Connecticut AI Alliance (CAIA)
AI Computing Access Program

Sample Applications: Faculty & Graduate Research Support Track

Award Period: Summer 2026 – Fall 2026

Note to Applicants: *The following two sample applications are provided to illustrate the level of detail and specificity expected in a competitive submission to the Faculty & Graduate Research Support track. All names, institutions, and contact information are fictional. These samples represent common application scenarios and are not intended to be copied verbatim. Reviewers value clear articulation of scientific merit, realistic computational justifications, and concrete plans for graduate student development and dissemination. Note that applicants need not be in computer science departments; this track welcomes AI-enabled research from any discipline.*

Sample Application 1: AI-Enabled Public Health Research (Non-CS Department)

Scenario: *A public health faculty member and her MPH/PhD students are using deep learning to forecast syndromic surveillance signals from emergency department data across Connecticut. The PI has no institutional GPU access and relies on this program to train spatiotemporal models on a large, multi-year dataset. This sample illustrates how applicants outside of computer science can present a compelling, technically grounded application.*

Section 1: Principal Investigator (PI) Information

1. Full Name
Dr. Amina Hassan

2. Email Address
ahassan@northshore.edu

3. CAIA Member Institution
Northshore University

4. Department / Program
Department of Epidemiology and Biostatistics, School of Public Health

5. Title / Rank
Associate Professor

6. Phone Number
(860) 555-0319

7. Personal Website or Google Scholar Profile URL
https://scholar.google.com/citations?user=EXAMPLE_ID (fictional)

Section 2: Research Team

8. Co-PI(s)
Dr. Samuel Kwon, Assistant Professor, Department of Computer Science, Northshore University (skwon@northshore.edu). Dr. Kwon provides technical guidance on deep learning architectures and

serves as methods co-advisor for the PhD student on this project.

9. Graduate Student Researchers

Maria Torres, PhD in Epidemiology (3rd year). Maria is the lead researcher and will use GPU resources for all model training and evaluation. This project constitutes Chapters 2 and 3 of her dissertation. Daniel Olsen, MPH candidate (2nd year). Daniel will assist with data preprocessing and geospatial feature engineering as part of his practicum requirement.

10. Total number of personnel who will actively use the computing resources

3 (PI, one PhD student, one MPH student). Dr. Kwon will advise but will not directly use the resources.

Section 3: Research Project Description

11. Project Title

Spatiotemporal Deep Learning for Early Detection of Respiratory Illness Surges in Connecticut Emergency Departments

12. Research Area(s)

- ■ Natural Language Processing / Large Language Models
- ■ Computer Vision / Image Processing
- ■ Reinforcement Learning / Multi-Agent Systems
- ■ AI Safety, Security, & Adversarial ML
- ■ Healthcare / Biomedical AI
- ■ Robotics / Autonomous Systems
- ■ AI for Science (climate, materials, physics, etc.)
- ■ AI for Education / EdTech
- ■ Generative AI / Foundation Models
- ■ Graph Neural Networks / Knowledge Graphs
- ■ Fairness, Accountability, & Transparency
- ■ Other

13. Project Abstract

Timely detection of emerging respiratory illness outbreaks is critical for public health preparedness, yet current syndromic surveillance systems in Connecticut rely primarily on rule-based thresholds applied to weekly aggregated counts, which often produce alerts too late for effective intervention. This project develops a deep learning framework that integrates daily emergency department (ED) chief complaint data, environmental covariates (temperature, humidity, air quality index), and spatial connectivity between hospital catchment areas to produce 7-day and 14-day forecasts of respiratory syndrome volumes at the hospital level across Connecticut.

Our approach combines graph neural networks (GNNs) to model spatial dependencies between ED facilities with temporal convolutional networks (TCNs) to capture seasonal and short-term trends in syndrome counts. The model is trained on six years of de-identified ED visit records (2018–2024) from 28 Connecticut hospitals, comprising approximately 4.2 million encounters with syndromic classification labels. We construct a hospital connectivity graph weighted by patient flow patterns and geographic proximity, then train a spatiotemporal GNN that jointly forecasts syndrome volumes across all nodes.

