

Pragyananda Sahoo

Junior Research Fellow @ *Swami Rama Himalayan University*

Haridwar, Uttarakhand, India

+91 6398665353 | pragyanandasaho@gmail.com | [pragyananda](#) | [pragyanandasaho](#)
[pragyananda.github.io](#) | [Google Scholar](#)

PROFESSIONAL SUMMARY

Junior Research Fellow and former AI Engineer with two years of applied research at CAIR Lab, Institute of AI, DSVV, specialising in generative AI, retrieval-augmented generation (RAG), large language models, and computer vision. Built an LLM-based legal case matching system for the Supreme Court of India and contributed to Innopat, an AI patent-drafting tool. Co-author of three peer-reviewed papers at COLING 2025, SemEval-2026 (ACL), and MBCC 2025, including a SemEval system ranked 3rd of 47; current research on child-safe advertising is under review at the NLLP Workshop. Full stack background with deployments on AWS, Azure, and Vercel.

SKILLS & EXPERTISE

GenAI & LLMs	RAG, LLM fine-tuning, prompt engineering, vector search, LangChain, Hugging Face Transformers, OpenAI API
Expertise	NLP, Computer Vision, OCR, Deep Learning, Generative Modelling, Meta Learning, Machine-generated text detection
Libraries	PyTorch, TensorFlow, OpenCV, docTR, NumPy, Pandas
Web	Next.js, React, Node.js, Express, Django, Flask, FastAPI, Tailwind CSS, Bootstrap
Data & Cloud	MySQL, PostgreSQL, MongoDB, SQLite, Supabase, Firebase, Redis, AWS, Azure, Docker, Vercel, Nginx
Languages & Tools	Python, JavaScript (ES6+), TypeScript, Java, C++, SQL; Git, GitHub Actions, Linux (Ubuntu), Anaconda, VS Code, Jira

PROFESSIONAL EXPERIENCE

Junior Research Fellow | *Swami Rama Himalayan University, Doiwala* Mar 2026 – Present

- **HistoAI:** Continuing development of the historical-books ecosystem started at CAIR Lab, DSVV: OCR with image extraction, structured data extraction, knowledge-graph generation, and a RAG chatbot over the digitised collection.
- **Child-safe advertising research:** Conducted an NLP study on child-safe advertisements; paper submitted to the Natural Legal Language Processing (NLLP) Workshop (under review).
- **Prototypes & research support:** Built web application prototypes and contributed to other research projects at the university.

AI Engineer | *CAIR Lab, Institute of AI, DSVV, Haridwar* May 2024 – Mar 2026

- **Document-automation systems:** Built LLM and RAG products for legal, patent, and historical-document workflows: Supreme Court legal case matching, Innopat patent drafting, and HistoAI (detailed in Key Projects).
- **Applied Research:** Generative AI for images, video, and text; transformer architectures and LLM fine-tuning; trained and deployed models for real-world use and mentored junior developers.

Project Intern (Computer Vision) | *CAIR Lab, Institute of AI, DSVV, Haridwar* Feb 2024 – May 2024

- **Handwritten OCR:** Researched handwritten document transformation with Document AI transformers and docTR; shipped a MERN + Flask web app for handwritten-text recognition.
- **Vision & LLM tooling:** Built a face-recognition attendance system (Tkinter, OpenCV, Haar cascades); integrated OpenAI APIs with LangChain and Azure services.

Internship Trainee (Web Development) | *Travel Vaidya, Haridwar* Jun 2023 – Jul 2023

- Migrated uttarakhandtrips.com from WordPress to a scalable Django platform with payment integration, MySQL, and a REST API, improving performance and reliability for real customers.

Summer Intern (Web Development) | *Dev Sanskriti Vishwavidyalaya, Haridwar* Apr 2022 – May 2022

- Built internal web applications supporting university operations with PHP and MySQL, including e-commerce features.

