

ARNAV GUPTA

Bandipur, Nepal

+977 9863744666

✉ arnavgupta6969@icloud.com

🌐 [Linkedin](#)

🐙 [GitHub](#)

Experience

BizLink

October 2025 – Present

AI/ML Engineer (NLP & Research Focus)

Remote

- Designed and deployed NLP systems matching SMBs with professional consultants using semantic search and embedding-based retrieval.
- Developed LLM-based recommendation pipelines, optimizing ranking accuracy and contextual relevance.
- Integrated AI models into production pipelines with REST APIs and Docker for scalable inference.

Suvidha Foundation

September 2025 – November 2025

Machine Learning Intern

Remote

- Researched Multi-Document Abstractive Summarization with transformers under mentorship of a PhD in NLP.
- Built Proof-of-Concept summarization system with hierarchical context aggregation.
- Participated in community outreach programs supporting underserved populations.

Research

Independent Research – NLP

2025 – Present

Multi-Document Abstractive Summarization

- Designed a modular, training-free pipeline combining entity-guided anchor extraction, anchor-conditioned generation, and dual-mode NLI/entity faithfulness auditing for multi-document summarization.
- Authoring a manuscript on improving multi-document summarization faithfulness through NLI and entity-gated iterative refinement (under review at ACL SRW 2026).
- Achieved a 6.3% relative SummaCConv inconsistency reduction over a 500-instance Multi-News evaluation; outperformed fine-tuned baselines (BART, PEGASUS, PRIMERA) across all LLM-as-judge dimensions.
- Empirically documented verifier exploitation in NLI-guided iterative refinement and proposed a three-condition multi-metric triangulation protocol for its detection.

Projects

Transformer from Scratch | *Python, PyTorch, Deep Learning, NLP*

2025

- Implemented full Transformer architecture from scratch: multi-head self-attention, positional encoding, feed-forward networks, and stacked encoder layers.
- Trained model for next-token prediction, demonstrating sequence modeling and attention mechanisms without high-level libraries.
- Modularized components with step-by-step Jupyter notebook bridging theory and code.

Financial Statement Analyzer AI | *Python, Streamlit, LLMs, Pandas, NumPy*

2024

- Built AI tool to analyze financial statements and generate strategic insights with voice-enabled explanations.
- Calculated 20+ financial ratios and performed trend analysis producing multi-format reports.

Ledger Management System | *Python, Pandas, Streamlit, Firebase*

2025

- Developed management system to track credits, debits, and transaction history with persistent storage.
- Implemented automated calculations and Excel export functionality to streamline financial record-keeping.

Emotion Detection System | *Python, Machine Learning, NLP, PyTorch/Scikit-learn*

2025

- Developed text-based emotion detection model classifying multiple emotional categories.
- Applied preprocessing, tokenization, and feature engineering to improve model accuracy.
- Optimized model performance via hyperparameter tuning and evaluation metrics.

Programs & Fellowships

Harvard Computer Society AI Bootcamp

2025

Artificial Intelligence Fellow

Remote

- Completed intensive AI program collaborating with Harvard students, focusing on frontier AI systems and real-world deployment challenges.
- Studied Transformers in depth and implemented one from scratch; covered embeddings, positional encoding, self-attention, cross-attention, feed-forward networks.
- Explored Reinforcement Learning: MDPs, Value Iteration, Policy Iteration, SARSA, Q-Learning, and Multi-Armed Bandits.
- Analyzed LLM architectures (ChatGPT, Claude, Gemini, Perplexity) and researched hallucination mitigation strategies.

- Completed hands-on projects: fine-tuned BERT, built mini-LLM, implemented Vision Transformers, and built neural networks from first principles.

Mathematics Coursework (MIT OpenCourseWare)

- **18.01 – Single Variable Calculus:** Differential and integral calculus with rigorous problem-solving.
- **18.06 – Linear Algebra:** Matrix theory, eigen decomposition, orthogonality, and applications to deep learning.
- **18.05 – Introduction to Probability and Statistics:** Probabilistic modeling, Bayesian reasoning, and statistical inference.

Technical Skills

Languages: Python, C++, SQL

Libraries/Frameworks: NumPy, Pandas, Scikit-Learn, Matplotlib, Seaborn, Plotly, Transformers, Hugging Face, LLMs, TensorFlow, PyTorch

Tools & Platforms: Git/GitHub, Docker, REST APIs, Jupyter, VS Code, Firebase, MySQL, CUDA

Certifications

- CS50 AI - Harvard University
- Supervised Machine Learning: Regression and Classification - DeepLearning.AI, Stanford Online
- Advanced Learning Algorithms - DeepLearning.AI, Stanford Online
- Linear Algebra for Machine Learning - DeepLearning.AI
- Calculus for Machine Learning - DeepLearning.AI
- Python Data Analytics - Meta
- Python for Data Science, AI & Development - IBM
- Scientific Computing with Python - FreeCodeCamp.org
- Feature Engineering - Kaggle

Honors & Awards

HackerRank: 5 Stars in Python Programming

HackerRank: 3 Stars in C++ Programming

Kaggle: 17 Badges for contributions and competitions