louvain community detection and degree-distribution entropy for structural comparison
- Benchmarked against a traditional NLP pipeline (NLTK MaxEnt NER + VADER) and BERT baselines (dslim/bert-base-NER, DeBERTa-v3 ABSA): ~87% sentiment accuracy vs. 73.0% and 68.8%; both baselines missed all 1,905 in-text occurrences of the 13 sacred-heritage terms tested
- Identified a sign inversion with direct management consequences: access fairness (65.1% negative), crowd management (66.8%) and environment (61.4%) are majority negative under LLM-SC, yet BERT ABSA scored the same aspects 74–86% positive
- Introduced a dual-valence sentiment class capturing co-present grief and transcendence at cremation-active sites — the first computational treatment of this state in heritage tourism

### Research Trainee

June – Sept 2024

*Nepal Research & Collaboration Center (NRCC), Kathmandu*

- Selected from a competitive national pool to conduct a structured literature review of deep-learning architectures for medical computer vision, surveying 40+ papers
- Co-reviewed a paper on deep learning and computer vision in the medical domain under a PhD candidate (University of South Dakota); contributed the gap analysis and methodology critique sections

## INDUSTRY EXPERIENCE

---

### AI/ML Engineer (Full-time)

Jan 2026 – Present

Crown Institute, Kathmandu

- Building LLM-based NLP systems for review integrity on Trustkie, a review and trust platform for local businesses in Nepal
- Developing the fake-review detection pipeline end to end — review ingestion, LLM-based authenticity scoring, and flagging of suspicious user-generated content (in active development)

### AI Instructor

May – June 2025

Kathmandu Metropolitan City

- Designed and delivered 300+ hours of applied AI and Python curriculum to 75 students across Kathmandu Metropolitan City schools
- Built hands-on labs on supervised learning, EDA, and LLM API integration; 92% of students completed capstone projects

### AI/ML Engineer Intern

Sept 2024 – Feb 2025

Scala Technologies, Kathmandu

- Developed a brain tumour segmentation pipeline (DenseNet + ResNet + attention) on the BraTS benchmark; Dice 92% on 2D axial and 97% on 3D volumetric modalities
- Engineered a tag-based POI recommendation and route-optimisation system using Dijkstra and A\*, matching destinations to user-selected interest tags
- Deployed the inference pipeline via Docker, Azure Container Registry and GitHub Actions CI/CD, cutting the deployment cycle from 3 hrs to 2 hrs

## PROJECTS

---

### NepSLM — Parameter-Efficient Adaptation for Romanized Nepali

Fine-tuned three 2–3B-parameter LLMs on a Romanized Nepali corpus using QLoRA, with a "Golden Point" checkpoint-recovery technique for training stability under low-resource conditions — a cross-scale study of how sub-7B models adapt to a language with no standard benchmark or tokenizer for its Latin-script register.

[github.com/pantpujan017/NepSLM](https://github.com/pantpujan017/NepSLM)

### Multi-Agent AI Credit Scoring

Five specialised agents (Document, Income, Scoring, Compliance, Decision) performing consensus-based verification of stated document income against e-wallet transaction history.

[github.com/sagarnewpane/autonomous-credit-and-lending-orchestrator](https://github.com/sagarnewpane/autonomous-credit-and-lending-orchestrator)

### Medical RAG Chatbot

Flask service retrieving evidence-grounded answers from medical PDFs via Pinecone vector search over a Groq-hosted LLM.

[github.com/pantpujan017/RAG-LLM-medical-chatbot](https://github.com/pantpujan017/RAG-LLM-medical-chatbot)

### Self-Driving Car — Embedded Autonomy

Physical prototype running real-time on-board object detection at 20 FPS with 50 ms end-to-end latency.

[github.com/pantpujan017/Self-Driving-Car](https://github.com/pantpujan017/Self-Driving-Car)

### Brain Tumor Segmentation — BraTS Benchmark

2D and 3D segmentation models with a dual-encoder + attention architecture on the BraTS benchmark dataset.

[github.com/pantpujan017/brain\\_tumor\\_segmentation](https://github.com/pantpujan017/brain_tumor_segmentation)

## CERTIFICATIONS

---

- Machine Learning Specialization — Stanford University / DeepLearning.AI (Coursera)
- Generative AI with LLMs — DeepLearning.AI (Coursera)
- Bayesian Algorithms for Self-Driving Cars — Coursera

## ACHIEVEMENTS AND HONOURS

---

- GIBL Hackathon 2026 — paper selected for the onsite final round; competed at Park Village Resort, Kathmandu (May 2026)
- Mahatma Gandhi Scholarship Scheme, 2018 — awarded by the Embassy of India, Government of India, for high-school studies