# Sarvagya Porwal

## SUMMARY

AI/ML Engineer and Researcher with expertise in deep learning, computer vision, generative AI, and NLP. Experienced in designing and implementing diffusion models, large language models, and retrieval-augmented generation systems. Skilled in building scalable AI pipelines using Python, PyTorch, and TensorFlow, with hands-on experience in cloud deployment and distributed training. Published research on diffusion models and spectral attention networks. Passionate about bridging research and production by translating cutting-edge AI methods into efficient, real-world solutions.

## PUBLICATIONS

**Smoothed Energy Guidance - Reproducibility Challenge** **Jan 2025 - Feb 2025**
*Diffusion Models — Code PyTorch, Hugging Face*

- Reproduced and enhanced SEG (Smoothed Energy Guidance) from NeurIPS 2024, addressing missing ablation studies on kernel size and blurring strategies.
- Optimized smoothing in diffusion models, replacing redundant operations with EMA and BoxBlur, improving efficiency without loss in quality.
- Tracked reverse diffusion trajectory using Frobenius Norm, Laplacian Variance, and Gradient Entropy on attention layers, ensuring better interpretability.
- The research paper is currently under review.

**Spectral Band Attention Network** **Sep 2024 - Dec 2024**
*Computer Vision — Code PyTorch, OpenCV*

- Classification of wheat seeds into 96 varieties, fine-tuned models like DenseNet-121, ResNet-50, and GoogleNet for RGB seed image classification (max 78% accuracy) and developed a DenseNet-121-inspired architecture for Hyperspectral data.
- Applied Spectral Band Attention Module (SBAM) for band selection, ranking bands by their contribution to the classifier's accuracy, achieving 87% accuracy on Hyperspectral images.
- Employed Regression-based Ensemble (using SVM) to combine model predictions, ensuring robustness and enhancing overall accuracy to 95%.
- The research paper has been accepted (view here).

## PROJECTS

**NeurIPS 2025 Pokemon AI Competition** **September 2025 – November 2025**
*Game AI — Leaderboard Imitation Learning, Offline RL*

- Trained an **Actor-Critic policy** using **Offline Reinforcement Learning** on game replay datasets, enhancing standard **Behavior Cloning** with exploration mechanisms.
- Engineered an **policy ensemble** to mitigate high action-entropy, dynamically selecting confident agents based on **TD error** to handle less explored scenarios during behaviour cloning.
- Secured **Global Rank 3** in Track-1 with a policy **ELO rating of 1793**, demonstrating superior generalization in unseen competitive scenarios.

**Enriched Bots-Clever Chat** **June 2024 – July 2024**
*Generative AI — Demo Python, Django, LlamaIndex, Hugging Face*

- Enhanced text-based interactions with multimedia using a novel graph-based architecture.
- Decomposed queries into acyclic graphs to store text and media metadata within nodes.
- Generated responses via topological traversal to ensure logical sequencing of enriched output.

**AI Agent 007: Tooling up for Success (Inter-IIT Techfest 2023)** **Dec 2023 – Dec 2023**
*Generative AI — Code Python, Langchain, GPT-4, Hugging Face*

- Built a query-aware agent capable of allocating and reviewing tool outputs
- Focused on creating autonomous tools for efficient parameter extraction and downstream function calls
- Implemented a self-reflective ReAct style agent, curated dataset using given tool descriptions

**Sentinel-2 Field Delineation** **July 2024 – August 2024**

*Computer Vision — Code PyTorch, OpenCV, Segmentation Models*

- Engineered a **field delineation** pipeline using high-resolution **hyperspectral imagery** from Solafune.
- Fine-tuned **UNet++, FPN, & Mask-RCNN**; extracted annotation polygons from patches via **OpenCV**.
- Designed an **ensemble model** by stacking prediction masks, achieving a segmentation **IoU of 0.96**.

# ACHIEVEMENTS

- Secured **Global Rank 3** in the **NeurIPS 2025 Pokemon AI Competition** (Track-1), competing as Team srsk-1729. [Leaderboard]
- Achieved **Rank 3** in the **AI Agent 007: Tooling up for Success (Inter-IIT Techfest 2023)** under the DevRev AI Agent Track, competing against top Indian technical institutes.

# EXPERIENCE

**Avathon — SDE-1** **July 2025 – Present**

→ *avathon.com*

- Architected and implemented distributed task processing infrastructure using Celery with message queuing, enabling asynchronous job execution, automatic retries with exponential backoff, and real-time task monitoring across microservices, significantly improving system throughput and reducing blocking operations.
- Contributed to the development of **Neon**, a proprietary *scala-based graph database* designed for supply chain solutions, implementing semantic search capabilities on customer data distributed across graph nodes and enabling efficient querying of complex relational data structures.
- Developed and maintained Django REST APIs and comprehensive test suites for the **Central Asset Manager** service, ensuring robust functionality, data validation, and compliance with system reliability standards.

**DeepLogic AI — AI Engineer Intern** **July 2024 – Dec 2024**

→ *Certificate*

- Contributed to the development of a Retrieval-Augmented Generation (RAG) pipeline for enterprise search, managing email and document embeddings in a Postgres vector-store on AWS, enabling high-performance information retrieval.
- Designed normalized database schemas and optimized scalable CRUD operations for metadata-filtered searches across millions of documents.
- Developed critical components such as the Retriever, Response Generator, and Re-ranker, and implemented caching strategies to enhance chatbot integration, improving overall system interaction, efficiency, and scalability.

# EDUCATION

**Indian Institute of Technology, Roorkee** **Roorkee, India**

*BTech in Mechanical Engineering* | ***CGPA: 7.1*** *Nov 2021 – July 2025*

# TECHNICAL SKILLS

**AI & ML**: PyTorch, TensorFlow, LangChain, LlamaIndex, Hugging Face, RAG, PEFT/LoRA, OpenCV
**Dev & Ops**: Docker, AWS, Linux, FastAPI, Django, LLMOps, Git, AsyncIO

# CERTIFICATIONS

- Machine Learning Certificate
- AWS Cloud Practitioner Certificate
- Scala Programming Certificate