

Bhaswata Kaushik

Patiala, India | bkaushik0602@gmail.com | +91-9101755315

GitHub | LinkedIn | Personal Website

EDUCATION

Thapar Institute of Engineering & Technology (TIET), Patiala, Punjab

Sep 2022 – May 2026

B.E. in Electronics and Communication Engineering

CGPA: 7.49/10

Relevant Coursework: Data Structures and Algorithms, Artificial Intelligence, Accelerated

Data Science, Digital Signal Processing

TECHNICAL SKILLS

Languages: Python, C/C++, SQL

Libraries & Frameworks: NumPy, Pandas, Scikit-learn, TensorFlow, PyTorch, Hugging Face Transformers, OpenCV

Cloud & MLOps Platforms: AWS (SageMaker, Lambda, EC2), GCP (Vertex AI, Colab), MLflow, Apache Airflow, Docker, Kubernetes, CI/CD (GitHub Actions, Jenkins), FastAPI

Tools & Environments: Git, Bash, Linux

PROFESSIONAL EXPERIENCE

Machine Learning Researcher (Intern)

Sep 2024 – Jun 2025

Thapar Institute of Engineering & Technology

- Conducted applied machine learning research under the guidance of Dr. Mohit Agarwal, focusing on classification and regression tasks in healthcare and renewable energy domains.
- Co-authored and published a research paper on enhancing solar cell efficiency using ML, achieving an R^2 score of 0.9998 and MSE of 0.036 ± 0.028 with XGBoost and ensemble techniques.
- Developed and validated breast cancer classification models, attaining 98.07% accuracy through stratified k-fold cross-validation and rigorous hyperparameter tuning.

PUBLICATIONS (GOOGLE SCHOLAR)

- N. Singh, B. Kaushik, M. Agarwal, *Performance enhancement of CHTS-based solar cells using machine learning optimization techniques*, **Journal of Physics and Chemistry of Solids**, Volume 201, 2025, Article 112642. ISSN: 0022-3697. <https://doi.org/10.1016/j.jpcs.2025.112642>

PROJECTS

End-to-End YouTube Semantic Search Engine with Embedding-Based Retrieval

GitHub

- Built a **production-ready semantic search engine** for YouTube videos using Sentence-BERT embeddings and neural information retrieval, enabling intelligent content discovery beyond keyword matching.
- Developed scalable MLOps pipeline with **FastAPI microservices**, Docker containerization, **AWS ECS deployment** with auto-scaling, and **automated CI/CD pipeline** using GitHub Actions for ETL orchestration and real-time transcript processing.
- Achieved **sub-second query response times** (100–300ms average) using optimized Manhattan distance similarity computation and 384-dimensional vector embeddings on Parquet-based storage.

Fine-Tuning an LLM with Direct Preference Optimization (DPO)

GitHub

- Implemented **Direct Preference Optimization** for fine-tuning Qwen2.5-0.5B model on YouTube title generation, eliminating complex RLHF pipelines through preference-based learning.
- Achieved **72.5% reward accuracy** (improved from 65%) and 33% increase in reward margins using automated preference data generation with Together AI and Google Gemini APIs.
- **Published model and dataset** on Hugging Face Hub with comprehensive evaluation framework, demonstrating 18.3% reduction in training loss and production-ready deployment capabilities.

CERTIFICATIONS

- **Fundamentals of Accelerated Computing with CUDA Python** NVIDIA Deep Learning Institute
- **Scaling Workloads Across Multiple GPUs with CUDA C++** NVIDIA Deep Learning Institute
- **Getting Started with Deep Learning** NVIDIA Deep Learning Institute