Key research questions include: (1) Does the spatiotemporal GNN architecture outperform site-independent baselines (ARIMA, single-site LSTMs) in forecast accuracy, particularly for early surge detection? (2) How does model performance vary across urban, suburban, and rural hospital settings, and what are the equity implications of differential accuracy? (3) Can an LLM-generated natural language summary layer make model outputs interpretable to non-technical public health decision-makers?

This research directly supports Connecticut's public health infrastructure and aligns with CDC priorities for modernized biosurveillance. Results will inform the design of an operational early warning dashboard for the Connecticut Department of Public Health.

14. How are cloud computing resources essential to this research?

Northshore University's institutional computing cluster is a shared, CPU-only resource managed by the IT department, with a maximum per-user allocation of 8 CPU cores and 32 GB RAM. Training our spatiotemporal GNN on the full 4.2 million encounter dataset with graph convolution operations requires GPU acceleration; a single training epoch takes approximately 4 hours on a T4 GPU but would take an estimated 40+ hours on the available CPU cluster, making hyperparameter search and ablation studies impractical. Our complete experimental plan (described in Section 4) requires approximately 500–600 GPU-hours over the award period. Additionally, the LLM summarization component requires API access to evaluate multiple prompt strategies for translating model outputs into actionable public health language. Without cloud GPU and API access, this dissertation research cannot proceed on its current timeline, and the PhD student's expected Spring 2027 defense would be delayed by at least one year.

15. Current Status of the Research

- ■ New project – not yet started
- ■ Early stage – preliminary results or pilot data collected
- ■ In progress – active experiments requiring additional/scaled compute
- ■ Thesis/Dissertation research – graduate student milestone work

16. Expected Project Timeline

- ■ Summer 2026 only
- ■ Fall 2026 only
- ■ Both Summer + Fall 2026

17. Target Venues for Dissemination

AMIA Annual Symposium (November 2026, abstract submission); Machine Learning for Health (ML4H) Workshop at NeurIPS 2026; American Journal of Public Health (manuscript, Spring 2027); CAIA Research Symposium 2026

Section 4: Computing Resource Requirements

18. What type(s) of computing resources are you requesting?

- ■ GPU Computing – Google Colab Pro/Pro+
- ■ GPU Computing – JupyterHub (managed cluster)
- ■ LLM API Access – Google Vertex AI

19. Describe your computational workloads in detail.

Our primary workload is training a spatiotemporal graph neural network that operates on a hospital connectivity graph of 28 nodes, with each node receiving a daily feature vector of 18 dimensions (syndrome counts by category, environmental covariates, day-of-week indicators). The full training set covers 2,192 days (2018–2024) and the model jointly forecasts 7-day and 14-day syndrome volumes across all hospitals.

Architecture: The model combines a 2-layer Graph Attention Network (GAT) for spatial message passing with a Temporal Convolutional Network (TCN) backbone (6 residual blocks, dilated convolutions). Total parameter count is approximately 3.8M. We also evaluate two baselines: a site-independent LSTM (1.2M parameters per site) and a simple spatial mean pooling variant.

Training details: Batch size 32 (batched over temporal windows), trained for 200 epochs with early stopping. A single training run takes approximately 2.5 hours on a T4 GPU. We plan 5-fold temporal cross-validation (5 runs per configuration), an ablation study removing spatial, temporal, and environmental components (4 ablation variants x 5 folds = 20 runs), a hyperparameter sweep over learning rate, GAT heads, and TCN depth (30 configurations x 1 fold for screening, top 5 x 5 folds for full evaluation = 55 runs), and baseline model training (2 baselines x 5 folds = 10 runs). Total estimated GPU-hours: approximately 225–250 hours for the core experiments, plus 50–75 hours for debugging, data pipeline validation, and inference evaluation. All workloads fit within T4 (16 GB VRAM); A100 is not required. Peak GPU usage is expected in July–August (main experimental runs) and November (final ablation and revision experiments).

