

Jinxi He

jhe44@u.rochester.edu | (585)-752-5010 | Website

Education

University of Rochester

B.S. Computer Science, B.A. Psychology

• cumulative GPA: 3.95/4.0, Dean's List (7 semesters)

• Schwartz Discover Grant

Rochester, New York

Sep. 2021 - May 2025 (*Expected*)

2023 Summer

Research Interests

• Adversarial Machine Learning, Trustworthy AI, Vision-Language Models (VLM), Diffusion models

Publications and Posters

• **Jinxi He**, Jing Bi, Pinxin Liu, Ali Vosoughi, Jiarui Wu, Chenliang Xu, I^2 : Illustration & Instructions Content[†]. *ICLR 2025 under review*.

• Zeliang Zhang, **Jinxi He**, Chenliang Xu, Guided Feature Fusion to Boost the Transferable Targeted Adversarial Attacks[†]. *ICLR 2025 under review*.

• **Jinxi He**, Zidian Xie, Dongmei Li, Public Discussions of Smoking Cessation Strategies and Challenges on Reddit. Poster presentation at the *SRNT Annual Meeting 2024*.

• **Jinxi He**^{*}, Jiarui Chen^{*}, Zidian Xie, Dongmei Li, Public Perceptions of Synthetic Nicotine Products on Twitter. *Frontiers in Public Health*.

• Jing Bi, **Jinxi He**, Jiarui Wu, Concept Binding: Learning Abstract Realities through Conceptual Fusion[†], *CVPR 2025 under review*.

• Mingqian Feng, Zeliang Zhang, **Jinxi He**, Investigating Catastrophic Forgetting in Neural Networks Through Adversarial and Clean Data Distributions[†], *in preparation for CVPR 2025*.

[†] The paper's title may differ from that in the actual submissions due to academic honesty reasons.

Academic Experiences

Department of Computer Science

University of Rochester, New York

Clinical & Translational Science Institute

University of Rochester Medical Center, New York

Research Assistant in Prof. Chenliang Xu and Prof. Dongmei Li's joint-lab

Sep. 2022 – Present

Project 1: I^2 : Automatic Generation Framework of Multimodal Instructional Content

• Achievement: Constructed an innovative generative framework to automatically generate continuous, instructive images or videos to augment current text-centric AI assistants.

• Assisted in proposing and construction of the 3-step framework: 1) Decomposed instructions into goal and step descriptions to mimic human logic. 2) Leveraged language models to generate textual instructions, and diffusion models to output relevant visuals conditioned on the text. 3) Incorporated image-text matching feedback to ensure consistency between modalities.

• Designed new evaluation metrics involving human assessment for comprehensive quality measurement.

Project 2: Guided Feature Fusion to Boost the Transferable Targeted Adversarial Attacks

• Achievement: Proposed a novel guided feature fusion method to generate adversarial examples with higher attack success rate and transferability against deep learning models.

• Assisted in designing the core techniques: 1) Retrieved semantically related images using CLIP to introduce target class features. 2) Fused the features of original and retrieved images via random splicing. 3) Optimized adversarial examples guided by the fused features.

• Compared various fusion schemes and selected the optimal one. Evaluated attack performance and transferability on Im-

geNet using multiple metrics.

Project 3: Concept Binding: Learning Abstract Realities through Conceptual Fusion

- Modify the cross-attention mechanism in the SDXL model, excluding irrelevant tokens and confine tokens attention to prevent semantic leakage and enhance layout control.
- Modify the self-attention mechanism in the SDXL model, allowing subject-background interaction to ensure seamless blending and apply the modifications to both conditional and unconditional noise estimation to avoid artifacts.
- Utilized LoRA to fine-tune two transformer blocks, achieving efficient style-content separation and enable various image stylization applications, such as image-based style transfer, text-based stylization, and consistent style generation.

Project 4: Leveraging Social Media Data to Study Smoking Behaviors

- Collected, process and analyzed around 100k user-generated posts about smoking from Twitter and Reddit to explore public perception of novel tobacco products, discussions of e-cigarette prevention, and challenges in smoking cessation.
- Cleaned, transformed, and integrated the multi-modal data (textual, visual, and video data) into structured datasets.
- Analyzed text data through LDA, BERTopic, VADER sentiment analysis, regex entity extraction; extracted visual features using ResNet and CLIP; performed statistical tests including ANOVA, regression, hypothesis testing; and created visualizations.

Work Experiences

Bank of Changsha

Changsha, China

Software Development Engineer, Intern

Jun. 2023 - Aug. 2023

- Worked closely with the business team to thoroughly understand their requirements, leveraged SQL's JOINs and complex window functions to extract data with optimized queries from Alibaba Cloud with more than 20000 statements efficiently.
- Developed and optimized a suite of anti-fraud computer vision models for 95% accuracy background separation in loan application photos, enabled the subsequent identification of fraudulent loan teams.
- Enhanced feature matching capabilities using the cutting-edge pretrained model, significantly boosted the system's recall rate to 80% against human-labeled data, effectively pinpointing fraud in loan applications.

Projects

Proactive AI Assistant | *Python, SQL, NLP, Bayesian Network*

- Improve the chatbot's related-questions recommendation feature by transitioning from hand-labeled relevant questions to intelligently suggesting five questions per conversation round based on a previous standardized question database.
- Leveraged NLP model word2vec and similarity algorithms to refine TOP5 recommendations for rare queries from the existing standard question database, increasing the user click rate and positive feedback rate.

CineFusion | *Android Studio, Java, Kotlin, Git*

- Designed and developed an Android application utilizing The Movie Database API for real-time multilingual movie data retrieval.
- Introduced a title-centric search in our app for precise results, paired with an embedded web view for allowing users to access web content directly within the app, enhancing user experience and promoting richer community engagement.

Text Searching Tool | *C, Valgrind*

- Introduce the concept of regular expressions and their relationship with Non-deterministic Finite Automata (NFAs).
- Design patterns such as finding words beginning with specific characters and ending with others (e.g., words starting with 'a' and ending with 'z').

Technical Skills

Programming Languages: C/C++, JAVA, Python, SQL, SAS, Kotlin, Swift, Shell, HTML/CSS, PHP

Technologies: Pytorch, OpenCV, NumPy, Android Studio, Git, Unix/Linux, TensorFlow, Google Cloud