Masum Hasan

PhD Student, Computer Science
University of Rochester, NY

🌐 https://masumhasan.net   in masum6   ✉ m.hasan@rochester.edu

RESEARCH INTERESTS

(LLM + ML) ∩ Game Engines = Conversational AI + Synthetic Data

TECHNICAL SKILLS

Languages: Python, SQL, Java, JavaScript, C++, C, Assembly, Microcontrollers

Technologies: LLMs, Diffusion, ViT, Vector Databases, PyTorch, TensorFlow, HuggingFace, AWS, Azure, 3D engines

PROFESSIONAL EXPERIENCE

University of Rochester   Rochester, NY
Graduate Research Assistant, ROC-HCI lab, Supervised by Prof. Ehsan Hoque   *Aug 2021 - Present*

- **SAPIEN:** *An Affective Virtual Agent Platform* Building and validating an LLM-driven anthropomorphic 3D avatar platform that can be used in a multitude of ways for skill development and education. Such as academic tutoring, language learning, training doctors on end-of-life communication, training therapists for PTSD patients, social interaction practice, job interviews, and others.
- **Hi5:** *Synthetic Dataset for Hand Pose Estimation* Using Unity 3D game engine we created a high-fidelity, realistic, diverse, and large synthetic dataset for human hand pose estimation.
- **AutoGait:** *Ataxia Risk Assessment using Computer Vision* Using object detection, pose estimation, and classification to predict the risk of ataxia from a video of a person walking. The classifier correlates with clinical knowledge without being explicitly programmed.

SAPIEN Coach, LLC   NY
Cofounder   *Mar 2023 - Present*

- Co-founded SAPIEN.Coach: an interactive, multilingual, emotionally responsive, virtual-human-based coaching platform for learning academic or soft skills, guided by our research.

Auburn University   Auburn, AL
TES Remote Research Assistant, Supervised by Prof. Shubhra (Santu) Karmaker   *Jun 2021 - Aug 2021*

- Semantic similarity in natural language, vector search, paraphrasing, Quora Question Pairs dataset.

Bangladesh University of Engineering and Technology (BUET)   Dhaka, Bangladesh
Research Assistant (RA), Applied Machine Learning Lab   *Nov 2018 - Jun 2021*

- Supervised by Prof. Rifat Shahriyar and Prof. Anindya Iqbal.
- **Machine learning for source code**: Created one of the largest Code-Natural Langauge parallel datasets at the time. Language models for source code Summarization/ Synthesis/ Search. Functional Android apps from Natural Language text. Automated program repair from code review comments using Neural Machine Translation.
- **NLP for low-resource Bengali language**: Trained a Bengali-GPT-2, helped create the largest dataset for Bengali-English Machine Translation.

EDUCATION

University of Rochester   Rochester, NY
Ph.D. Computer Science — Full Scholarship & Research Assistantship   *August 2021 - Present*

University of Rochester   Rochester, NY
M.Sc. Computer Science   *August 2021 - August 2023*

Bangladesh University of Engineering and Technology (BUET)   Dhaka, Bangladesh
B.Sc. Computer Science & Engineering   *February 2013 - October 2018*

Publications

- **SAPIEN: Affective Virtual Agents Powered by Large Language Models** – **Masum Hasan**, Cengiz Ozel, Sammy Potter, Ehsan Hoque. (ACIIW 2023)

- **Auto-Gait: Automatic Ataxia Risk Assessment with Computer Vision on Gait Task Videos** – Wasifur Rahman, **Masum Hasan**, Md Saiful Islam, Titilayo Olubajo, Jeet Thaker, Abdelrahman Abdelkader, Phillip Yang, Tetsuo Ashizawa, Ehsan Hoque. (UbiComp 2023)

- **CoDesc: A Large Code–Description Parallel Dataset** – **Masum Hasan**, Tanveer Muttaqeen, Ishtiaq Niloy, Kazi Mehrab, Tahmid Hasan, Mahim Pantho, Wasi Ahmad, Rifat Shahriyar, Anindya Iqbal. (ACL Findings 2021)