20. GPU Hardware Preferences

- ■ T4 (16 GB VRAM) – suitable for inference, small model training
- ■ A100 (40 GB VRAM) – suitable for large model training, fine-tuning
- ■ A100 (80 GB VRAM) – required for very large models, multi-GPU training
- ■ No specific preference / flexible

21a. Which models do you plan to use?

Gemini Pro for generating natural language summaries of model predictions. We will evaluate whether LLM-generated surveillance briefs are interpretable and actionable for public health officials compared to raw numerical outputs.

21b. Estimated monthly usage

Approximately 800K–1.2M input tokens and 300K–500K output tokens per month, concentrated in Fall 2026 during the interpretability evaluation phase.

21c. Research use case

Prompt engineering experiments to translate spatiotemporal model forecasts into structured surveillance briefs (e.g., "Hospital X is projected to see a 35% increase in respiratory presentations over the next 7 days; recommend activating surge protocols"). We will evaluate multiple prompt strategies using a rubric scored by public health practitioners.

22. Do you require persistent storage for datasets or model checkpoints?

- ■ Yes – approximately 50 GB (de-identified ED dataset, preprocessed graph features, model checkpoints for reproducibility)
- ■ No

23. Estimated total cloud credit budget requested

$6,500 ($4,500 for GPU compute at approximately 300–325 GPU-hours on T4 instances; $2,000 for Vertex AI API credits for the LLM interpretability study)

Section 5: Research Impact & Outcomes

24. What are the expected research outcomes?

- ■ Peer-reviewed publication(s)
- ■ Conference presentation or poster
- ■ Thesis or dissertation chapter(s)
- ■ Open-source software, tools, or library
- ■ Public dataset or benchmark release
- ■ Patent or invention disclosure
- ■ Grant proposal (using results as preliminary data)

■ Presentation at CAIA Research Symposium

■ Other

25. How does this project contribute to graduate student development?

This project is central to Maria Torres's dissertation. The computing resources will enable her to complete Chapters 2 (spatiotemporal model development and evaluation) and 3 (LLM interpretability study) and defend by Spring 2027. Through this work, Maria will develop expertise in graph neural networks, temporal forecasting, and responsible deployment of AI in public health settings. She will gain practical experience with cloud computing workflows and cost management. Daniel Olsen's MPH practicum contribution (geospatial feature engineering) will satisfy his applied practice requirement and expose him to research-grade data science practices. Both students will present findings at the CAIA Research Symposium and will be co-authors on resulting publications.

26. Does this project involve collaboration with other CAIA institutions or external partners?

■ Yes – another CAIA institution

■ Yes – industry partner

■ Yes – national lab, government, or nonprofit

■ No

The PI has an active data use agreement with the Connecticut Department of Public Health (CT DPH), which provides the de-identified emergency department syndromic surveillance dataset used in this project. A CT DPH epidemiologist serves on the dissertation committee as an external member and will participate in the LLM interpretability evaluation as a domain expert reviewer.

27. Will this research generate preliminary results for future grant applications?

■ Yes – NSF Smart and Connected Health (SCH) program. The PI plans to submit a full proposal in Spring 2027 using results from this project as preliminary data. Additionally, the CT DPH partnership may support a CDC Epidemiology and Laboratory Capacity (ELC) cooperative agreement application.

