

Syed Arbaaz Qureshi (he/him/his)

✉ arbaaz.qureshi29@gmail.com

🔗 [Google Scholar](#)

🌐 [linkedin.com/in/arbaazqureshi/](https://www.linkedin.com/in/arbaazqureshi/)

🎓 arbaazqureshi.me

EDUCATION

University of Massachusetts Amherst
Master of Science in Computer Science. GPA: 4.0/4.0

September 2021 - May 2023

Indian Institute of Technology Patna
Bachelor of Technology in Computer Science and Engineering. GPA: 8.2/10

August 2015 - May 2019

SKILLS

Concepts: Natural Language Processing, Large Language Models, Deep Learning, Machine Learning, Data Science, Retrieval Augmented Generation, Transformers

Software/Toolkits: TensorFlow, Keras, PyTorch, LangChain, LlamaIndex, NumPy, SciPy, Pandas, Scikit-learn, Matplotlib, Plotly, MySQL

Programming languages: Python, Java, C#, C, MySQL, CosmosDB, Distributed Systems, MS Azure, AWS, Shell Scripting

WORK EXPERIENCE

Lowe's Companies, Inc. Data Scientist.

July 2023 - Present

• Advisor: [Dr. Surya Kallumadi](#)

• Projects:

1) LLM-based conversation guided recommender system:

- Built a multi-modal chatbot from scratch, for product discovery/purchase/comparison and home improvement question answering for the shopping website. Also built evaluation suites for various modules of this chatbot.
- Experimented with different large language models (GPT, Gemini, LLaMA etc), and with different paradigms of in-context learning (few-shot learning and chain-of-thought reasoning for using tools and making appropriate API calls, retrieval-augmented generation for question answering) to realize an LLM agent which guides customers throughout their online shopping journey.

2) Context-aware query auto-completion system:

- Built a query re-ranker that ranks auto-completions based on semantic similarity with the session's previous queries. This system is deployed on Lowe's shopping website.
- This framework shows improved recall and a reduced average number of characters typed for the ground truth to appear in the auto-complete suggestions, with an insignificant increase in end-to-end latency. A/B test reveals a 1% improvement in product conversion rate and about 2% improvement in user engagement with the auto-complete suggestions.
- Currently working towards replacing the existing most popular completion (MPC) based query retriever with a context-aware query retriever, to provide semantically similar prefix-appropriate query suggestions to the aforementioned re-ranker.

IBM. Graduate Student Researcher.

February 2023 - June 2023

• Advisors: [Dr. Taesung Lee](#) and [Dr. Youngja Park](#)

• Project: **Towards generating informative textual description for neurons in language models [AAAI-24]:**

- Devised a method to obtain text descriptions per neuron, of concepts that activate those neurons in a BERT model. Our method alleviates the requirement of Human in the loop, to generate the text descriptions.
- Created a dataset of 72K Amazon reviews and their annotated concepts, using various open-source LLMs like Flan-T5 XL and Pythia 12B. Analyzed which annotated concepts trigger neurons in the BERT language model using this dataset.

Google. Engineering Intern in the Pixel Watch Ambient Compute team.

September 2022 - December 2022

• Advisor: [Dr. Cac Nguyen](#)

• Project: **Low latency off-body detection on the Google Pixel Watch**

- Created an end-to-end framework which loads and processes wear data from the cloud, trains any generic neural network on the wear data to detect whether the watch is on-wrist or off-wrist, and evaluates how the network performs in a real-world setting.

- Performed extensive experimentation (building, debugging, training and evaluation) with different convolutional neural networks and different off-body detection algorithms, and achieved a performance which competes with the deployed heuristic algorithm.

Microsoft Research. Research Fellow.

August 2019 - August 2021

- Advisors: [Ms. Sonu Mehta](#), [Dr. Rahul Kumar](#) & [Dr. Ranjita Bhagwan](#)
- Project: **ML for DevOps tasks involving code-edits**

- Tasked with generating commit messages, and automatically classifying edit, based on the content of the commit.
- Scraped commit-commit message pairs from more than 1000 GitHub repositories. Built ACMG (Automatic Commit Message Generator), based on an existing work [code2vec](#), and trained it on the scraped dataset.
- Built edit2vec (an extension of code2vec) for automatic code edit classification. Trained and analyzed it on the ManySStuBs4J dataset, and on another dataset that we collected.
- Achieved an accuracy of over 99% on code edit classification. Discovered code2vec isn't generalizable to other downstream tasks.

AI-NLP-ML lab, IIT Patna. Research Assistant.

August 2018 - May 2019

- Advisors: [Dr. Sriparna Saha](#), [Dr. Gaël Dias](#) & [Dr. Mohammed Hasanuzzaman](#)
- Projects:

1) Multitask representation learning for multimodal estimation of depression severity [IEEE IS 2019]

- Developed different multitask learning model architectures to learn representations of individual modalities, by simultaneously predicting depression severity score and class.
- Outperformed the state of the art by 4.93% on RMSE and 1.50% on MAE. Set new baseline for depression classification, 66.66% accuracy and 0.53 F1-score.

2) Multitask learning to concurrently estimate emotion intensity and depression severity [IEEE CIM 2020]

- Designed and trained various multitask learning model architectures (fully shared, shared private and adversarial shared private) to concurrently predict depression score and emotion intensity using text data.
- Showed that substantial performance improvements in predicting the depression severity can be achieved by using emotion-aware models.

SELECTED PROJECTS

Research Paper Tagger (RPT), with [Dr. Mohit Iyyer](#), at UMass Amherst.

Objective: Automatically tagging the research track of an NLP paper, given the title, abstract and the authors.

- Built a dataset of 1744 research paper-research track pairs from ACL 2021, 2020 and 2019. Fine-tuned a BERT-based classifier on the collected dataset, on various combinations of title, abstract and the authors of the research papers.
- Achieved a top-1 accuracy of 70% and a top-3 accuracy of 83% on the test split of the collected dataset. Working on publishing this work in a conference/workshop.

PUBLICATIONS

1. **Towards generating informative textual description for neurons in language models**
Association for the Advancement of Artificial Intelligence (AAAI) ReLM workshop, 2024 [[link](#)]
2. **Gender-Aware Estimation of Depression Severity Level in a Multimodal Setting**
International Joint Conference on Neural Networks (IJCNN), 2021. [[pdf](#)]
3. **Improving Depression Level Estimation by Concurrently Learning Emotion Intensity**
IEEE Computational Intelligence Magazine (IEEE CIM), Volume 15, Issue 3, 2020. [[pdf](#)]
4. **Multitask Representation Learning for Multimodal Estimation of Depression Level**
IEEE Intelligent Systems, Volume 34, Issue 5, 2019. [[pdf](#)]
5. **Automatic Prediction of PHQ-8 Questionnaire Scores using Artificial Intelligence**
French Journal of Psychiatry, Volume 1, Supplement 2, 2019 [[pdf](#)]