Mohammad Reza Taesiri

1808, 4720 Lougheed Hwy. Burnaby, BC, Canada (V5C 0M8)
mtaesiri@gmail.com • +1 (438) 303-8905 • https://taesiri.ai • Google Scholar • GitHub • Hugging Face

RESEARCH HIGHLIGHTS

- My research focuses on evaluating and improving the fundamental perceptual and reasoning capabilities of Vision-Language Models (VLMs) through rigorous, large-scale benchmarking.
- My work has been recognized and utilized by leading AI research labs to evaluate flagship models:
  - **ZeroBench** was featured in Google's **Gemini 2.5 Pro** technical report.
  - **VLMsAreBlind** has been used for evaluation by **OpenAI**, **ByteDance**, and others.
- **Keywords**: VLMs, Post-training, Evals

CURRENT POSITION

EA Sports, Vancouver, BC, Canada
- Machine Learning Scientist — Jun 2025 – Present

EDUCATION

University of Alberta, Edmonton, Alberta, Canada
- Ph.D. in Software Engineering and Intelligent Systems — Sep 2021 – Sep 2024

Sharif University of Technology, Tehran, Tehran, Iran
- M.Sc. in Computer Software Engineering — Sep 2015 – Sep 2017

Amirkabir University of Technology, Tehran, Tehran, Iran
- B.Sc. in Pure Mathematics — Sep 2009 – Jun 2015

EXPERIENCE

ASGAARD Lab, University of Alberta
- Postdoctoral researcher — Oct 2024 – May 2025
  - Supervisor: Dr. Cor-Paul Bezemer
  - Focus: Vision-language models for understanding the world and detecting anomalies

La Forge, Ubisoft Montreal
- Research and Development Intern — Aug 2022 – Dec 2022
  - Supervisor: Dr. Sarra Habchi
  - Focus: Robustness of Foundation Models, Image and Video Retrieval

ASGAARD Lab, University of Alberta
- Graduate Research Assistant — Sep 2021 – Sep 2024
  - Supervisor: Prof. Cor-Paul Bezemer
  - Focus: Foundation Models for Video Games

Nguyen Lab, Auburn University
- Visiting Researcher — Mar 2021 – Present
  - Supervisor: Prof. Anh Totti Nguyen
  - Focus: Robust and Explainable Machine Learning

PUBLICATIONS

CONFERENCES

[1] Mohammad Reza Taesiri, A Ghildyal, S Zadtootaghaj, N Barman, and CP Bezemer – "VideoGameQA-Bench: Evaluating Vision-Language Models for Video Game Quality Assurance" in *The Thirty-Ninth Annual Conference on Neural Information Processing Systems (NeurIPS 2025)*

[2] A Vo, Mohammad Reza Taesiri, D Kim, and AT Nguyen – "B-score: Detecting biases in large language models using response history" in *Forty-Second International Conference on Machine Learning (ICML 2025)*

[3] Mohammad Reza Taesiri, and Cor-Paul Bezemer – "VideoGameBunny: Towards vision assistants for video games" in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision 2025. (WACV 2025)*

[4] Rahmanzadehgervi, Pooyan, Logan Bolton, Mohammad Reza Taesiri, and Anh Totti Nguyen – "Vision language models are blind" in *Proceedings of the Asian Conference on Computer Vision 2024. (ACCV 2024)*

[5] Mohammad Reza Taesiri, Tianjun Feng, Anh Nguyen and Cor-Paul Bezemer – "GlitchBench: Can large multimodal models detect video game glitches?" in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024. (CVPR 2024)*

[6] Mohammad Reza Taesiri, Giang Nguyen, Sarra Habchi, Cor-Paul Bezemer, and Anh Nguyen – "ImageNet-Hard: The Hardest Images Remaining from a Study of the Power of Zoom and Spatial Biases in Image Classification" in *Thirty-Seventh Annual Conference on Neural Information Processing Systems (NeurIPS 2023)*

