

Shubhashis Roy Dipta

Last updated: April 2024

🏠 RoyDipta.com
☎ (+1) 443 889 3961
✉ sroydip1@umbc.edu
📄 notes.RoyDipta.com
Socials: [G](#) [in](#) [8](#)

Research Interest

Multimodal Event Understanding & Generation
Vision Language Model, Large Language Model
Artificial Intelligence, Machine Learning, Natural Language Processing

Education

- Jan 2021–June 2026 **Ph.D. in Computer Science**
University of Maryland, Baltimore County (UMBC), USA
Topic: Multimodal Event Understanding & Generation
Advisor: Dr. Frank Ferraro
GPA – 4.00/4.00
- Jan 2021–May 2023 **M.Sc. in Computer Science**
University of Maryland, Baltimore County (UMBC), USA
Phi Kappa Phi (Top 10% of STEM)
GPA – 4.00/4.00 (Top 1% of Class)
- Jan 2013–Dec 2016 **B.Sc. in Computer Science & Engineering**
Military Institute of Science & Technology (MIST), Bangladesh
Ranked 9th in University Rover Challenge 2015, USA (Top 3%)
Participated in 30+ National & International Programming Competitions
GPA – 3.51/4.00

Publications

Total Citations: 76; h-index: 5; (Source: Google Scholar) (*Equal Contribution)

- NAACL 2024 [10] **Shubhashis Roy Dipta***, and Sai Vallurupalli*. “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)
- NAACL 2024 [9] **Shubhashis Roy Dipta**, and Sadat Shahriar. “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)
- ACL 2023 [8] **Shubhashis Roy Dipta**, Mehdi Rezaee, and Francis Ferraro. “Semantically-informed Hierarchical Event Modeling.” Proceedings of the 11th Joint Conference on Lexical and Computational Semantics (*SEM 2023)
- Springer 2022 [7] Sadia Islam, Shafayat Bin Shabbir Mugdha, **Shubhashis Roy Dipta**, [4 other Co-Authors]. “MethEvo: an accurate evolutionary information based methylation site predictor.” Neural Computing and Applications
- Preprint (arXiv) [6] Sourajit Saha*, and **Shubhashis Roy Dipta***. “SeeBel: Seeing is Believing.” arXiv preprint arXiv:2312.10933
- Elsevier 2020 [5] **Shubhashis Roy Dipta**, [5 other Co-Authors]. “SEMAl: Accurate protein malonylation site predictor using structural and evolutionary information.” Computers in biology and medicine
- Genes 2020 [4] Md Easin Arafat, [9 Co-Authors including **Shubhashis Roy Dipta**]. “Accurately predicting glutarylation sites using sequential bi-peptide-based evolutionary features.” Genes 11, no. 9
- IEEE Access 2020 [3] Md Wakil Ahmad, [7 Co-authors including **Shubhashis Roy Dipta**]. “Mal-light: Enhancing lysine malonylation sites prediction problem using evolutionary-based features.” IEEE access

IEEE 2017 [2] Md Mainul Hasan Polash, [13 Co-authors including **Shubhashis Roy Dipta**]. “Explorer-0100: An autonomous next generation Mars rover.” 20th international conference of computer and information technology (ICIT) - Project Paper for The European Rover Challenge 2017, Poland (**Ranked 22nd**)

IEEE 2015 [1] Tanvir Ahmed Fuad, [13 Co-authors including **Shubhashis Roy Dipta**]. “MAYA: A fully functional rover designed for the mars surface.” 18th International Conference on Computer and Information Technology (ICIT) - Project Paper for The University Rover Challenge 2015, USA (**Ranked 9th**)

Research Projects

Jan 2023–Ongoing

Multimodal Object State Tracking & Generation

*Which of the **States** of the **Entities** will change? and **How**?*

Advised by Dr. Frank Ferraro

- Extending the previous SHEM^[8] work, we are exploring if the (noisy) image can be used as an external knowledge source or guide the model to generate state changes.
- To be robust to the noisy or missing images during testing, we have used controlled semi-supervision during training time.
- The model can be used on Search & Rescue (SAR) robots, where tracking state changes is crucial, even if the images are noisy or blurred.
- To make it feasible for edge devices, we are exploring small ($\leq 9B$) vision-language models for this task.

