Alvi Ishmam

alvi@vt.edu | 5404492513 | LinkedIn | alvi.me | Google Scholar | Research Gate

EDUCATION

VIRGINIA TECH
PhD in Computer Science
January 2021 - 2025 (December)
M.Sc. in Computer Science
January 2021 - November 2023
GPA: 3.92 / 4.0

SKILLS

LANGUAGES
Python • C# •Java •C++ •SQL •Assembly •LaTeX

TOOLS AND FRAMEWORKS
HTML • CSS •Angular • React • Weblogic • Oracle• Apache Tomcat • RESTful API •Spring MVC • Spring Boot • Android • Oracle ADF • .Net Core • Scikitlearn • WEKA • Facebook Graph API

MACHINE LEARNING FRAMEWORKS
Pytorch • Keras • Tensorflow

PUBLICATION

(check out google scholar for complete list)

- Semantic Shield: Defending Vision-Language Models Against Backdooring and Poisoning via Fine-grained Knowledge Alignment **[CVPR'24]**
- Learning Universal Adversarial Perturbations for Multi Image Tasks via Pretrained Models **[AAAI'26]**
- M3D: MultiModal MultiDocument Fine-Grained Inconsistency Detection **[EMNLP'24]**
- JourneyBench: A Challenging One-Stop Vision-Language Understanding Benchmark of Generated Images **[Neurips'24]**
- Hateful Speech Detection in Public Facebook Pages for the Bengali Language. **[ICMLA'19][Asian CHI'19]**

EXPERIENCE

FUTUREWEI TECHNOLOGIES | Research Intern
May 2025 – Aug 2025 | Framingham, MA, USA

- Develop self evolving RAG system with a focus on advanced storage system/data abstraction process.
- Design reference free evaluation method for hallucination mitigation for enterprise. **NeuripsW '25**

GE HEALTHCARE | AI/ML PhD Intern
May 2024 – Aug 2024 | San Ramon, CA, USA

- Generate the largest medical image-text pairs **2M** from modalities CT, MR, US.
- Design the largest medical **CLIP** style model achieving SOTA performance.

PROJECTS

DEFENDING VISION-LANGUAGE MODEL AGAINST IMAGE/TEXT POISONING ATTACK | [CVPR2024]

- The goal is to design a robust multimodal model to defend against any sort of data poisoning attack that involved image. (Foundation models, e.g., Vicuna, LLaMa, Multimodal models)
- Proposed a robust and trustworthy finetuning strategy that can defend vision-language model (CLIP) against any image based backdoor and text based poisoning attack.

LEARNING UNIVERSAL ADVERSARIAL PERTURBATIONS FOR MULTI IMAGE TASKS VIA PRETRAINED MODELS | [AAAI'26]

- We propose the first adversarial attack for multi-image multi-model models by learning universal adversarial perturbations.
- Our method increases the attack success rate by 20% in various multi-image tasks across MLLMs compared to SOTA.

MULTIMODAL DISINFORMATION DETECTOR USING FINE GRAINED VISUAL ENTAILMENT | [EMNLP'24]

- Visual Entailment is a reasoning task where the logical relationship between text, image, and video is predicted.
- The goal of the project is to determine disinformation in multimodal news articles by representing the semantic inconsistency in knowledge elements in text, images/videos.

PROBING AND ROBUSTNESS ANALYSIS OF BROWSER-BASED WEB AGENTS

- Proposed adaptive vulnerabilities in multimodal web agents.
- Threat models (white-box and black-box) incorporating stealth, controllability, and delayed triggers to systematically probe agent decision-making under adversarial conditions.

AUDIO ATTACKS AGAINST VIDEO AUDIO MODELS

- Imperceptible audio attack to video audio models.
- Optimized audio signal to attack the video audio model.