[7] Mohammad Reza Taesiri*, Giang Nguyen*, and Anh Nguyen (* Denotes Equal Contribution) – "Visual correspondence-based explanations improve AI robustness and human-AI team accuracy." in *Thirty-sixth Annual Conference on Neural Information Processing Systems (NeurIPS 2022)*

[8] Finlay Macklon, Mohammad Reza Taesiri, Markos Viggiato, Stefan Antoszko, Natalia Romanova, Dale Paas, and Cor-Paul Bezemer – "Automatically Detecting Visual Bugs in HTML5 <canvas> Games." in *International Conference on Automated Software Engineering (ASE 2022)*

[9] Mohammad Reza Taesiri, Finlay Macklon, and Cor-Paul Bezemer – "CLIP meets GamePhysics: Towards bug identification in gameplay videos using zero-shot transfer learning." in *The Mining Software Repositories conference (MSR 2022)*

JOURNALS

[9] Giang Nguyen, Valerie Chen, Mohammad Reza Taesiri, and Anh Totti Nguyen – "PCNN: Probable-Class Nearest-Neighbor Explanations Improve Fine-Grained Image Classification Accuracy for AIs and Humans" in *Transactions on Machine Learning Research (TMLR)*

[11] Mohammad Reza Taesiri, Finlay Macklon, Sarra Habchi, and Cor-Paul Bezemer – "Searching bug instances in gameplay video repositories" in *IEEE Transactions on Games (ToG 2024)*

[12] Mohammad Reza Taesiri, Moslem Habibi, and MohammadAmin Fazli – "A Video Game Testing Method Utilizing Deep Learning" in *Journal on Computer Science and Engineering (JCSE 2021)*

WORKSHOPS

[13] Giang Nguyen, Mohammad Reza Taesiri, Sunnie S. Y. Kim, and Anh Totti Nguyen – "Allowing humans to interactively guide machines where to look does not always improve a human-AI team's classification accuracy." in *The 3rd Explainable AI for Computer Vision (XAI4CV) Workshop (CVPR 2024)*

PREPRINTS

[14] A Vo, KN Nguyen, Mohammad Reza Taesiri, VT Dang, AT Nguyen, and D Kim – "Vision Language Models are Biased" in *arXiv Preprint*

[15] Mohammad Reza Taesiri, B Collins, L Bolton, VD Lai, F Dernoncourt, T Bui, and AT Nguyen – "Understanding Generative AI Capabilities in Everyday Image Editing Tasks" in *arXiv Preprint*

[16] T Nguyen, L Bolton, Mohammad Reza Taesiri, and AT Nguyen – "HoT: Highlighted Chain of Thought for Referencing Supporting Facts from Inputs" in *arXiv Preprint*

[17] J Roberts, Mohammad Reza Taesiri, A Sharma, A Gupta, S Roberts, I Croitoru, et al. – "Zerobench: An impossible visual benchmark for contemporary large multimodal models" in *arXiv Preprint*

[18] Mohammad Reza Taesiri, Finlay Macklon, Yihe Wang, Hengshuo Shen, and Cor-Paul Bezemer – "Large Language Models are Pretty Good Zero-Shot Video Game Bug Detectors." in *arXiv Preprint*

[19] MohammadAmin Fazli*, Ali Owfi*, and Mohammad Reza Taesiri* (* Denotes Equal Contribution) – "A Data-Driven Analysis on NFT Auctions: Assessment, Opportunities and Fraudulent Activities." in *arXiv Preprint*

AWARDS & SCHOLARSHIPS

- Alberta Innovates Scholarship 2024
- Alberta Graduate Excellence Scholarship (AGES) 2023
- Upper Bound Talent Bursary 2023
- NeurIPS Scholar Award 2022