■ No / Not yet determined

Section 6: Responsible AI & Data Practices

28. Does your research involve any of the following?

■ Human subjects data (requires IRB approval)

■ Protected health information (PHI / HIPAA)

■ Personally identifiable information (PII)

■ Biometric data

■ Synthetic media generation

■ Dual-use AI research

■ None of the above

29. Ethical safeguards and institutional approvals

The ED visit dataset is provided by CT DPH under a data use agreement (DUA #2023-0847, renewed annually). All records are de-identified in accordance with HIPAA Safe Harbor provisions prior to transfer: no names, dates of birth, addresses, or medical record numbers are included. Hospital identifiers are replaced with coded IDs. The Northshore University IRB reviewed the project protocol and issued a determination of "not human subjects research" (IRB #2024-0312) on the basis that the data are fully de-identified and not individually linkable. Despite this determination, the PII box is checked above because the dataset includes visit dates and zip-code-level geography, which could theoretically enable re-identification in combination with external data sources. Accordingly, the data are stored on an encrypted, access-controlled institutional server and are never uploaded to public repositories.

30. Data management and security plan

The de-identified dataset is stored on Northshore University's encrypted research server with access restricted to the PI and named graduate students. For cloud-based training, preprocessed feature matrices (which contain no individual-level records) are uploaded to a private Google Cloud Storage bucket with encryption at rest and in transit. Raw de-identified data never leaves the institutional server. Model checkpoints and code are version-controlled in a private GitHub repository. API keys are stored using Google Secret Manager and are never committed to version control. Upon project completion, cloud storage will be purged and a local archival copy retained per university data retention policy.

Section 7: Prior Support & Related Funding

31. Have you previously received computing resources through CAIA?

- ■ Yes
- ■ No

32. List any active grants or funding that support this research.

NIH/NIGMS T32 Training Grant (T32-GM123456, "Training in Computational Epidemiology," 2023–2028, PI: Dr. R. Chen). Maria Torres is supported as a T32 trainee; the training grant covers her stipend and tuition but does not include computing resources. No other external funding directly supports this project.

33. Is this project currently supported by any cloud computing credits from another source?

- ■ Yes
- ■ No

Section 8: Supporting Materials (Optional)

34. CV / Biosketch

Uploaded (2-page NIH biosketch format).

35. Supplementary Material

Uploaded: 2-page supplement containing (1) preliminary results from a single-site LSTM baseline trained on a 10% data sample using CPU resources, demonstrating feasibility, and (2) a system architecture diagram of the proposed spatiotemporal GNN pipeline.

Section 9: Applicant Certification

36. Electronic Signature

Amina Hassan

37. Date

May 8, 2026

Sample Application 2: PhD Dissertation Research in Adversarial ML / AI Safety

Scenario: *A CS faculty member requests GPU resources to support a PhD student's dissertation on adversarial robustness of vision-language models. The project requires A100 GPUs for fine-tuning large multimodal models and red-teaming experiments. This sample illustrates a compute-intensive request with clear hardware justification, a strong dissemination plan, and use of results as preliminary data for a federal grant proposal.*

Section 1: Principal Investigator (PI) Information

1. Full Name

Dr. Yuna Park

2. Email Address

ypark@cedarhill.edu

3. CAIA Member Institution

Cedar Hill University

4. Department / Program

Department of Computer Science

5. Title / Rank

Assistant Professor

6. Phone Number

(203) 555-0461

7. Personal Website or Google Scholar Profile URL

https://yunapark.github.io (fictional)

Section 2: Research Team

8. Co-PI(s)

None.

9. Graduate Student Researchers

Rakesh Iyer, PhD in Computer Science (2nd year). Rakesh is the primary researcher and will conduct all fine-tuning, adversarial evaluation, and defense experiments. This project forms the core of his dissertation (expected defense: Spring 2028). Sofia Chen, MS in Data Science (1st year). Sofia will contribute to dataset curation and automated red-teaming pipeline development as part of her thesis research.