Jan 2023–Ongoing

Multimodal Counterfactual Event Understanding & Generation

*What will happen to the **Bike Race** if the **Weather** is **Stormy** instead of **Sunny**?*

Advised by Dr. Frank Ferraro

- Exploring a novel dataset generation technique that uses a state-of-the-art VLM model as a generator and humans as evaluators.
- Developing a novel dataset for counterfactual generation with text and images from real-life events.
- Exploring small ($\leq 9B$) open-source VLM models that can be used in edge devices for crucial situations where the machine has to understand the real-life situation, find out all alternate situations that can hinder the goal completion, and act accordingly, i.e., SAR robots.
- Present a comprehensive ablation on the importance of different modalities in predicting crucial events and generating alternative outcomes.
- Present a comprehensive analysis of the performance of current SOTA models (open vs. closed, small vs. large) in Counterfactual generation.

Jan 2021–Jan
2023

Semantically-informed Hierarchical Event Modeling & Abstraction^[8]

“Bill went to hospital, Doctors started treatment” – compressed down to “CURE”

Advised by Dr. Frank Ferraro

- The project leverages semantic frames and FrameNet relations to guide event representation and compression.
- Experimental results demonstrate the effectiveness of the hierarchical model in event modeling tasks, outperforming existing approaches and showing improvements in various event modeling tasks.
- The model incorporates structural and ontological hierarchy in event sequences, showcasing the importance of leveraging semantic ontologies for event modeling.
- This project highlights the benefits of structured and semantic hierarchical knowledge for event modeling.
- The proposed model, SHEM, is a doubly hierarchical, semi-supervised event modeling framework that outperforms previous state-of-the-art methods by up to 8.5% on two datasets and four evaluation metrics.

Jan 2023–Ongoing

Question-Answering using Semi-Supervised Graph Injection

*Can we answer questions **from a graph**? Or can we **induce the graph** even if the graph is absent?*

Advised by Dr. Frank Ferraro

- Text to (directional) graph generation and graph to question-answering.
- There are an abundant number of question-answering datasets, but question-answering datasets with induced graphs are scarce.
- With the above assumption, we have explored how to use the small-size graph dataset with semi-supervision so that our model can generate the graph event it's absent.
- This semi-supervision setting is important, as the graph datasets are small or sometimes mislabeled (noisy) in real life. This setting makes our model robust to those noisy & small datasets.

See more projects at the end of this CV (Page 5).

Academic Experience

- May 2021–Ongoing **Graduate Research Assistant**^[8,9,10], UMBC
Advisor: Dr. Frank Ferraro
- See (1) Multimodal Object State Tracking & Generation, (2) Multimodal Counterfactual Event Understanding & Generation, (3) Semantically-informed Hierarchical Event Modeling & Abstraction and (4) Question-Answering using Semi-Supervised Graph Injection in “Research Projects” Section (Page 2).
 - Mentoring: Supervising and providing research guidance to an undergraduate who is a member of an underrepresented group in CS.
- Jan 2021–May 2021 **Graduate Teaching Assistant**, UMBC
Course: Operating System
- Responsible for helping students, evaluating class projects, and final papers.
 - Designed kernel programming challenge as a class project.
- Jan 2019–Jan 2021 **Research Assistant**^[3,4,5,7], VOLUNTEER
Advisor: Dr. Iman Dehzangi
- See Language Model + Rotation Forest on Protein Sequences in “Project (cont’d)” Section.
 - Explored Language Models in Bio-informatics protein sequences (**Published 4 journals**).
- Jan 2014–Dec 2016 **Competitive Programming Trainer**, MIST
- Trainer of Competitive Programming in MIST Programming Club (2015–16).
 - Trainer of Data Structures & Algorithm in MIST Computer Club (2014–15).
 - Solved **3000+** problems in various online judges – 
- Jan 2015–Feb 2015 **Teaching Assistant**, MIST
- *Courses Taught:* C, C++, Data Structures, Algorithms