- Graduate Research Assistant, University of Alberta 2021
- Ranked 10th, National entrance exam in Software Engineering, Iran 2015
- Ranked 11th, National entrance exam in Algorithms and Theory of Computation, Iran 2015

INVITED TALKS & ACTIVITIES

Foundation Models for Video Game Quality Assurance, 2024
Honours Seminar, University of Alberta, Edmonton, Canada
A Brief Tutorial on Large Language Models, 2023
ISAIC, University of Alberta, Edmonton, Canada
Zoom Is What You Need: An empirical study of the power of zoom and spatial biases in image classification, 2023
Samsung SAIT AI Lab (SAIL), Montreal, Québec, Canada

PROJECTS & DATASETS

VideoGameQA-Bench
- A benchmark to evaluate vision language models for the task of video game quality assurance. May 2025

VLMsAreBiased
- A benchmark for evaluating bias in vision language models. May 2025

VLMsAreBlind
- A benchmark for evaluating vision-language models on basic visual perception primitives. May 2024

GlitchBench
- A benchmark to evaluate large multimodal models for the task of video game testing. Dec 2023

ImageNet-Hard
- Introduced a challenging dataset to rigorously assess the robustness of diverse vision models. Apr 2023

Claude Reads ArXiv
- Harnessing the power of the *Claude-v1.3-100k* to answer questions about academic papers. Apr 2023

Intelligent Image Captioner
- Empowering ChatGPT with the ability to see and interpret images, using Detic. Dec 2022

CLIP Meets GamePhysics, Hugging Face Spaces
- Developed a CLIP-based video retrieval system for video games. Mar 2022

The GamePhysics Dataset, Hugging Face Datasets
- A dataset of video game bugs Jan 2022

TEACHING EXPERIENCE

University of Alberta, Edmonton, Alberta, Canada
- Teaching Assistant Sep 2023 – May 2024
  - Served as a teaching assistant for multiple courses.
  - ECE 447 - Data Analysis and Machine Learning for Engineers - Winter 2024
  - ECE 342 - Probability for Electrical and Computer Engineers - Winter 2024
  - ECE 325 - Object-Oriented Software Design - Fall 2023
  - ECE 321 - Software Requirements Engineering - Fall 2023

Sharif University of Technology, Tehran, Tehran, Iran
- Teaching Assistant - Head Jan 2016 – Jun 2016
  - Led a team of teaching assistants in the Discrete-Event Simulation course.
  - Collaborated with the professor to develop lesson plans, assess students' performance

CERTIFICATES & ONLINE COURSES

- Deep Reinforcement Learning Nanodegree, Udacity 2020
- Reinforcement Learning Specialization, Coursera, University of Alberta 2020
- Computational Neuroscience, Coursera, University of Washington 2020
- Deep Learning Specialization, Coursera, DeepLearning.AI 2018
- Image and video processing, Coursera, Duke University 2014
- Heterogeneous Parallel Programming, Coursera, University of Illinois Urbana-Champaign 2014
- Programming Languages, University of Washington 2014

SKILLS

Machine Learning: PyTorch, Keras, JAX
Programming: Python, C#, Java, C/C++, Swift, Objective-C, Scheme, Racket, ML, CUDA
Other Technologies: Docker and Kubernetes, NodeJS, MongoDB, Neo4j, Wolfram Mathematica
Game Engines: Unity, Unreal Engine

HOBBIES

Photogrammetry, Digital Photography, Hiking

REFERENCES

- **Dr. Anh Totti Nguyen**
  Associate Professor, Auburn University
  anhnguyen@auburn.edu
- **Dr. Trung Bui**
  Senior Research Scientist & Research Manager, Adobe Research
  bui@adobe.com
- **Dr. Cor-Paul Bezemer**
  Associate Professor, University of Alberta
  bezemer@ualberta.ca
- **Dr. Marek Reformat**
  Professor, University of Alberta
  reformat@ualberta.ca

[*CV compiled on 2025-10-19*]