Arijit GUPTA

in linkedin.com/in/arijitgupta42   github.com/arijitgupta42
@ arijitgupta2000@gmail.com   arijitgupta42.github.io

EDUCATION

Present September 2024	**The University of Sheffield** | MSc in Computer Science with Speech and Language Processing **Classification :** Distinction (Predicted)
May 2022 August 2018	**Birla Institute of Technology and Science, Pilani, Goa** | BE in Electronics and Instrumentation **Classification :** First Class

WORK EXPERIENCE

August 2024
July 2022

ZS Associates | Advanced Data Science Associate
- Developed scalable NLP pipelines using Haystack and PyTorch, reducing manual analysis time by 100+ hours/week across medical datasets.
- Delivered an LLM-based drug-side effect analysis pipeline with Flan-T5, surfacing co-occurrence patterns in 5K+ patient reports—informing treatment strategy optimizations.
- Engineered a document segmentation workflow with few-shot learning for 10K+ physician notes to extract decision-driving factors using extractive QA and clustering.

`LangChain` `HuggingFace` `AWS` `Python`

July 2021
May 2021

Microsoft | Software Engineer Intern
- Was part of the Microsoft Search, Assistance and Intelligence team, and worked on adding voice skills to a state-of-the-art chatbot
- Added the complete pipeline for new, essential features to the bot, using TypeScript and Express to communicate with the bot and YAML to write training data

`TypeScript` `Express` `YAML`

July 2020
May 2020

- Was one of the select 33 second year interns among all 500+ interns
- Migrated a static website to a dynamic website which implemented systematic polling of APIs and real time update of components in ReactJS and deployed the web service on Azure

`JavaScript` `ReactJS` `C#` `AWS`

RESEARCH EXPERIENCE

CENTER FOR LANGUAGE EVOLUTION | THE UNIVERSITY OF EDINBURGH   AUGUST 2021 - MAY 2022
Visiting Research Student | Advisor : *Prof. Kenny Smith*
- Trained vector representations based on contextual and formal properties of words
- Experimented extensively on the usability of both types of embeddings in a task of noun gender prediction

COGNITIVE NEUROSCIENCE LAB | BITS PILANI   JUNE 2020 - MAY 2021
Research Assistant | Advisor : *Dr. Veeky Baths*
- Used word vector representations to study the semantic and syntactic lexical development in children
- Explored the usability of diachronic distributed word representations in cognitive modeling and analysis of the lexical development in children

DEPARTMENT OF EEE | BITS PILANI   AUGUST 2020 - JULY 2022
Research Student | Advisor : *Dr. A. Amalin Prince*
- Worked on inferring depth information from a single RGB image to infer scene geometry in 2D
- Implemented recent vision models not based on convolutions and evaluating their performance for the task
- Performed analysis to investigate the efficiency of these models for edge devices

📄 Publications

Transformer-based Models for Supervised Monocular Depth Estimation 📄

Arijit Gupta, A. Amalin Prince, A. R. Jac Fredo, Femi Robert
IEEE ICICCSP

Towards Using Diachronic Distributed Word Representations as Models of Lexical Development 📄

Arijit Gupta, Rajaswa Patil, Veeky Baths
arXiv (ACL'22 Student Research Workshop)

💻 Select Projects

Rasoee

June 2020 - August 2020

Won First Runner-Up at Global PyTorch Summer Hackathon 2020

🐙 Code   🔗 Website

> ❯ Web and android application that identifies food dishes from images. The image classification model is trained in Pytorch using a EfficientNet B2 model as the backbone, and some pre-processing done on the images
> ❯ The data of 308 unique classes of food items was scraped from the web and cleaned. The mobile application, built with Kotlin, runs completely on the edge and no active internet connection is required to identify the dishes
> ❯ The application also takes an image input, and returns the identified food item, its primary ingredients, link to a recipe and link to video tutorials for that dish
> ❯ The website is built with Django and hosted on Heroku

Python | Django | Kotlin | PyTorch

ChainVoter

June 2019 - July 2019

Won First Runner-Up at Microsoft CodeFunDo++ Hackathon

🐙 Code

> ❯ Web and mobile application for secure online voting using blockchain.
> ❯ Implements two-factor authentication of voters using face recognition with Microsoft Face API and MongoDB database
> ❯ In the blockchain the creation of block uses User ID(containing Name, Party, Constituency, Aadhaar, Voter ID, and Signature(Name, Aadhaar, Voter ID) and hashes them. Also checks if the vote is a duplicate before adding it to the ledger

Javascript | Angular | Typescript | CSS

📚 Relevant Coursework

> ❯ ML and Adaptive Intelligence
> ❯ Natural Language Processing
> ❯ Speech Technology
> ❯ Scalable Machine Learning
> ❯ Text Processing

> ❯ Speech Processing
> ❯ Team Software Project
> ❯ OOP Programming
> ❯ Data Structures and Algorithms
> ❯ Discrete Mathematics

☰ Technical Skills

Programming	Python , C++ , C# , JavaScript , LaTeX, MATLAB , TypeScript , Shell
Frameworks	LangChain , PyTorch , TensorFlow ,Keras , Angular , ReactJS
Tools	Git , NumPy , OpenCV , Pandas , Scikit-Learn

👥 Conferences and Workshops

60th Annual Meeting of the Association for Computational Linguistics (ACL '22)
May 2022 • Organized by The Association for Computational Linguistics

> Presented our work on models of child lexicon development at the student research workshop
> Was an online student volunteer for the virtual part of the conference

Research Week with Google
Feb 2022 • Organized by Google Research, India

> Among the 50 participants selected as a part of the NLU track across undergraduate, master's and Ph.D. students across India

International CCCP (Cognition, Communication, Computation, and Perception) Symposium
Dec 2020 • Organized by The Institute for Cognitive Neuroscience, Higher School of Economics, Moscow

> Was one of the selected speakers at the iBrain Erasmus+ round-table, part of the *'Data Issues in Cognitive Neuroscience'* track
> Spoke on issues in child language data with respect to availability and privacy among other aspects

⑁ Other Projects

> OCR Sudoku Solver
> Double DQN for Atari Pong in OpenAI Gym
> Text Cipher Algorithms in C++

> Object Detection in Real Time and in Images using YOLO v3
> Forest Cover Binary Classification using an Ensemble Model
> Russian AI Cup 2019 Bot (1st Round Qualified)

🏆 Achievements

Scholarships :
> 2017 • Uttar Pradesh Science Talent Search Exam **(UPSTSE) Scholarship** by Govt. of India

Competitions :
> 2020 • PyTorch Global Summer Hackathon : **First Runners-Up** in Web/Mobile Application Category out of **2,500 participants from 114 different countries**
> 2021 • OpenCV Spatial AI Competition : Among the **top 210 finalists** from **over 1500** submissions worldwide and selected as one of the **25 general teams** in the **Central and South Asia** Region.

ⓘ Miscellaneous

TECHNICAL WRITER @ Towards Data Science | You can find my articles here
HOBBIES | **Drums** (Performing member of the institute music society), **Quiz Bowls** (Active member of the univeristy quiz society)