

Daksh Agarwal

dakshagrawal76@gmail.com | [Linkedin](#) | [Github](#) | [+91 8433901840](#)

EXPERIENCE

Machine Learning Intern | Peakflo (YC W22)

August 2026 – Present

- Fine-tuned and optimized LLM-powered AI agents using GPT-4o, Gemini, and Claude through prompt engineering, reducing hallucinations, improving response consistency, and integrating external tools for enterprise finance workflows.

Machine Learning Intern | Avawatz

August 2025 – Present

- Built predictive failure detection models for Flint (USA) bus operational systems using real-world sensor, travel, and operational data; generated 100,000+ synthetic records with distribution-preserving techniques to address severe class imbalance in rare failure events
- Benchmarked Graph Neural Networks against Linear Regression, Logistic Regression, Random Forest, ANN, RNN, and GraphSAGE, demonstrating superior modeling of temporal inter-sensor dependencies; conducted feature ablation studies and Grad-CAM visualizations to improve model interpretability and support responsible AI
- Curated and annotated a domain-specific dental imaging dataset of 800+ radiographs and fine-tuned YOLOv8 for incisor detection, achieving 91% mAP50 on the test set

Research Associate | New York University

July 2025 – Present

- Co-leading a research group benchmarking a custom RAG system against Open Deep Research and AutoGPT, generating structured comparative metrics across retrieval quality, answer grounding, and hallucination rate
- Orchestrated multi-LLM evaluation pipelines using GPT-4o, Gemini 3.0 Flash, Llama 3.3 70B, and Claude 3.5 across 5+ retrieval configurations, identifying optimal model-retriever pairings for research-grade tasks
- Improved answer accuracy by 9% through prompt and retrieval optimization, while extending the pipeline with citation attribution, session memory, and reproducible version-controlled benchmarking

Machine Learning Intern | Grant Thornton (Walker Chandiok)

June 2026 – August 2026

- Enhanced and debugged **GenMobius**, an enterprise AI MVP, and developed intent-aware AI agents integrating Azure AI Search for document retrieval and external API services
- Built an AI-powered proposal generation system and analyzed enterprise agent usage over 1 month, identifying active deployments and business use cases for partners and senior leadership
- Developed and benchmarked ML/DL models on the NASA C-MAPSS dataset for predictive maintenance, comparing model performance for remaining useful life (RUL) prediction

Data Science Intern | HCL Guvi

Feb 2026 – April 2026

- Designed and delivered 10+ cloud and AI curriculum modules covering generative AI, LLMs, prompt engineering, and MLOps, enabling knowledge transfer and upskilling across 10+ learners
- Co-developed an internal MVP product for AI-generated avatars and synthetic voice (a HeyGen alternative), integrating generative AI APIs and audio synthesis pipelines into a deployable web application — demonstrating hands-on experience building production-grade generative AI systems

Data Science Intern | Australian Institute of Advanced Technologies

July 2025 – October 2025

- Mentored students across AI domains, delivering structured lectures and designing project briefs that improved learner outcomes
- Designed and reviewed project statements covering supervised learning, NLP, and computer vision workflows, sharing best practices and fostering a culture of technical excellence

Undergraduate Researcher | IIT Tirupati

May 2025 – August 2025

- Built a real-time underwater object detection system using YOLOv8, constructing a custom dataset by combining proprietary camera footage with open-source data, covering the full ML lifecycle from data acquisition to deployment
- Validated and stress-tested the model in live underwater environments on physical hardware, demonstrating experience deploying models beyond controlled lab settings
- Designed and optimised the data pipeline for ingestion, annotation, and augmentation, improving model robustness across diverse real-world conditions

Undergraduate Researcher | IIIT Delhi

April 2025 – August 2025

- Engineered an end-to-end deep learning pipeline for malware classification, extracting opcode and API call sequences from 20+ malware families and modelling them using transformer-based NLP techniques, achieving 96.4% classification accuracy
- Applied SHAP-based feature importance analysis to interpret model decisions, aligning with responsible AI and model auditability standards
- Decompiled and disassembled 10,000+ window malware binaries using static analysis tools (IDA and Ghidra) to build a labelled training corpus from scratch

SKILLS

Languages: Python, C, C++, SQL, JavaScript, HTML, CSS, R, Bash
Cloud Platforms: Microsoft Azure (Azure ML, Azure AI Foundry, Azure OpenAI Service, Azure VMs), GCP (Vertex AI), Oracle Cloud Infrastructure (OCI)
Machine Learning: TensorFlow, Scikit-learn, Keras, Model Evaluation, Hyperparameter Tuning, Feature Engineering, XGBoost, LightGBM, CatBoost, FastAI, PyTorch
LLM / NLP: Transformers, LangChain, LangGraph, LlamaIndex, RAG, Prompt Engineering, Embedding Models, NLTK, LLM Evaluation
Computer Vision: OpenCV, YOLO, Object Detection, Image Segmentation
Vector Databases: FAISS, Chroma, Pinecone
MLOps & Deployment: Docker, Kubernetes, MLflow, Flask, FastAPI
Soft Skills: Teamwork, Adaptability, Collaboration, Flexibility, Leadership, Communication

KEY PROJECTS

- Prompt Sensitivity Analysis in LLM Code Generation**
- Orchestrated approximately 6,000 automated evaluation jobs across 6 LLMs (GPT-4o, Claude, Gemini, Qwen-Coder, LLaMA, Phi) using Azure AI Foundry and Azure OpenAI Service, building a reproducible experimentation framework with version-controlled job configs and structured result logging.
 - Designed a prompt perturbation taxonomy across three axes (structural, linguistic, stylistic) and introduced a custom metric suite including PromptVar, WorstPrompt, BestPrompt, and SensRange to quantify model brittleness beyond standard benchmarks like pass@k, surfacing robustness gaps invisible to conventional evaluation
 - Applied Spearman correlation analysis across model families to distinguish whether prompt sensitivity is driven by problem difficulty or model architecture, and validated findings with bootstrap confidence intervals and paired t-tests to ensure statistical rigour
- Domain-Specific AI Chatbot**
- Architected a domain-constrained RAG chatbot using LangChain and a vector database for grounded retrieval, implementing query filtering logic to ensure responses remained within scope
 - Implemented persistent session memory to maintain multi-turn conversation context, improving answer consistency across interactions; applied prompt engineering techniques to enforce domain boundaries and reduce off-topic generation, reflecting bias-aware and governable AI design principles

EDUCATION

Guru Gobind Indraprastha University	Delhi,India
<i>Bachelor of Technology(B.Tech) in Artificial Intelligence and Machine Learning (CGPA: 8.0)</i>	<i>Aug. 2023 – May 2027</i>
Indian Institute Of Technology - Madras	Madras,India
<i>Bachelor Of Science(B.S) in Artificial Intelligence and Data Science (CGPA: 8.0)</i>	<i>June 2023 – June 2027</i>

PUBLICATIONS & LEADERSHIP

Publication in ASIACCS (A Ranking)	2026
Accepted publication in MDPI Electronics (Q2 Ranking)	2026
Certifications : Link	
ML Lead Google Development Group	Oct. 2024 – Present