EDUCATION

Sep 2022 – Jul 2024	Master of Computer Applications (Data Science) , DSVV , Haridwar, India CGPA: 8.20/10
Aug 2019 – Jul 2022	Bachelor of Computer Applications (Computer Science) , DSVV , Haridwar, India CGPA: 7.42/10

KEY PROJECTS

Generative AI & Document Automation

- **HistoAI: Ecosystem for Historical Books** [OCR, Knowledge Graphs, RAG]
 - End-to-end pipeline for digitising historical books, developed at CAIR Lab, DSVV and continued at SRHU: OCR with image extraction for degraded scans, structured data extraction, knowledge-graph generation, and a RAG chatbot for context-aware queries over the collection.
- **Innopat: AI Patent Drafting Assistant** [RAG, LLMs, LangChain]
 - Turns claim lists and Invention Disclosure Forms (IDF) into full patent specifications through a modular generation workflow, cutting the labour-intensive parts of patent preparation.
 - Iteratively refined prompts for consistent technical terminology, improved parsing of PDF and DOCX inputs, and better handling of figure mappings and definitions for professionally structured drafts.
- **Supreme Court Legal Case Matching System** [RAG, LLMs, vector search]
 - Similarity and matching model for legal workflows: vector-based RAG with LLMs automatically identifies core commonalities, semantic similarities, and legal precedents between newly filed and historical cases.

Financial Automation

- **AURA: Tax Automation for Ride-share Drivers (Australia)** [Data pipelines, reconciliation]
 - Accounting automation system that streamlines taxation and financial data processing for UBER and DIDI drivers: extraction modules ingest ride-sharing activity statements alongside automated pipelines for major Australian banks (CBA, NAB, ANZ).
 - Phase 1 delivered a scalable architecture with a Business Activity Statement (BAS) automation engine for tax compliance, transaction reconciliation algorithms, and a dynamic UI with Excel-integrated reporting for Australian financial-year requirements.

PUBLICATIONS

- **Pragyananda Sahoo** et al. Paper on child-safe advertising (title to be confirmed). *Natural Legal Language Processing (NLLP) Workshop*, 2026. *Under review*.
- Vishnu Tripathi, Prakhar Joshi, **Pragyananda Sahoo**, Gaurav Kumar, Piyush Arora, Neel Mani, et al. AI-Monitors at SemEval-2026 Task 4: A Hybrid Embedding and LLM Ensemble for Narrative Similarity. *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, ACL, 2026, pp. 2139–2148.
- **Pragyanand Sahoo**, Neel Mani, Shastri Nimmagadda, and Amit Das. Marma Point Detection and Personalized Ayurvedic Diagnosis Using Computer Vision. *Indian Knowledge System and Wellbeing: Proceedings of MBCC 2025*, Vol. 1, Springer Nature, 2026, p. 228.
- Azad Singh, Vishnu Tripathi, Ravindra Kumar Pandey, **Pragyanand Saho**, Prakhar Joshi, Neel Mani, Richa Alagh, Pallaw Mishra, and Piyush Arora. [AI-Monitors at GenAI Detection Task 1: Fast and Scalable Machine Generated Text Detection](#). *GenAIDetect Workshop, COLING*, 2025, pp. 230–235.

ACHIEVEMENTS

- **COLING 2025 Competition on Binary Text Classification:** Competed with the AI-Monitors team in the GenAI Content Detection shared task; results published at the GenAIDetect workshop.
- **2nd MBCC 2025 Conference, IIT Mandi:** Presented our Marma Point Detection paper, later published in the conference proceedings.
- **ICFAI University Ideathon, Dehradun:** Pitched an idea and presented its impact and feasibility to the judges, receiving positive feedback.
- **Tula's 24-Hour Hackathon:** Participated in the 24-hour hackathon at Tula's Institute, Dehradun.

CERTIFICATIONS & ACTIVITIES

- **Certifications:** AWS Builders Online Series (Amazon Web Services); Deploy a Django Web App with AWS Elastic Beanstalk (2023); Data Analytics with Specialization in Tableau; Intro to Programming (Kaggle).
- **Activities:** National Art Competition participant; social internship in community development; badminton player.