

Shubhashis Roy Dipta

No visa sponsorship required; Green Card holder.



sroydip1@umbc.edu | [in/dipta007](https://in.dipta007) | RoyDipta.com | dipta007 | [\(443\) 889-3961](tel:(443)889-3961)

EDUCATION

Ph.D. in Computer Science, UNIVERSITY OF MARYLAND, BALTIMORE COUNTY

Jan 2021 – December 2026 (Expected)

Specialization: **Decomposition-based Multimodal Reasoning, Tool Calling LLM, Math Reasoning**

- Phi Kappa Phi Award - **Top 10% of STEM**. GPA 4.00/4.00 – **Top 1% of class**.

EXPERIENCE

Applied Research Scientist Intern, AMAZON

June 2025 – Aug 2025

Tech Stack: Tool-calling LLM, Agentic LLM, PyTorch, Multi-Stage SFT, PPO, GRPO, RL, Weight & Biases

- **Injecting Tool-use Knowledge**: Developed multi-stage continual training to inject tool-use policies, outperforming τ -Bench baselines by **20 points** without in-context policies while decreasing the **context length by 40%**.
- **GRPO RL**: Introduced a novel **PolicyRecall** reward with Jaccard score and hallucination penalty for precise policy adherence in LLM.
- **Tau-Override**: Developed a synthetic test benchmark, τ -Override, to evaluate the ability of LLMs to override already learned rules.

Machine Learning Research Intern, SCALE.AI

June 2024 – Aug 2024

Tech Stack: PyTorch, DeepSpeed, Hugging Face, Kubernetes, dagster, RLHF, SQL, Weight & Biases

- **RLHF text2sql**: Used Online KTO (a novel method) and data augmentation to improve the BIRD benchmark accuracy by **8 points** over the same size SFT model (featured in [Scale.AI Blog](#)).
- **Auto Eval**: Implemented AutoCOT, Self Critique to improve the current Auto Eval system by 6 points (**F1 score: 83 → 89**).
- **Many-Shot text2sql**: Implemented Many-Shot *text2sql* system, improving the total accuracy by **8%** and per use case accuracy by **12%**.

Graduate Research Assistant, UNIVERSITY OF MARYLAND, BALTIMORE COUNTY

Jan 2021 – Present

Tech Stack: PyTorch, Python, Hugging Face, Scikit-Learn, LLM, SQL, Spacy, NLTK, Pandas, Numpy, Matplotlib, Seaborn, Weight & Biases, Hydra

- **Reasoning**: Conducting research on Decomposition-based Reasoning using text and vision data (Publications [1](#), [2](#), [3](#), [4](#))
- **Event Understanding**: Developed a hierarchical variational autoencoder for Event Representation Learning (Publication [5](#))

Machine Learning Engineer, BACKPACKBANG.COM

Oct 2018 – Mar 2019

Tech Stack: PyTorch, Keras, Hugging Face, Large Language Model, NLTK, Python, MySQL, Elasticsearch, AWS EC2, Node.js, React

- **Boosted the sale by $\approx 23\%$** by improving existing product recommendation system using *Product2Vec* embedding.
- **Decreased server cost by $\approx 10\%$** by implementing an AWS Lambda-based ML pipeline for online learning.

Founder & Chief Technology Officer, UNISHOPR.COM

Jan 2019 – Jan 2021

Tech Stack: Python, Node.js, React, PostgreSQL, GraphQL, AWS Lambda, AWS Lightsail

- Single-handedly led a cross-functional team of 10, achieving **1,000+** active e-commerce users and **\$100,000+/month** in orders.

PUBLICATIONS (FIRST AUTHORED)

For a comprehensive list, visit my [Google Scholar Profile](#)

1. "If We May De-Presuppose: Robustly Verifying Claims through Presupposition-Free Question Decomposition", 14th Joint Conference on Lexical and Computational Semantics, EMNLP (2025)
2. "VC-Inspector: Advancing Reference-free Evaluation of Video Captions with Factual Analysis" ([arXiv](#))
3. "Q2E: Query-to-Event Decomposition for Zero-Shot Multilingual Text-to-Video Retrieval" ([arXiv](#))
4. "Semantically-informed Hierarchical Event Modeling", 12th Joint Conference on Lexical and Computational Semantics, ACL (2023)
5. "UMBCLU at SemEval-2024 Task 1A and 1C: Semantic Textual Relatedness with and without machine translation", Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2024), NAACL (2024)
6. "HU at SemEval-2024 Task 8A: Can Contrastive Learning Learn Embeddings to Detect Machine-Generated Text?", Proceedings of the 17th International Workshop on Semantic Evaluation, (SemEval-2024), NAACL (2024)
7. "Accurate protein malonylation site predictor using structural and evolutionary information", Computers in biology and medicine (2020)

COLLABORATIONS

1. Sarwar, Nobin, and Shubhashis Roy Dipta "FedMentor: Domain-Aware Differential Privacy for Heterogeneous Federated LLMs in Mental Health" ([arXiv](#))

RELEVANT PROJECTS

For a comprehensive list, visit my [Portfolio](#)

Bird Chirping Identification [🔗](#)

Tech Stack: PyTorch, Keras, TensorFlow, Python, Matplotlib, Seaborn, EfficientNet

- Secured **Bronze Medal** in Kaggle (**Top 6%**) by using EfficientNet on the spectrogram with a weighted ensemble of framed timespan.

Amazon [Crawler] [Search Engine]

Tech Stack: Elasticsearch, Node.js, Express.js, Python, Multithreading, GCP

- Designed a distributed data pipeline using GCE to extract and index over **10M products**, enhancing resource-efficient data retrieval.

PROFESSIONAL SERVICES

- **Reviewer**: Reviewed 28+ papers in top ML conferences (NeurIPS, CVPR, ACL, NAACL, CoNLL) & Bioinformatics journals.
- **Competitions**: 2 International Robotics Competitions (URC, USA; ERC, Poland), 2 ACM-ICPC, 25+ National Programming Competitions.