Arijit Gupta

arijitgupta2000@gmail.com | +447342887938 | Github | Linkedin | Website

EDUCATION

THE UNIVERSITY OF SHEFFIELD
MSc Computer Science with Speech and Language Processing
September 2024 - Present | Sheffield, UK
Classification: Distinction (Predicted)

BITS PILANI
B.E. Electronics and Instrumentation
August 2018 - May 2022 | Goa, India
Classification: First Class

WORK EXPERIENCE

ZS ASSOCIATES | Data Scientist
Jul 2022 – Aug 2024 | Bangalore, India
- Designed and deployed scalable NLP pipelines using Haystack, LangChain, and PyTorch
- Developed a GenAI-based side-effect analysis system to analyze 5K+ patient reports, leveraging prompt chaining and LLM scoring heuristics.
- Built a few-shot learning pipeline for extracting and segmenting medical entities from 10K+ physician notes using extractive QA and few-shot RAG based methods.
- Collaborated with product teams to define metadata schema and enrich text data for LLM downstream tasks.
- Supported LLM evaluation through quality scoring metrics and comparisons between chained prompt versions.

MICROSOFT | Software Engineer Intern
May 2020 – Jul 2020 | May 2021 – Jul 2021 | Hyderabad, India
- Built chatbot voice assistant features using TypeScript and Express; managed structured training via YAML on the Microsoft Search & Intelligence team.

RESEARCH

THE UNIVERSITY OF EDINBURGH | Visiting Research Student
Aug 2021 – May 2022 | Edinburgh, UK
- Conducted empirical analysis on noun gender prediction using contextual and formal embeddings.
- Worked under Prof. Kenny Smith; findings contributed to deeper understanding of grammatical cue learning in L2 acquisition.
- Explored diachronic word embeddings to model lexical development — a precursor to language evolution modeling.

SELECTED PROJECT

RASOEE | Python, Django, PyTorch |
Aug 2020
- Rasoee is an application for food dish identification using image classification.
- Designed an on-device food classification app across 308 dishes, trained via CNN architecture benchmarks.
- Delivered a full-stack pipeline (mobile + web) optimized for online inference—enabling real-time use.

SKILLS

PROGRAMMING
Python • C++ • C# • JavaScript • LaTeX• MATLAB • TypeScript • Shell

FRAMEWORKS
LangChain • PyTorch • TensorFlow • Keras • Angular • ReactJS

TOOLS
Git • NumPy • OpenCV • Pandas • Scikit-Learn

PUBLICATIONS

- Transformer-based Models for Supervised Monocular Depth Estimation |
- Towards Using Diachronic Distributed Word Representations as Models of Lexical Development |

SCHOLARSHIPS AND COMPETITIONS

2017 UPSTSE Scholarship – Govt. of India - Top 1000 students from my home state
2020 PyTorch Global Summer Hackathon - First Runners-Up out of 2,500 participants from 114 countries
2021 OpenCV Spatial AI Competition - Top 210 finalists from 1500+ worldwide

COURSEWORK

ML and Adaptive Intelligence
Natural Language Processing
Speech Technology
Scalable Machine Learning
Text Processing
Speech Processing
Team Software Project
Object Oriented Programming
Data Structures and Algorithms
Discrete Mathematics

CONFERENCES

ACL 2022 - Student Research Workshop
Research Week with Google
International CCCP Symposium - Cognition, Communication, Computation, and Perception