Syed Nazmus Sakib

Research Assistant, Cortex AI Lab | B.Sc. Student, University of Dhaka

Portfolio | Scholar | GitHub | LinkedIn | Mail

Research Interests

Multi-Modal AI · Agentic AI · Reinforcement Learning · Physical Intelligence · Robot Learning · Embodied AI · Vision–Language–Action Models · Multi-Agent Systems

Education

B.Sc. in Robotics and Mechatronics Engineering *Sep 2022 – Oct 2026 (expected)*
University of Dhaka Dhaka, Bangladesh

Research Experience

Research Assistant *Mar 2025 – Present*
CortexAI Lab, University of Dhaka Dhaka, Bangladesh

- First-authored research on multimodal reasoning and autonomous-agent safety, including papers in Findings of ACL 2026 and Scientific Data, and preprints under review at NeurIPS 2026.
- Built large-scale evaluation resources, including PlantInquiryVQA (138K QA pairs), PlantExpertVQA (193K QA pairs) and agentic moderation benchmarks.
- Trained MemeAgent, a 2B-parameter verifier reaching 87.8% in-distribution auditing accuracy and outperforming zero-shot frontier baselines.

Research Intern *Nov 2024 – Present*
Data and Design Lab, CARS, University of Dhaka Dhaka, Bangladesh

- Analyzed large-scale billing and distribution data for power-quality assessment in an industry-focused research project on power supply around Dhaka City.
- Developed machine learning pipelines for failure prediction, anomaly detection and load/demand forecasting from power-quality meter data.
- Contributed to an electricity-domain conversational assistant that answers billing, outage and power-quality queries in natural language.

R&D Engineer *Nov 2023 – Jul 2025*
Tech Topia Dhaka, Bangladesh

- Designed robotics and IoT automation systems using Arduino and ESP32, including hardware–software integration for industrial applications.

Selected Publications

Thinking Like a Botanist: Challenging Multimodal Language Models with Intent-Driven Chain-of-Inquiry
Syed Nazmus Sakib, Nafiul Haque, Shahrear Bin Amin, H. M. Abdullah, M. M. Hasan, Mohammad Zabed Hossain, Shifat E. Arman
***Accepted**, Findings of ACL 2026* [ACL Anthology] [arXiv]
Introduces PlantInquiryVQA with 24,964 expert-curated images and 138,078 QA pairs; structured inquiry improves diagnostic correctness and reduces hallucination in vision–language models.

Do Web Agents Investigate Before They Decide?
Syed Nazmus Sakib, Nafiul Haque, Tapodhir Karmakar Taton, Shahrear Bin Amin, Shifat E. Arman
Preprint [arXiv]
Benchmark of 750 multi-step tasks isolating investigative competence across eight LLM agents; reveals navigation–discovery gaps, contradiction failures and investigative hallucination.

MemeEconomy: Do LLM Agents Trade Ethics for Survival?
Syed Nazmus Sakib, Nafiul Haque, Ahnaf Tahmid Manan, M. M. Mahabub Morshed, Shifat E. Arman
***Preprint** · Under Review, NeurIPS 2026*
Multimodal multi-agent market simulation across 300 events; develops a compact verifier for auditing harmful agent behavior under survival pressure.

PlantExpertVQA: A Visual Question Answering Dataset for Benchmarking Vision–Language Models in Plant Science
Syed Nazmus Sakib, Nafiul Haque, Mohammad Zabed Hossain, Shifat E. Arman
***Published**, Scientific Data (Nature), 2026* [Nature] [arXiv]
Large-scale agricultural VQA resource containing 193,609 QA pairs over 55,448 images, 14 crop species and 38 disease conditions.

The Surface You Test Is Not the Surface That Breaks
Syed Nazmus Sakib, Nafiul Haque, Shahrear Bin Amin, Shifat E. Arman
Preprint [arXiv] [code]
Shows that prompt-injection vulnerability is a model-by-surface interaction; a surface-adaptive attacker improves attack success across 13 language models.

PhyDrawGen: Physically Grounded Diagram Generation from Natural Language
Nafiul Haque, **Syed Nazmus Sakib**†, Shifat E. Arman
Preprint [arXiv]
Neuro-symbolic system coupling language-based scene extraction with deterministic geometric and physical constraint solving over 1,449 diagram-generation problems.

SugarcaneShuffleNet: A Very Fast, Lightweight CNN for Diagnosis of 15 Sugarcane Leaf Diseases
Shifat E. Arman, H. M. Abdullah, **Syed Nazmus Sakib**, R. M. Saiem, S. N. Asha, M. M. Hasan, et al.
Under Review*, Computers and Electronics in Agriculture* [arXiv]
Lightweight visual model reaching 98.02% accuracy at 4.14 ms per image for deployment on resource-constrained devices.

† Equal contribution (co-first author). Complete publication list: Google Scholar.

Research Projects

Long-Horizon Mobile Manipulation and Robust Skill Composition *2026 – Present*
Independent research using ManiSkill / MS-HAB, PyTorch and simulation-based evaluation

- Reproduced and evaluated a long-horizon TidyHouse pipeline using 21 pretrained Pick, Place and Navigate policies, with strict success, reachability and failure-stage diagnostics.
- Designed controlled handoff experiments showing that robot-state mismatch, rather than object-state mismatch, dominated chained Pick→Place degradation; localized major physical failures to base/torso contacts and force-limit trips.
- Improved full-chain success from 0.7% to 2.2% by tuning global base damping and force limits, validated over three seeds with confidence intervals, per-skill analysis and representative rollout videos.

Mobile Differential-Drive Robot *2025*
ROS 2, Gazebo, Xacro, URDF and autonomous exploration GitHub

- Built a complete ROS 2 simulation package covering robot modeling, differential-drive control, sensor integration and autonomous exploration in Gazebo.

Hexapod Robot *2024*
C++, CMake, inverse kinematics and gait generation GitHub

- Implemented inverse-kinematics-based leg control, coordinated gait generation and motion planning for a six-legged robotic platform.

Technical Skills

Programming	Python, C++, SQL, JavaScript, TypeScript
Agentic AI	LLM agents, tool use, multi-agent simulation, prompt-injection robustness, benchmark design
Robot Learning	PyTorch, Gymnasium, ManiSkill / MS-HAB, reinforcement learning, policy evaluation
Simulation & Robotics	Isaac Sim, Isaac Lab, MuJoCo, ROS 2, Gazebo, URDF, Xacro, kinematics, motion planning
Perception & ML	OpenCV, Transformers, vision–language models, TensorFlow, Scikit-learn
Hardware & Engineering	Arduino, ESP32, Autodesk Fusion 360, Docker, Git, Linux, MLflow

Honors & Awards

- **Global Nominee,** NASA Space Apps Challenge 2024
- **Runner-up,** KUET Datathon 2025; **Runner-up,** DU AI Challenge 2025; **Runner-up,** Technocrats V2 IUBAT Hackathon 2024
- **Regional Champion,** National High School Programming Contest 2019
- **Kaggle Expert** with multiple competition podium finishes