10. Total number of personnel who will actively use the computing resources

3 (PI and two graduate students)

Section 3: Research Project Description

11. Project Title

Evaluating and Improving Adversarial Robustness of Vision-Language Models Under Multimodal Perturbations

12. Research Area(s)

- ■ Natural Language Processing / Large Language Models
- ■ Computer Vision / Image Processing

■ Reinforcement Learning / Multi-Agent Systems
■ AI Safety, Security, & Adversarial ML
■ Healthcare / Biomedical AI
■ Robotics / Autonomous Systems
■ AI for Science
■ AI for Education / EdTech
■ Generative AI / Foundation Models
■ Graph Neural Networks / Knowledge Graphs
■ Fairness, Accountability, & Transparency
■ Other

13. Project Abstract

Vision-language models (VLMs) such as LLaVA, Qwen-VL, and GPT-4V are rapidly being deployed in safety-critical applications including medical image interpretation, autonomous driving scene understanding, and content moderation. However, the adversarial robustness of these multimodal systems remains poorly understood compared to their unimodal counterparts. Prior work on adversarial attacks has largely focused on single-modality perturbations (e.g., image-only adversarial patches or text-only prompt injections), while real-world threat models increasingly involve coordinated multimodal attacks.

This project investigates three interrelated research questions. First, we develop a taxonomy and benchmark of multimodal adversarial attacks on VLMs, including joint image-text perturbations, cross-modal transfer attacks (where perturbations in one modality exploit vulnerabilities in the other's processing pathway), and semantic-preserving attacks that evade perceptual detection. Second, we conduct a systematic robustness evaluation of five open-weight VLMs of varying scale (2B to 13B parameters) under our proposed attack suite, quantifying the gap between single-modality and multimodal attack success rates. Third, we propose and evaluate a defense framework based on adversarial fine-tuning with multimodal perturbation augmentation, combined with a lightweight input verification module that detects inconsistencies between visual and textual inputs.

The significance of this work lies in establishing principled evaluation methodology for VLM safety that accounts for the multimodal attack surface, and in producing open-source tools that the community can use to red-team and harden VLMs before deployment. This research contributes directly to the growing body of work on AI safety and aligns with NIST AI Risk Management Framework priorities for adversarial testing of foundation models.

14. How are cloud computing resources essential to this research?

Cedar Hill University's on-campus GPU cluster consists of four NVIDIA RTX 3090 cards (24 GB VRAM each), shared across the entire CS department. Fine-tuning a 7B-parameter VLM using LoRA requires at minimum 40 GB VRAM (A100 40 GB), which exceeds the capacity of our institutional hardware. Full adversarial fine-tuning of 13B-parameter models requires 80 GB VRAM. Furthermore, the adversarial evaluation pipeline involves running inference on thousands of perturbed image-text pairs across five model variants, generating substantial throughput requirements that would monopolize departmental hardware for weeks. Cloud A100 access through CAIA would allow us to conduct these experiments in parallel without disrupting other research groups, and would be the enabling resource for this dissertation.

15. Current Status of the Research

■ New project – not yet started
■ Early stage – preliminary results or pilot data collected
■ In progress – active experiments requiring additional/scaled compute
■ Thesis/Dissertation research – graduate student milestone work

16. Expected Project Timeline

- ■ Summer 2026 only
- ■ Fall 2026 only
- ■ Both Summer + Fall 2026

17. Target Venues for Dissemination

USENIX Security 2027 (submission deadline: late January 2027); IEEE Symposium on Security and Privacy (S&P) 2027; NeurIPS 2026 Workshop on Adversarial Robustness (if announced); CAIA Research Symposium 2026; IEEE Transactions on Information Forensics and Security (journal manuscript, Spring 2027)

Section 4: Computing Resource Requirements

18. What type(s) of computing resources are you requesting?

- ■ GPU Computing – Google Colab Pro/Pro+
- ■ GPU Computing – JupyterHub (managed cluster)
- ■ LLM API Access – Google Vertex AI

19. Describe your computational workloads in detail.

This project has three main computational phases.

