

Nabaraj Subedi

📧 subedinab | 💻 nabaraj-subedi | 🌐 subedinab.github.io | ✉ nsubedi1@uwyo.edu | 📞 +1 (307) 357-6341

Research Interests: Multimodal RAG · LLMs & Agentic AI · Vision-Language Models · Multimodal Learning

EDUCATION

University of Wyoming

M.S. in Computer Science · Graduate Research Assistant · GPA [4.00/4.00]

Laramie, WY

Aug 2025 – May 2027 (expected)

- Advisors: Dr. Shivanand V. Sheshappanavar (EECS) and Dr. Ahmed Abdelaty

Tribhuvan University, Institute of Engineering

B.E. in Electronics, Communication & Information Engineering

Nepal

2019 – Mar 2024

PUBLICATIONS

- [1] **N. Subedi**, Y. Majeed, A. Abdelaty, S. V. Sheshappanavar. ConstructGPT: BIM-Conditioned 3D Diffusion and World Models for Automated Construction Site Monitoring and Compliance Checking. *CVPR 2026 Workshop on 3D-LLM/VLA* (poster).
- [2] **N. Subedi**, D. Suvo, A. Abdelaty, S. V. Sheshappanavar. PlanSightRAG: A Visual-First Multimodal RAG for Automating Question Answering and Compliance Checking for Civil Standard Plans. *Automation in Construction*, Under Second revision.
- [3] S. Regmi, **N. Subedi**, A. Pillai, Y. Y. Wu, Y. Chen, M. V. Heinz, L. Marsch, N. C. Jacobson, A. Campbell. MontageFlow: Physically Valid High-Fidelity EEG Generation. *ICLR 2027* (under review).
- [4] **N. Subedi**, A. Abdelaty, S. V. Sheshappanavar. When More Documents Hurt RAG: Mitigating Vector Search Dilution with Domain-Scoped, Model-Agnostic Retrieval. Under review, *ACL Rolling Review*, 2026.
- [5] **N. Subedi**, A. Abdelaty, S. V. Sheshappanavar. CivilAny: Multimodal Model for Civil Infrastructure Inspection and Design. *3DV 2027* (under review).

RESEARCH EXPERIENCE

Graduate Research Assistant, Geometric Intelligence Research Lab, University of Wyoming

Aug 2025 – Present

- Developed **PlanSightRAG**, **MASDR-RAG**, **Civil-NEXT**, **3D CoT Dataset**, and **ConstructGPT**, spanning multimodal RAG, 3D vision-language models, world models, multimodal generation, and agentic AI for engineering applications.

Summer Research Intern, University of Wyoming

Summer 2026

- Developed a modular GPU-accelerated computer vision pipeline for real-time traffic and parking analytics, integrating object detection, tracking, parking occupancy, and vehicle speed estimation.

Research Collaborator, Dartmouth College

Summer 2026

- Contributed to **MontageFlow**, a flow-matching model for physically valid EEG generation, including EEG preprocessing, baseline and ablation experiments, conditioning studies, physiological analysis, and manuscript writing.

SELECTED PROJECTS

- **MRAG 4.0: Multi-Agent Cognitive Digital Twin** (2025–present) – Integrated video, CAD, LiDAR, and operational data in a multi-agent pipeline for visual inspection, CAD reasoning, and compliance checking on WYDOT infrastructure.
- **WYDOT Multimodal Knowledge-Graph Chatbot** (2025–present) – Built multi-agent RAG over Neo4j (150K+ nodes, 330K+ relationships) indexing 1,656+ engineering documents and deployed on Google Cloud.
- **Nepali Image Captioning** (2023–2024) – Developed Transformer + Inception V3 captioning on 20,350 image-text pairs (BLEU-4: 0.59).

EXPERIENCE

• **Junior Machine Learning Engineer**, PalmMind Technology, Nepal

(Apr – Aug 2024)

Built production RAG chatbots for EV and insurance clients using OpenAI, LangChain/LangGraph, OCR, and Redis caching; improved latency and multi-hop reasoning.

• **Teaching Assistant**, Paschimanchal Campus, IOE, Tribhuvan University

(Sep 2024 – Jul 2025)

Taught C Programming, OOP, Operating Systems, and Data Mining; designed labs, guided semester projects, and supported 100+ students.

HONORS & TALKS

• **Graduate Scholar Award** (\$5,000), School of Computing, University of Wyoming, 2026 – 2027

• Lightning talk, UW AI Summit, Sept 2026: *Teaching Models to Work in 3D: Perceiving, Predicting, and Generating 3D Infrastructure Geometry*

TECHNICAL SKILLS

ML / LLM: PyTorch, HF Transformers, DeepSpeed ZeRO-3, PEFT/LoRA, DiT, Qwen2.5-VL, InternVL, LLaVA, LangChain, LangGraph, ColPali

Vision & languages: Llava-Next, World Models, point clouds; Python, C/C++, CUDA, SQL

Infrastructure: NVIDIA H100/A100, SLURM/HPC (ARCC), Docker, GCP Cloud Run, Neo4j, PostgreSQL, BigQuery