Aneesh Bose

(983)119-0785 | aneeshbose2296@gmail.com | LinkedIn | GitHub

Education

- **Carnegie Mellon University, School of Computer Science** — Pittsburgh, PA
  Master of Science in Artificial Intelligence and Innovation — *Fall 2025*
  Relevant Coursework: Foundations of Computational Data Science, Intro to Machine Learning

- **Manipal Institute of Technology** — Manipal, IN
  Bachelor of Technology in Computer Science, Minor in Intelligent Systems; CGPA: 8.44/10 — *Jul 2015-2019*
  Relevant Coursework: Artificial Intelligence, Machine Learning, Natural Language Processing, Soft Computing

Work Experience

- **Microsoft & GitHub** — Bengaluru, IN
  Software Engineer / Data and Applied Scientist 2 — *Jul 2019-Present*
  - Deployed a Knowledge Graph endpoint in Sales Copilot leveraging RAG from an HNSW-indexed vector DB, boosting CTR (click through rate) by **51%** and diversifying plugin invocations by **3X for 35K sellers across 50 plugins**.
  - 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 launch.
  - Fine-tuned a BERT model for multi-class intent classification that maps a user's STT (speech-to-text) transcribed instruction to VS Code API commands with a validation accuracy of **93%** on a custom generated dataset.
  - Created an explanations pipeline with TreeSHAP Explainer on Gradient-Boosted Trees for **4.03M** solutions, surfacing results in 7 hours - a processing rate of **160.15 rows (with 1028 features per row)** per second.
  - Achieved a **10%** conversion improvement **(12.5M USD)** in rule-based recommendations from Daily Recommender by applying frequent itemset mining and Apriori algorithm metrics (support, confidence, lift, leverage & conviction).
  - Published a custom SHAP Explainer package to PyPi.org for generating locally interpretable Shapley values for **any** tree-based model. The package is actively used by my team to interpret ML models for stakeholders. Link

- **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 through co-occurrence clustering and tf-idf weighting in word graphs, constructing a 1.5-degree egocentric network with query words as nodes in the NetworkX library and plotted them in Gephi.
  - Achieved a **Mean Average Precision (mAP) of 79.33%** ranking top-10 documents for domain question answering on the 20newsgroups dataset by applying 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 - Published on 2 May, 2024* Link
- **Large Language Model Utterance Augmentation:** *Patent Status - Published on 2 May, 2024* Link
- **Attributing meaning to Utterance Terms based on Context:** *Patent Status - Published on 2 May, 2024* Link
- **Retrieval-Augmented Prompt for Intent Detection:** *Patent Status - Filed on 11 May, 2023*

Skills

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

Awards

- **Microsoft Global Hackathon:** Secured **first** place in "Hack For Developers" & "Hack For Observational Assistance" Executive Challenges (out of 6400 teams across Microsoft) for **Copilot Voice**—an AI-assisted VS Code extension powered by voice, targeted towards helping developers with accessibility needs. (Dec 2021)
- **Exact Sciences OCR Hack, HackerEarth:** Won **first** place (out of 730 teams, with **2500 USD** prize money) for PixelWave, an OCR solution extracting text from handwritten forms for colon cancer patients. (Aug 2020)
- **CSA Open Day Data Science Hackathon, IISc Bengaluru:** Placed **first** (out of 200 teams) for Image Forgery Localization - an image segmentation task to detect fake objects and generate their masks. (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)