Phase 1: Attack Generation and Benchmark Construction (Summer 2026). We will generate multimodal adversarial examples against five open-weight VLMs: LLaVA-1.6 (7B and 13B), Qwen-VL (7B), InternVL (6B), and CogVLM (17B). Our attack suite includes six attack types (gradient-based image perturbation, typographic injection, cross-modal transfer, joint image-text optimization, semantic-preserving paraphrase + patch, and instruction injection). For each model-attack combination (30 total), we generate adversarial examples on 2,000 image-text pairs sampled from VQAv2 and COCO Captions, yielding 60,000 perturbed samples. Attack generation for gradient-based methods requires forward and backward passes through the target VLM; for the 7B models this takes approximately 3 seconds per sample on an A100 40 GB, and for the 13B/17B models approximately 6 seconds on an A100 80 GB. Estimated Phase 1 GPU-hours: 200–250 hours (A100 40 GB) + 80–100 hours (A100 80 GB).

Phase 2: Robustness Evaluation (Summer–Fall 2026). Run all five VLMs against the full adversarial benchmark, compute attack success rates, and analyze cross-modal transfer patterns. This phase is inference-heavy: approximately 300,000 model inferences across 5 models. Estimated GPU-hours: 150–180 hours (A100 40 GB).

Phase 3: Defense Development and Evaluation (Fall 2026). Fine-tune LLaVA-7B and Qwen-VL-7B using LoRA (rank 16) with our adversarial augmentation strategy, then re-evaluate robustness. Each LoRA fine-tuning run requires approximately 8 hours on an A100 40 GB. We plan 10 fine-tuning runs per model (hyperparameter variations) plus 5 ablation configurations. Estimated GPU-hours: 200–250 hours (A100 40 GB). Total estimated GPU-hours across all phases: approximately 650–780 on A100 40 GB and 80–100 on A100 80 GB.

20. GPU Hardware Preferences

- ■ T4 (16 GB VRAM)
- ■ A100 (40 GB VRAM) – primary requirement for most workloads
- ■ A100 (80 GB VRAM) – required for 13B/17B model attack generation only
- ■ No specific preference / flexible

21a. Which models do you plan to use?

Gemini Pro and Gemini Ultra. We will include Gemini models as closed-source comparison points in our robustness evaluation (black-box attacks only) and use Gemini Pro for automated evaluation of attack output quality (e.g., scoring whether adversarial responses are coherent, harmful, or off-topic).

21b. Estimated monthly usage

Approximately 3–4M input tokens and 1–1.5M output tokens per month during the evaluation phases (July–November). Usage will be lower in June (~500K input tokens) during pipeline development.

21c. Research use case

Black-box adversarial robustness evaluation of closed-source VLMs (Gemini); automated evaluation of adversarial output quality using LLM-as-judge methodology; and systematic red-teaming experiments to compare attack transferability between open-weight and closed-source models.

22. Do you require persistent storage for datasets or model checkpoints?

■ Yes – approximately 200 GB (adversarial benchmark dataset with 60K perturbed image-text pairs; LoRA adapter checkpoints for fine-tuned defense models; evaluation logs and intermediate results for reproducibility)
■ No

23. Estimated total cloud credit budget requested

$18,000 ($14,000 for GPU compute: approximately 550–680 hours on A100 40 GB at ~$1.50/hr and 80–100 hours on A100 80 GB at ~$2.50/hr, with buffer for debugging and reruns; $3,000 for Vertex AI API credits; $1,000 for cloud storage)

Section 5: Research Impact & Outcomes

24. What are the expected research outcomes?

■ Peer-reviewed publication(s)
■ Conference presentation or poster
■ Thesis or dissertation chapter(s)
■ Open-source software, tools, or library
■ Public dataset or benchmark release
■ Patent or invention disclosure
■ Grant proposal (using results as preliminary data)
■ Presentation at CAIA Research Symposium
■ Other

25. How does this project contribute to graduate student development?

Rakesh Iyer's dissertation will consist of three papers corresponding to the three research phases described above. The CAIA computing resources are essential to producing the experimental results for all three papers. Through this project, Rakesh will develop deep expertise in adversarial machine learning, multimodal model architectures, and responsible red-teaming methodology. He will gain experience with large-scale distributed GPU computing and cloud cost optimization. Sofia Chen's MS thesis will focus on the automated red-teaming pipeline (Phase 1), giving her a publishable contribution and practical experience in ML security research that will strengthen her applications for PhD programs or industry research positions.