- **LowResource at BLP-2023 Task 2: Leveraging BanglaBert for Low Resource Sentiment Analysis of Bangla Language** – Aunabil Chakma, **Masum Hasan**. (EMNLP 2023, BLP Workshop)

- **A User-Centered Framework to Empower People with Parkinson's Disease** – Wasifur Rahman, Abdelrahman Abdelkader, Sangwu Lee, Phillip Yang, Md Saiful Islam, Tariq Adnan, **Masum Hasan**, Ellen Wagner, Sooyong Park, E. Ray Dorsey, Catherine Schwartz, Karen Jaffe, Ehsan Hoque. (IMWUT, UbiComp 2024)

- **Using a Balanced Scorecard to Identify Opportunities to Improve Code Review Effectiveness: An Industrial Experience Report** – **Masum Hasan**, Anindya Iqbal, Amiangshu Bosu, Mohammad Rafid Ul Islam, A.J.M. Imtiajur Rahman. (EMSE Journal)

- **Review4Repair: Code Review Aided Automatic Program Repairing** – Faria Huq, **Masum Hasan**, Mahim Pantho, Sazan Mahbub, Anindya Iqbal, Toufique Ahmed. (IST Journal)

- **Hitting your MARQ: Multimodal ARgument Quality Assessment** – Md Kamrul Hasan, James Spann, **Masum Hasan**, Md. Saiful Islam, Kurtis Haut, Rada Mihalcea, Ehsan Hoque. (EMNLP 2021)

- **Not Low-Resource Anymore: Aligner Ensembling, Batch Filtering, and New Datasets for Bengali-English Machine Translation** – Tahmid Hasan, Abhik Bhattacharjee, Kazi Samin, **Masum Hasan**, Madhusudan Basak, M. Sohel Rahman, Rifat Shahriyar. (EMNLP 2020)

- **Text2App: A Framework for Creating Android Apps from Text Descriptions** – **Masum Hasan**, Kazi Sajeed Mehrab, Wasi Uddin Ahmad, Rifat Shahriyar. (NLP4Prog, ACL-IJCNLP 2021)

- **Recognition of Bengali Handwritten Digits Using Convolutional Neural Network Architectures** – **Md Mahmudul Hasan**, Md Rafid Ul Islam, Md Tareq Mahmood. (ICBSLP 2018, **Best Student paper award**)

- **Early Detection of Earthquake Using Satellite Based Quantum Computing** – Akhter Al Amin, **Mahmudul Hasan**, Kazi Sinthia Kabir, Tanzila Choudhury, ABM Alim Al Islam. (ICCSNT, 2015)

Public Speaking

- (Upcoming talk, Feb 15, 2024) Future of Interactive Education with Generative AI at New York Academy of Sciences
- (Upcoming talk, Feb 21, 2024) Privacy, Security, Ethics of Generative AI, University of Rochester
- Radio appearance on WXXI News, "Connections with Evan Dawson" talking about the future of human-AI interactions, December 2023.
- Speaker at FlowerCity.AI conference, December 2023.
- Speaker on 4-part workshop series on Generative AI, Fall 2023 at University of Rochester.
- Conducted "Introduction to ChatGPT" workshop, Spring 2023 at University of Rochester.
- Created an introductory Machine Learning course in Bengali, 2021.

Relevant Courses

Artificial Intelligence: Frontier of Deep Learning, Knowledge Representation in AI, Introduction to Artificial Intelligence, Machine Learning, Deep Learning, Natural Language Processing Specialization, Computer Vision Specialization

Computer Science: Database Management Systems, Advanced Algorithms, Theory of Computation, Computational Modeling, Computer Organization. Object Oriented Programming, Structured Programming Languages, Operating Systems, Networking, Algorithms and Data Structures, Human Computer Interactions

Teaching Experience

University of Rochester Rochester, NY

Department of Computer Science : Graduate Teaching Assistant

- CSC 282/482 - **Advanced Algorithms** (Fall 2023)
- CSC 261/461 - **Database Systems** (Spring 2022)
- CSC 211/411 - **Human Computer Interaction**, (Spring 2022)

Extracurricular Activities

Audiobook addict, worst performer at my improv club, in a toxic relationship with my OneWheel, a sporadic rock climber, existential dread (not because of AI).