

Sourav Sharan

Uppsala, Sweden – Email: souravthesharan@gmail.com

☎ +46 766065362 • 🌐 souravsharan.com • 🐙 [SouravSharan](https://github.com/SouravSharan) • [in](https://www.linkedin.com/in/souravsharan) [souravsharan](https://www.linkedin.com/in/souravsharan)

Research Interests

Continual Learning; Causal Reasoning; Alternative Learning Mechanisms. Exploring how systems can learn robust, interpretable, and transferable representations of the world.

Education

Uppsala University

Uppsala, Sweden

M.Sc. in Data Science (Machine Learning and Statistics)

2024–2026

- **Relevant Coursework:** Deep Learning, Statistical Machine Learning, Optimization, Data Engineering, Large Language Models and Societal Consequences of Artificial Intelligence, Functional Programming I
- **Teaching Assistant:** Programming (1DL042), Computer Programming I (1TD433)

SRM Institute of Science and Technology

Chennai, India

B.Tech. in Computer Science

2016–2020

- **Undergraduate Thesis:** “Fast Enhanced Underwater Object Detection on Embedded Systems” — Implemented real-time object detection for marine robotics on resource-constrained hardware
- **Head, AI at Next Tech Lab:** Led a multi-disciplinary research lab of 150 students focused on emerging technologies; collaborated with advisers from Google Brain, MIT, and CMU
- **Community Leadership:** Treasurer for ACM Student Chapter (organized Cybersecurity Winter Camp); Organizer for PyData Meetup (convened meetings for ML enthusiasts).

Professional Experience

Billion Hearts Software Technologies

Remote

Machine Learning Consultant

Jun 2025 – Aug 2025

Designed and implemented computer vision and deep learning solutions for Picsee, a photo-sharing application. Optimized models for production deployment, focusing on inference efficiency and user experience.

HyperVerge Inc.

Bangalore, India

Machine Learning Engineer II

Dec 2019 – Apr 2024

- Promoted twice based on performance, advancing from a campus-recruited Intern (Dec 2019) to ML Engineer I (Aug 2020) and ML Engineer II (Jan 2022).
- Led the R&D of multi-modal Transformer architectures for NER, achieving a **30% accuracy improvement** on identity documents.
- Optimized and deployed core models for a system impacting **280 million customer onboardings** across India, Vietnam, Indonesia, and Malaysia.
- Replaced a legacy OCR pipeline with a state-of-the-art model, increasing **end-to-end accuracy by 10%** while reducing **GPU usage by 80%**.
- Took end-to-end ownership of the information extraction pipeline for KYC automation in the ASEAN region, including document quality models (blur/glare).
- Built and scaled data preprocessing/annotation pipelines for large-scale datasets (0.6 billion data points).
- Led weekly AI team paper reading sessions and mentored new hires.

Crea.Vision

Remote (Part-time)

Computer Vision Engineer (Freelance)

Aug 2018 – Mar 2019

Deployed end-to-end employee monitoring system in workspaces. Developed deep learning models for face recognition, object detection, and tracking.

Research and Development Projects

Academic Research:

- **A Practical Investigation into Mixture of Experts (MoE)** (2025): Graduate project supervised by Prof. S. Toor & Prof. P. Singh. Conducted systematic empirical study of MoE architectures on CIFAR-10.
- **Do Vision Transformers Encode Data Augmentations?** (2022): Investigated if augmentation strategies are encoded in ViT representations. Mentored by Dr. Been Kim (Google Brain); supported by ICLR 2022 CoSubmitting Initiative grant.

Personal Projects & Infrastructure:

- **Personal Research & MLOps Compute Cluster** (2021–Present): Designed and maintain a multi-node Proxmox cluster with heterogeneous GPUs and NAS for DL experiments, hybrid-cloud integration, and bare-metal (Pi-KVM) management.
- **Konf.dev** (2025): Graph-based execution engine for multi-agent workflows (DAGs) with multi-tier cognitive memory. Implements ReAct, CoT, etc.
- **Custom Transformer Implementation** (2024): Built Transformer from scratch to experiment with attention mechanisms, layer configurations, and positional encodings.

Coursework & Competitions:

- **Distributed Text Summarization Pipeline** (2024): Implemented distributed pipeline for summarizing 3.8M Reddit posts using Hadoop and Spark.
- **VQA Hateful Meme Detection** (2021): Implemented transformer-based multi-modal models for Facebook's Hateful Meme Challenge 2021.

Programming Skills

Languages: Python, C/C++, Bash, Haskell

ML Frameworks: PyTorch, TensorFlow, Hugging Face, Scikit-learn, PyTorch Lightning, LangGraph

Data & MLOps: Apache Spark, Hadoop, Proxmox, Docker, Kubernetes, PostgreSQL, Weights & Biases, TensorRT, ONNX, ChromaDB

Cloud & Edge: AWS, Google Cloud, Raspberry Pi

Core Tools: NumPy, Pandas, Git, Streamlit, LaTeX

Awards and Recognition

2021: ICLR Co-Submitting Summer Initiative: Awarded \$250 grant for research proposal on Vision Transformer augmentation encoding, accepted under ICLR's competitive CoSubmitting Summer program.

2019: Best Submission, Vivli Microsoft Data Challenge: Built privacy-preserving solution for clinical trials.

2018: ACM Student Grant: Received INR 8,000 funding for ACM India's largest data science conference.

2017: Innovation Award, Smart India Hackathon: Team won INR 30,000 and Innovation Award for autonomous video surveillance system at India's largest hackathon.

References (Available Upon Request)

- **Dr. Salman Toor**, Professor, Department of Information Technology, Uppsala University
- **Hariprasad P S**, Head of AI, HyperVerge Inc.
- **Dr. Subalalitha C. N.**, Professor, Department of Computing Technologies, SRM IST