

# Saumya Bhandari

✉ [saumyabhandary22@gmail.com](mailto:saumyabhandary22@gmail.com) | 📞 +977-9843866411 | 📍 Mulpani, Kathmandu, Nepal  
🌐 [saumyabhandari.com.np](http://saumyabhandari.com.np) | [in saumya-bhandari](https://www.linkedin.com/company/saumya-bhandari) | [SaumyaBhandari](https://www.github.com/SaumyaBhandari)

## EXPERIENCE

### • Co-Founder & COO

Jul 2025 – Present

*PointRef Inc. (DE C-Corp) – AI Sports-Tech*

*Apprenticeship:* [Mosalam Ebrahimi](#) (Meta Reality Labs)

- Shipping world's first real-time AI pickleball officiant iPhone/macOS → Cloud; 99% foul-call accuracy at **150 ms** E2E latency.
  - Scaled multi-object tracking fused with IMU; homography-based 2D/3D court registration; **3x throughput** via TensorRT-INT8 + selective frame-skip + frame buffering.
  - Working with global cross-functional pod (AI+EE+FW); owned roadmap, sprint cadence, OKRs, fundraising; 10% founder equity
- Endorsements:* [Mike White](#), [Deno Dean](#), [Gary Brown](#)

### • Co-Founder & Leading AI

Oct 2024 – Present

*The Legalities (InstaComply) – Legal-Tech / Compliance*

- Architected **SelfCompl.ai** – a modular multi-agent RAG system powering PDPL, ISO-27001, and ISO-9001 audits; reduced manual audit cycles from **6 months → 20 days** (80% effort cut), auto-generating compliance templates.
- Designed complete AI blueprint: RAG pipeline, vector DB strategy, HITL feedback, XAI audit-trails, and cost-optimized multi-LLM orchestration (Claude, GPT-4o, Gemini, Mixtral).
- Authored 5+ research-backed articles on AI and compliance published at [selfcomplai.com/blogs.html](https://selfcomplai.com/blogs.html)
- Stack: Python, FastAPI, Django, PyTorch, LangChain, Pinecone, AWS Textract, RDS; led 6-person AI pod, defined OKRs, hiring, and infra strategy.

*Endorsement:* [Emmet Kelly](#), [Sushrant Rijal](#)

### • Machine Learning Engineer

Jan 2024 – Oct 2024

*Naamche (ReAlpha) – PropTech AI*

- Virtual Staging Platform: Fine-tuned SegFormer for room segmentation; diffusion-based inpainting pipeline increased buyer-engagement metric **86%**; served on SageMaker serverless multi-GPU endpoint.
- Released "Claire" – RAG real-estate advisor (GPT-4 + Pinecone) answering 50k+ buyer Q's; cut support tickets 35%.

### • Machine Learning Engineer

Sep 2022 – Oct 2023

*Wiseyak Inc. – Medical AI*

- Delivered chest X-ray multi-label classifier (10 diseases) AUROC **0.91** on highly-imbalanced data; Grad-CAM explainability integrated in radiologist cockpit.
- Designed CNN-Transformer hybrid for medical video denoising + next-frame prediction; custom Convolutional Transformer with Transformer encoder for temporal understanding.
- Domain-adversarial network (DANN) for Nepali plant disease detection aligned with Meta's SAM, EfficientNet.

*Endorsements:* [Prof. Dr. Suresh Manandhar](#), [Suraj Prasai](#)

### • Lecturer

Nov 2024 - Present

*Malpi International College*

*Pokhara University: CSIT (Semesters I, II, and V)*

- **Subjects:** Programming Language: C, Object Oriented Programming: Java, and Artificial Intelligence and Machine Learning

## EDUCATION

### B.Sc. (Hons) Computer Science (First-Class)

2019 – 2022

*Herald College Kathmandu, Naxal, Kathmandu*

- \* Thesis: *Pathfinder: An Image Segmentation-Based Navigation System for Self-Driving Systems*
- \* Student Academic Representative (StAR) **2019–2021** (2 Years)
- \* Hult Prize 2021 On-Campus Winner
- \* Actively lead 3 student run communities

### Higher Secondary (Mathematics & Computer Science)

2017 – 2019

*Uniglobe College*

- \* IT Club President
- \* National Debate (Mahasangram) Finalist 2018

- \* High School Hackathon 2018 Winner (Android Application Development)
- \* High School Hackathon 2019 Organizer
- \* Honored in **16+ events** related to public speaking, debate, presentations, quiz, and design thinking over 2 years

## TECHNICAL SKILLS

---

- **Deep Learning:** PyTorch, TensorRT, ONNX, LoRA, Diffusion, YOLO, SAM, SegFormer
- **LLM / NLP:** RAG, LangChain, Claude-3, GPT-4, Llama-2, Pinecone, Chroma
- **Vision:** Multi-object tracking, 3D homography, CUDA, INT8 quant, OpenCV
- **Back-end:** FastAPI, Django, gRPC, PostgreSQL, Redis, Celery
- **Cloud / MLOps:** AWS (SageMaker, EC2, S3, Textract), Docker, CI/CD, OTA
- **Edge:** Jetson Orin / Nano, DeepStream, GStreamer, selective inference

## ORGANISATIONAL SOFT SKILLS

---

- Excellent Public Speaking and Presentation Skills
- Strong Communication, Brainstorming, and Critical Thinking abilities
- Skilled in Planning, Decision Making, and leading collaborative teams

## SELECTED SELF-LEARNING

---

- **Stanford:** CS231N (CNN), CS224N (NLP) | **MIT:** 6.034 (AI), 18.06 (Linear Algebra), Deep Learning Bootcamp, 9.13 (Human Brain), 8.04 (Quantum I) | **Harvard:** Stat110 | **UW:** CS480/680 (Deep Learning)