

SANTOSH GAIRE SHARMA

Osaka, Japan · +81 90-9719-3462 · gairesantosh509@gmail.com
github.com/Santey624 · [Portfolio](#) · Permanent Resident of Japan

EDUCATION

Doshisha University, Kyoto, Japan Apr 2025 – Sep 2027 (expected)
M.S., Information and Computer Science, Graduate School of Science and Engineering (ISTC)
GPA: 4.10 / 4.50 · Advisor: Prof. Kimiaki Shirahama (CCILAB)

Tribhuvan University, Kathmandu, Nepal 2019 – 2023
B.E., Computer Engineering, Advanced College of Engineering and Management, Institute of Engineering
First Division, 71.76% ($\approx 3.6 / 4.0$) · Thesis advisor: Dr. Surendra Shrestha

RESEARCH INTERESTS

Low-resource and cross-lingual NLP · Medical NLP and patient-facing clinical text generation · Evaluation of large language models · Text summarization · Multimodal and multi-sensor representation learning

PUBLICATIONS

[1] Sharma, S. G., Pujara, R., Aryal, P., & Shrestha, S. (2024). Refinetograph: A Machine Learning Approach Toward Image Enhancement. *International Journal of Information Communication Technology and Digital Convergence*, 9(1), 49–60.

Manuscripts in Preparation

[2] Sharma, S. G., & Shirahama, K. NepMedJP: A Cross-Lingual Summarization Approach Toward Japanese–Nepali Medical Text Generation. Target venue: COLING 2027.

[3] Sharma, S. G., & Shirahama, K. Cross-Sensor Attention for Diffusion-Based Wearable Sensor Reconstruction. In preparation.

RESEARCH EXPERIENCE

NepMedJP: Cross-Lingual Medical Summarization (Japanese–Nepali) 2025 – Present
Independent research project, Doshisha University · Advisor: Prof. Kimiaki Shirahama

- Constructed **NepMedJPBench**, the first benchmark for patient-facing Japanese–Nepali medical summarization: 100 parallel source–summary pairs with clinician-authored reference summaries.
- Developed a clinician-grounded evaluation rubric, drawing candidate attributes from the medical summarization, factuality, cross-lingual evaluation, and patient-communication literature, then validating them through a clinician survey of importance ratings and solicited criteria. Introduced two novel cross-lingual attributes: *clinical meaning preservation* and *medical terminology transfer*.
- Ran a multi-rater clinician evaluation across the full attribute set, establishing gold-standard human judgments for the benchmark.
- Benchmarked task-specific and frontier systems to characterize architectural behavior on cross-lingual medical summarization: pretrained and fine-tuned mT5 (encoder–decoder), Qwen3-4B in thinking and non-thinking configurations (decoder-only), Qwen3.5-4B (multimodal), and frontier LLMs.
- Meta-evaluated the evaluation methods themselves, testing whether automatic metrics (ROUGE-1/2/L, sacreBLEU, chrF++, BERTScore, BARTScore, medical entity-level P/R/F1) and LLM-as-a-judge approximate clinician judgment, and identifying where physician assessment remains irreplaceable.

Research Assistant | CCILAB, Doshisha University May 2026 – Present
Cross-sensor diffusion for wearable sensor reconstruction · Advisor: Prof. Kimiaki Shirahama

- Architected a cross-sensor diffusion framework that reconstructs missing wearable IMU signals, enabling reliable Human Activity Recognition (HAR) under incomplete sensing conditions.
- Built a complete training and evaluation pipeline over 14 body-worn IMU sensors from the OPPORTUNITY dataset, with locomotion and high-level activity labels.
- Designed three missing-sensor training regimes to model realistic sensor-failure patterns: 35% whole-body-location dropout, 40% single-sensor dropout, and 25% random 2–3 sensor dropout.

- Trained a DDPM-based reconstruction model with a downstream C-LSTM-A activity classifier, benchmarking reconstructed signals against mean-fill imputation.
- Demonstrated that diffusion-based reconstruction significantly outperforms statistical imputation when missing sensors are strongly correlated with observed ones, while identifying regimes where simple imputation remains competitive.
- Extending the evaluation to the CogAge and PAMAP2 datasets in collaboration with a visiting researcher.

TEACHING & MENTORING

Student Tutor | Doshisha University

Apr 2026 – Present

- Mentor incoming master's students on research-software setup, Docker environments, and laboratory computing infrastructure.
- Advise on research implementation, experimental design, code debugging, and academic paper writing.

HONORS & AWARDS

- Nominated by the Faculty of Science and Engineering (ISTC) as one of four students for the KMMF Scholarship, Doshisha University.
- Tuition Reduction Scholarship (50%) for privately financed international students, Doshisha University.
- Government-registered Computer Engineer, Nepal Engineering Council.

PROFESSIONAL EXPERIENCE

AI Engineer Intern | SoranoAI (Stanford-affiliated startup)

Sep 2025 – Apr 2026

- Built an LLM router using Claude Haiku 4.5 to classify incoming queries and dispatch them to appropriate tools, reducing inference cost relative to invoking a large model on every request.
- Designed an LLM-as-a-judge evaluation pipeline that scores agent responses against expected outputs, enabling systematic quality monitoring and iterative improvement.
- Built an end-to-end geospatial data pipeline that converts Dynamical.org datasets into GeoTIFF, generates map tiles, and renders them as interactive React Leaflet maps. Deployed the pipeline on Google Cloud Run with scheduled execution for automated daily refresh.

Software Engineer | TeamOne Technologies

May 2024 – Dec 2024

- Full-stack development for tourism-sector clients, including an interactive mapping application built on a structured JSON schema for routes, accommodations, and trekking data.

TECHNICAL SKILLS

Programming & Tooling: Python, PyTorch, JavaScript/TypeScript, SQL/PostgreSQL, Bash, LaTeX, Git, Docker, Linux, Google Cloud Platform

Modeling: Transformers (encoder-decoder and decoder-only), mT5, Qwen, HuggingFace ecosystem, SentencePiece tokenization, continued pre-training, supervised fine-tuning (SFT), LoRA and parameter-efficient tuning, prompt engineering, zero-/few-shot inference, denoising diffusion probabilistic models (DDPM), contrastive learning, C-LSTM-attention architectures

Evaluation & Experimental Methods: Human evaluation study design, multi-rater annotation protocols, inter-annotator agreement, expert survey instrument design and validation, LLM-as-a-judge, meta-evaluation of metrics, generation metrics (ROUGE, sacreBLEU, chrF++, BERTScore, BARTScore), entity-level precision/recall/F1, benchmark construction and ablation design

Systems & Applied Areas: Retrieval-augmented generation, Model Context Protocol (MCP), agentic LLM pipelines and routing, computer vision and image enhancement, wearable sensor and time-series modeling, geospatial data pipelines (GeoTIFF, tiling, Leaflet)

Software Engineering: React, Next.js, Node.js, REST APIs, webhooks, authentication systems, payment integration, ChromaDB, scikit-learn, NumPy, Pandas, Matplotlib, Jupyter

LANGUAGES

Nepali (native) · English (fluent; medium of instruction throughout undergraduate and graduate study) · Japanese (elementary, JLPT N4 in progress)