

# DEBANJAN MONDAL

☎ +1 413-472-6298 ✉ [debanjanmond@umass.edu](mailto:debanjanmond@umass.edu) [in linkedin.com/in/debanjanmondal702](https://www.linkedin.com/in/debanjanmondal702) [github.com/Demon702](https://github.com/Demon702)

## Education

### University of Massachusetts Amherst

Master of Science in Computer Science (Data Science Concentration) | GPA: 3.89/4  
Courses: NLP, Computer Vision, Reinforcement Learning, Systems for Deep Learning

Sep 2022 – May 2024  
Amherst, MA, USA

### Indian Institute of Technology Bombay

Bachelor of Technology with Honors in Computer Science | GPA: 8.21/10  
Courses: Data Structures & Algorithms, Database, Distributed Systems, Programming Languages

Jul 2016 – May 2020  
Mumbai, India

## Experience

### Allen Institute for AI, Student Researcher

Natural Language Processing, Generative AI, LLM, GPT, LLaMA, Python | Advisor: [Niket Tandon](#)

Sep 2023 – Present  
Amherst, MA, USA

- Observed strong correlation between human moral judgment and GPT-3.5 scores by exploiting existing moral theories.
- Built a multi-cultural values dataset with 20M datapoints where GPT-3.5 and LLaMA 2 showed subpar performance.
- Developing LLM compatible moral theories through refining open-source LLMs like Mixtral using RL techniques.

### UMass Amherst Center for Data Science, Machine Learning Intern

Python, PyTorch, Azure, Computer Vision, Object Detection, Statistics

Jun 2023 – Aug 2023  
Amherst, MA, USA

- Collaborated with [Red Cross](#) to count damaged buildings with 4 severity levels from post disaster satellite images.
- Applied importance sampling to enhance accuracy by 10% for U-Net based building detector model with few labels.
- Deployed the detector and damage classification model in Azure and built a user-friendly labeling tool interface.

### Allen Institute for AI, Graduate Student Researcher

Natural Language Processing, GPT, PyTorch, WandB | Advisor: [Xiang Lorraine Li](#), [Niket Tandon](#)

Feb 2023 – May 2023  
Amherst, MA, USA

- Demonstrated a strong correlation between commonsense judgments and early layer MLP parameters in GPT-2 models.
- Proposed *MEMIT<sub>CSK</sub>*, a model parameter editing strategy to edit commonsense mistakes in GPT-2 Large and XL models that outperform baseline fine-tuning strategies by more than 10.5% accuracy on PEP3k and 20Q datasets.
- Constructed a dataset of 21000 challenging commonsense statements for model editing evaluation using GPT-3.

### LG Ad Solutions, Software Engineer

React, Node.js, MySQL, MongoDB, Kubernetes, Spark, AWS S3, Hadoop, Elasticsearch, SSH

Aug 2020 – Jun 2022  
Bangalore, India

- Designed and implemented a central User Management Server, enabling user authentication flow across 10+ services.
- Developed 6 applications in React, Node.js, MySQL, set up Jenkins pipeline to automate deployment in Kubernetes.
- Composed and executed periodic Spark jobs to process 100 GBs of data in Hadoop and ingested in ElasticSearch.

### LG Ad Solutions, Engineering Intern

React, Node.js, Redux

May 2019 – Jul 2019  
Bangalore, India

- Built a React-based web application for browsing over 10000 television shows and ads aired on US Television.
- Developed a resource allocation system in Node.js to manage more than 100 concurrent requests.

## Publications

### “Editing Common Sense in Transformers” [\[Code\]](#)

EMNLP '23

- Enhanced GPT-2 Large and XL models by rectifying commonsense errors through editing early layer MLP parameters.

### “Understanding Code Semantics: An Evaluation of Transformer Models in Summarization” [\[Code\]](#)

EMNLP '23

- Evaluated LLM's code summarization ability with transformations like renaming identifiers, adding commented code.

### “VALUEALIGN: A Large-scale Dataset for Multi-Cultural Value Alignment” (*In Review*)

LREC-COLING '24

- Curated a dataset of 20 million question-answer pairs for socio-cultural values with corresponding user demographics.
- Demonstrated that, provided user demography, GPT-3.5 and LLAMA 2 exhibit poor alignment with human answers.

## Academic Projects

### Optimization and Evaluation of RL algorithms [\[Code\]](#) | Reinforcement Learning, PPO

Oct 2023 – Dec 2023

- Composed PPO, Actor-Critic algorithms from scratch with optimizations and evaluated on diverse environments.

### Robust Point Cloud Classification with Point Transformer [\[Code\]](#) | 3D Vision, PyTorch

Feb 2023 – May 2023

- Conducted evaluation of the Point Transformer model on the corrupted Point Cloud Classification benchmark [ModelNet40-C](#), assessing 5 different data augmentation strategies while training on clean data.
- Noticed that RSMix training augmentation yielded a 26% increase in accuracy over the baseline training strategy.

## Skills

**ML Tools:** PyTorch, TensorFlow, Keras, OpenCV, spaCy, NLTK, Sklearn, WandB, LLM, Pandas, Prompt Engineering

**Languages:** Python, C++, C, C#, PySpark, Java, Javascript, Spark, Scala

**Web Development:** React, Node.js, MySQL, MongoDB, Flutter, TypeScript, Django, ASP.NET, Elasticsearch

**Technologies/Frameworks:** Linux, Jenkins, Git, Docker, Kubernetes, Hadoop, AWS, S3, Azure, SSH, Unity, OpenGL