

# Sharu Kumo

[sharu.kumo@gmail.com](mailto:sharu.kumo@gmail.com) ♦ (555) 555-5555 ♦ San Jose, CA

## EDUCATION

### Carnegie Mellon University

Master of Science in Information Systems

Coursework: Machine Learning, Deep Learning, Natural Language Processing, Large Language Models **GPA: 3.99/4.0**

Pittsburgh, PA

May 2023 Selected

## WORK EXPERIENCE

### Amazon

Boston, MA Applied Scientist Intern

May 2023 – Aug 2023

- Finetuned a **Multilingual DeBERTa** model to auto-moderate answers submitted to the Alexa Answers website.
- Developed model outperformed the production **BERT** model and improved the automation rate by **15.81%** which resulted in auto-approving 500k answers and **saved 100,000 USD** in data labelling.
- Designed a novel **Informative-Diverse Algorithm** to select the optimal in-context examples for few-shot prompting of **Large Language Models (LLM's)** such as **LLaMA**, **Falcon** and **GPT-4** and improved the F1-score by **8%**.
- Proposed an innovative **Counterfactual Data Augmentation** method to facilitate the Chain-of-Thought (CoT) reasoning of LLM's and further improved the F1-score by **4%**. Published the findings from this work to **Amazon ML Conference**.

### Carnegie Mellon University

Research Assistant

Pittsburgh, PA

Aug 2022 – Dec 2023

- Leveraged **ChatGPT** for **QA** and improved the **Question-Answering** ability by **30%** through eliminating the irrelevant information in multi-turn conversations. Published this work in **ACL 2023 Conference**.
- Introduced a DeMUX framework to optimally select the most informative datapoints from pool of unlabelled multilingual datapoints. DeMUX supports Uncertainty, KNN Uncertainty, Distance based **Active Learning Algorithms**.
- DeMUX improved 11-points in F1-Score on token-level benchmarks and 5-points on more complex tasks over the previous state-of-the-art baselines. Findings of this work are submitted to **NAACL 2024 Conference**.

### ServiceNow

Software Engineer

Hyderabad, India

Jan 2020 – Jul 2022

- Developed **Natural Language Understanding (NLU)** module for ServiceNow's **ChatBot** by finetuning **BERT Transformer** model using **PyTorch Framework** and **HuggingFace Library** on ServiceNow's domain data. Improved **Intent Classification** score by **10%** over previous **Bi-LSTM** approach.
- **Summarized** long customer tickets by using **T5 Transformer** and obtained strong **BERTScore** of 0.95.
- Pioneered an incident-problem-change correlation modelling in ServiceNow's ATG Hackathon using **K-Means Clustering** and converted the clustering model to an end-to-end product within **2 sprints**.
- Collaborated with cross-team(DevOps) and engineered a **XG-Boost** classifier to predict build failures and obtained a strong **Recall** of **0.95**.

## PROJECTS

### Enhancing the Multilingual Instruction Following Capabilities of LLM's

- Curated two **Multilingual Instruction Following** datasets by machine translating the **Alpaca Dataset** and performing **Few-Shot Prompting** of ChatGPT to generate new instruction, output pairs in 10 languages.
- Finetuned the various variants of LLaMA-2 and PolyLM models on both datasets and improved the **Cross-Lingual** and **Multilingual** instruction following ability by **30%** and **35%** on Question Answering (QA) benchmarks.
- Performed the detail analysis and identified the optimal setting of multilingual instruction finetuning for **Low-Resource** and **High-Resource** languages using curated datasets. SKILLS

- **Programming Languages:** Python, Java, SQL, C++, Javascript
- **AI & ML:** PyTorch, Hugging Face, LangChain, Spacy, TensorFlow, Pandas, Numpy, Scikit-learn, Matplotlib, Git, AWS

## SELECTED PUBLICATIONS

- **Sharu Kumo\***, **Sean William\***, Monu Sinki, Sri Bunny, T Mils, Eric N. “**QA Systems**”. ACL 2023
- **Sharu Kumo**, Sean K, J Steve. “**Identification of XXXXX and Change Requests**”. IEEE ICIRCA 2021.