Industrial Experience

- June 2024–Aug 2024 **Incoming Research Intern**, SCALE.AI
- Will be working in the Vision-Language Team.
- Jan 2019–Jan 2021 **Founder & Chief Technology Officer**, UNISHOPR.COM
Amazon-like e-commerce site with cross-border shipping
- Single-handedly led a cross-functional team of 10, achieving **1,000+** active e-commerce users and **\$100,000+/month** in orders.
- Apr 2019–Jan 2021 **Full Stack Software Engineer**, SAPIEN.NETWORK
USA-based Startup, Worked on decentralizing our social experience data
- Implemented numerous backend services, i.e., Multi-factor Authentication, SSO, Caching (with Redis), Google ReCaptcha, Push Notification (iOS, Android, Web).
 - Implemented an Elixir-based scalable chat system.
- Oct 2018–Mar 2019 **Machine Learning Engineer**, BACKPACKBANG.COM
USA-based Y Combinator (YC) Startup
- Implemented search using a fusion of product and word embeddings.
 - Boosted sales by ~23% by developing a product recommendation system using Product2Vec embedding.
 - Engineered a Chatbot combining AI algorithms with logic-based if-else, decreasing response time by ~1 hour.
 - Decreased server cost by ~10% by implementing an AWS Lambda-based ML pipeline for online learning.
 - Build a model to predict the dimension of a product from the known datasets, which helped the traveler team allocate the luggage by ~20% more efficiently.
 - Implemented a microservice to refresh the inventory every 12 hours, which is later fetched by Facebook Ads, decreasing marketing labor cost by ~10%.
 - Worked with deployments on Google Cloud using Docker & Kubernetes.
 - Implemented crawler scripts to automatically curate data from Amazon, eBay, or other US-based e-commerce sites.
- Feb 2018–Sep 2018 **Frontend Engineer**, SELISE
- Implemented robust, scalable & reusable UI components that were used throughout the company projects, enabling fast production release.

Open-Source Contributions

PyTorch-Lightning
LeetCode

Academic Services

Reviewed **9 papers** at NLP/AI conferences and Bioinformatics journals.

- Mar 2024 The 18th International Workshop on Semantic Evaluation (SemEval)
- Jan 2024 Scientific Reports, Nature
- Jan 2024 Plant Methods, Nature
- Jan 2024 The 9th Workshop on Noisy and User-generated Text (W-NUT), EACL
- Dec 2023 Plant Methods, Nature
- May 2023 *SEM 2023, ACL (Secondary Reviewer)
- Mar 2023 Computational and Structural Biotechnology, ScienceDirect

Honors & Awards

- 2020 **70th** in Cornell Birdcall Identification, KAGGLE (**Top 6% - Bronze Medal**)
- 2020 Featured on the DocuSign Blog, DOCUSIGN
- 2018 **8th** in ACM ICPC Dhaka Regional, BANGLADESH (**Top 3%**)
- 2016 **13th** in ACM ICPC Dhaka Regional, BANGLADESH (**Top 10%**)
- 2016 **22nd** in European Rover Challenge^[2], POLAND (**Top 11%**)
- 2015 **9th** in European Rover Challenge^[2], POLAND (**Top 3%**)

Extra Curriculars

- 2023–Ongoing Maintains a Note-Garden [↗](#) based on ML, NLP, Research
- 2018 **Judge** in National High School & College Programming Contest, BANGLADESH
- 2018 **Judge** in MIST Intra Programming Contest, BANGLADESH
- 2014–2016 **Instructor** of Data-structures & Algorithms in MIST Computer Club, BANGLADESH

Leadership Experience

- 2016 **President** of MIST Programming Club, BANGLADESH
- 2015 **Head of Volunteer** of MIST Programming Club, BANGLADESH
- 2012 **Vice-President** in Notre Dame Science Fair, BANGLADESH
- 2011 **General Volunteer** in Notre Dame Science Fair, BANGLADESH

Skills

- Languages Python, SQL, Pandas, C/C++
- Frameworks PyTorch, PyTorch Lightning, Scikit-Learn, Huggingface, Streamlit, Gradio
- Visualization Matplotlib, Altair, Seaborn
- Web Dev. Flask, HTML/CSS, JavaScript, Node.JS, React
- Database PostgreSQL, MongoDB
- Utilities Conda, Git, Jupyter Notebook

