

Aneesh Bose

+91-9831190785 | aneeshbose2296@gmail.com | [LinkedIn](#) | [GitHub](#)

EDUCATION

- Manipal Institute of Technology** Manipal, IN
Bachelor of Technology in Computer Science, Minor in Intelligent Systems; CGPA: 8.44/10 Jul 2019
Relevant Coursework: Artificial Intelligence, Machine Learning, Natural Language Processing, Soft Computing

WORK EXPERIENCE

- Microsoft** Bengaluru, IN
Data and Applied Scientist 2 Dec 2021-Present
 - Designed and implemented a RAG + Knowledge Graph architecture for personalized trailing prompts in Sales Copilot within MSX D365 to facilitate exploration of **59K** Microsoft sellers across **40 plugins** published in MSX Skills Studio.
 - Launched **'Copilot Voice', the first voice-based AI pair-programmer, at GitHub Universe 2022**, along with four members from GitHub Next & Microsoft, that had **5000+** sign-ups for the waitlist within 2 days of the launch.
 - Fine-tuned a PyTorch BERT model using Azure ML for intent classification of a user's natural language utterances to VS Code APIs with a validation accuracy of **93%** on a custom dataset generated using OpenAI and Streamlit.
 - Created an explanations pipeline using TreeSHAP Explainer on Gradient-Boosted Trees for **4.03M** solutions that were surfaced in 7 hours - a processing time of **160.15 rows (with 1028 features per row) per second**.
 - Published a custom SHAP Explainer package to **PyPi.org** that generates locally interpretable Shapley values for any tree based model. The package is being used by all the team members to interpret their ML models to stakeholders. [Link](#)
 - Achieved a **10%** conversion improvement (**12.5M USD**) in rule-based recommendations churned out by Microsoft Daily Recommender using frequent itemset mining and Apriori metrics like support, confidence, lift, leverage & conviction.
- Software Development Engineer* Jul 2019-Nov 2021
 - Implemented bulk CRUD operations using Cosmos DB and Azure Service Bus topics to introduce Virtual Position Number (VPN) based seller assignments which accounted for **82% of the 26M** transactional records.
 - Owined the refactoring of underlying monolithic Azure Analysis Services and Azure Function Apps which reduced processing & refresh failures of PowerBI business reports by **10X** and minimized Azure spend by **30%**.
 - Served as primary on-call individual for customer reported and system livesite issues across **17** geographies during business critical planning and rollover phase that saw a reduction in tool downtime from **2 weeks to 23 hours**.
- Centre for AI & Robotics, DRDO - Intelligent Systems (ISRD) Lab** Bengaluru, IN
Research Intern, Advised by Ashish Srivastava Jan 2019-Jun 2019
 - Performed semantic query search using co-occurrence clustering and tf-idf weighting in word graphs by constructing a 1.5 degree egocentric network using query words as nodes in the NetworkX library and plotting them using Gephi.
 - Achieved a **Mean Average Precision (mAP) of 79.33%** ranking top-10 documents for domain question answering on the 20newsgroups dataset by using cosine similarity and Okapi Best Match (BM25) on the pruned subgraph. [Project](#)

PUBLICATIONS

- HaVQA: A Dataset for Visual Question Answering and Multimodal Research in the Hausa Language**
Shantipriya Parida, Idris Abdulmumin, Shamsuddeen Hassan Muhammad, Aneesh Bose, Guneet Kohli, Ibrahim Said Ahmad, Ketan Kotwal, Sayan Deb Sarkar, Ondej Bojar, Habeebah Kakudi
Accepted at **ACL 2023** as a long paper (Findings) - [Paper](#)

PATENTS

- Slot Extraction for Intents using Large Language Models:** *Patent Status - Filed on 31 Oct, 2022*
- Large Language Model Utterance Augmentation:** *Patent Status - Filed on 31 Oct, 2022*
- Attributing meaning to Utterance Terms based on Context:** *Patent Status - Filed on 31 Oct, 2022*
- Retrieval-Augmented Prompt for Intent Detection:** *Patent Status - Filed on 11 May, 2023*

TECHNICAL SKILLS

- Languages & Frameworks:** Python, SQL, C++, C#, Java, TypeScript, Flask, PyTorch
- Technologies:** Git, Azure Stack - [Databricks, Data Factory, Data Lake, Synapse, Cognitive Services, Service Bus, CosmosDB, Speech Studio, ML Studio, AI Studio, Prompt Flow], Docker, Grafana
- Blogs:** [Parallelize your massive SHAP computations with MLlib and PySpark](#) - Published in Towards Data Science
- Certifications:** Coursera - [NLP Specialization](#), [Neural Networks and Deep Learning](#), IIT Mandi - [Deep Learning Certification](#)

PROJECTS

- **Meeting360** | *PySpark, Databricks, Azure Cognitive Services, MS Teams Toolkit* *Sep 2021*
 - Developed the backend for Teams Toolkit VS Code Extension that informs participants about their meeting performance based on metrics like filler word usage, sentiment, words per minute & inclusivity.
 - Integrated generation of meeting summaries given keywords and translation of those summaries into English, Hindi, French & Spanish using Microsoft Translator and DeepAI's Text Generation API based on OpenAI GPT-2.
- **PixelWave** | *Flask, Firebase, Heroku, Azure Cognitive Services, Computer Vision* *Aug 2020*
 - Led a team of 4 to develop an OCR solution that extracts text from handwritten requisition forms, stores them in Firebase and digitizes them using Microsoft's Computer Vision Read API for patients suffering from colon cancer. [Project](#), [Video](#)
- **Image Forgery Localization** | *TensorFlow, Keras, Numpy* *Mar 2019*
 - Trained a VGG16 U-Net architecture with pre-trained ImageNet weights using Binary Cross Entropy loss and Adam optimizer for an image segmentation task to identify fake objects in real images and generate their masks.
 - Achieved a training accuracy of **94.22%** using the ReLU and Softmax activations in the convolution and output layers respectively. [Project](#)

AWARDS

- **Microsoft Global Hackathon:** **First** in Hack For Developers, **Fourth** in Hack For Observational Assistance for **Hello Cody**, that was finally launched as **Hey, GitHub!** **Second** in Intelligent Employee Experience, **Third** in Observational Assistance and Honorable Mention in Global Language Usability for **Meeting360**. Number of participating teams - 6400. (Dec 2021)
- **Exact Sciences OCR Hack, HackerEarth:** Overall **First** place for **PixelWave OCR** with a **2500 USD** prize money. Number of participating teams - 730. (Aug 2020)
- **CSA Open Day Data Science Hackathon, IISc Bengaluru:** Overall **First** place for **Image Forgery Localization**. Number of participating teams - 200 (Mar 2019)
- **Competitive Programming:** (Cleared 2 rounds for all) **Google CodeJam 2020** - Scored 42/100 in Qualifiers. **Codechef Snackdown 2019** Round 1 Rank - 436/15856, 2021 Round 1A Rank - 1646/28912. (2019-2021)