26. Does this project involve collaboration with other CAIA institutions or external partners?

■ Yes – another CAIA institution
■ Yes – industry partner
■ Yes – national lab, government, or nonprofit
■ No

Dr. Marcus Webb, Associate Professor of Computer Science at Eastgate College (a CAIA member institution), collaborates on the project in an advisory capacity. Dr. Webb's group has developed text-only adversarial attack methods that we are extending to the multimodal setting. Dr. Webb's PhD student, Lena Volkov, will contribute to the prompt injection and typographic attack components and will be a co-author on the resulting publication.

27. Will this research generate preliminary results for future grant applications?

■ Yes – NSF CAREER (SaTC program, submission deadline July 2027). The PI plans to use the adversarial benchmark and defense framework as preliminary results in a CAREER proposal on Trustworthy Multimodal AI. Additionally, the PI intends to submit to the NSF CISE Core Small program in Fall 2027 with Dr. Webb as co-PI.

■ No / Not yet determined

Section 6: Responsible AI & Data Practices

28. Does your research involve any of the following?

■ Human subjects data (requires IRB approval)

■ Protected health information (PHI / HIPAA)

■ Personally identifiable information (PII)

■ Biometric data

■ Synthetic media generation

■ Dual-use AI research (security, surveillance, autonomous weapons)

■ None of the above

29. Ethical safeguards and institutional approvals

This research involves generating adversarial examples that could, in principle, be used to attack deployed VLMs. We treat this work under a responsible disclosure framework consistent with standard practice in the computer security research community. Specifically: (1) All adversarial experiments are conducted against locally hosted open-weight models, not against production APIs (the Gemini evaluation uses only standard API queries, not adversarial exploits targeting API infrastructure). (2) The adversarial benchmark dataset will be released under a research-only license with a responsible use agreement. (3) Any vulnerabilities discovered in closed-source models during black-box evaluation will be reported to the model provider before public disclosure, following a 90-day responsible disclosure window. (4) The Cedar Hill University IRB has determined this project does not involve human subjects (IRB exemption #2025-0198).

30. Data management and security plan

All experiments use publicly available datasets (VQAv2, COCO Captions) and open-weight models. Generated adversarial examples and model checkpoints are stored in a private cloud storage bucket with project-level access controls; only the PI and named graduate students have access. API keys are managed through Google Secret Manager. The adversarial benchmark will be hosted on Hugging Face under a gated access model requiring agreement to responsible use terms. Code and evaluation scripts will be released on GitHub with an MIT license. The project follows Cedar Hill University's Research Data Management Policy and CAIA's Acceptable Use Policy.

Section 7: Prior Support & Related Funding

31. Have you previously received computing resources through CAIA?

■ Yes

■ No

32. List any active grants or funding that support this research.

Cedar Hill University Faculty Startup Fund (2024–2027, $50,000 total). Approximately $15,000 remains and is allocated to student travel and equipment, not computing. No external grants currently fund this project.

33. Is this project currently supported by any cloud computing credits from another source?

- ■ Yes
- ■ No

Section 8: Supporting Materials (Optional)

34. CV / Biosketch

Uploaded (2-page NSF-format biosketch).

35. Supplementary Material

Uploaded: 3-page supplement containing (1) preliminary single-modality attack results on LLaVA-7B conducted using the departmental RTX 3090 cluster (image-only PGD attack achieving 72% success rate), (2) proposed system architecture diagram for the multimodal attack and defense pipeline, and (3) a table of the proposed adversarial attack taxonomy.

Section 9: Applicant Certification

36. Electronic Signature

Yuna Park

37. Date

May 15, 2026