Relevant Courses

- Graduate
- Introduction to Natural Language Processing (A+)
 - Introduction to Machine Learning (A+)
 - Principles of Artificial Intelligence (A+)
 - Knowledge Graph (A+)
 - Data Visualization (A+)
 - Design and Analysis of Algorithms (A+)
 - Crowd Sourcing & Computing (A+)
- Undergraduate
- Artificial Intelligence (A+)
 - Basic Graph Theory (A+)
 - Computer Graphics (A)
 - Pattern Recognition (A)
 - Numerical Analysis (A)

Projects (cont'd)

Machine Learning

- 2019–2020 **Language Model + Rotation Forest on Protein Sequences^[5]** 
- Are protein sequences as easily encoded by SOTA LLM models as regular text?*
- Developed a novel predictor named SEMal for predicting Malonylation sites in proteins.
 - SEMal combines structural & evolutionary features and protein sequence embedding to encode the features.
 - Outperformed existing SOTAs like kmal-sp, MaloPred, and LEMP in terms of sensitivity, specificity, precision, accuracy, F1-score, and Matthews Correlation Coefficient (MCC) for both Human and Mouse species.
 - Utilizes Rotation Forest (RoF) as the classifier on top of the encoded features.
 - The web server for SEMal is available online.
- 2020 **Kaggle - Cornell Birdcall Identification** 
- Bird chirping detection in complex soundscape recordings*
- **70th** among 1391 teams (**TOP 6% - Bronze Medal**).
 - Used EfficientNet on the spectrogram images with an ensemble of framed timespan.
- 2020 **Kaggle - Quora Insincere Questions Classification** 
- Identify and flag Insincere Questions on Quora Dataset*
- Combination of multiple word embedding with smart pre-processing to increase the coverage.
 - Explored the performance of the Convolutional Neural Network (CNN) on the language word embedding to exploit the local knowledge.
 - The local features output from CNN is used through LSTM to exploit the global context knowledge.
 - Achieved a F1-score of 0.68.

Data Science

- 2019 **SeeBel: Seeing is Believing^[6]** 
- A innovative way to visualize vision segmentation during training per step/epoch*
Advised by Dr. Rebecca Williams
- **Increased interpretability by ~60%** (user survey) in computer vision segmentation tasks by designing a real-time visualization tool for semantic segmentation.
 - Dataset statistics gives us the analysis of the dataset. On the other hand, AI model performance visualization gives us an idea of the model's capability.
 - But there is a gap between the input statistics and output visualization – the training time visualization.
 - This visualization tool bridges the gap between dataset statistics and AI model performance by visualizing the task during training.
- 2019 **Amazon-Crawler  & Search Engine **
- A cost-efficient Amazon crawler to refresh the product inventory on a daily basis*
- Designed a distributed web crawler using 200 Google Compute Engine instances to extract 1M products.
 - Explored and analyzed different cost-efficient and scalable strategies for 10M to 100M items.
 - **Enhanced the retrieval of 1M data** by implementing a resource-efficient search engine using Elastic-search.

2022 **N-Puzzle Solver** – **Poster** , **Report**

Compared various search strategies, identified optimal approaches based on minimal time complexity

Advised by Dr. Adam W. Bargteil

- Explored and analyzed various search strategies, including Uninformed, Informed, and Local Search, for solving N-Puzzle problems.
- Implemented algorithms such as Breadth-First Search, Depth-First Search, Dijkstra's algorithm, Best-First Search, A*, Iterative Deepening A*, and Hill Climbing Search.
- Evaluated strategies based on completeness, admissibility, time complexity, and space complexity.
- Conducted experiments on 8-puzzle and 15-puzzle problems with diverse initial configurations.
- Presented detailed analysis including path cost, time, and number of nodes expanded for each algorithm.
- Identified BFS, Dijkstra, A*, IDA*, and IDDFS as optimal and promising approaches for the 8-Puzzle problem based on minimal time complexity.

2013-2017 **A Large Collection of Algorithms**

A large collection of algorithm templates – implemented in C++ & Python

- This repository includes algorithm templates for various topics, such as graphs, dynamic programming, number theory, data structure, advanced search techniques, game theory, string, mathematics, and geometry.
- This repository served as the foundation template for numerous online and in-person programming contests I participated in & **3000+** problems I solved – 