Raymond (Zerui) Wang, Ph.D.

AI Research Scientist — Vision–Language & Multimodal Model Evaluation

Moving to San Jose, CA | +1 263 382 8882 | xaiport@gmail.com
LinkedIn | Google Scholar | GitHub | xaiport.com

Summary

AI research scientist specializing in evaluation of vision–language and multimodal models (video, image, text, audio). Ph.D. Concordia 2025; 12 publications; first-author at ICSE (top conf in SE), IEEE Transactions, ACM Transactions. Built the first single-pass spatio-temporal attention-attribution method for video transformers — perceptual *faithfulness* and *monotonicity* metrics at <150 ms — and the first joint spatio-temporal adversarial evaluation of video models across 20,000 clips. Designed cloud-agnostic evaluation infrastructure benchmarking heterogeneous vision and multimodal backends across Azure, GCP, and AWS behind unified contracts. Currently shipping production multimodal generation pipelines (vision + language) at Maket.AI; 40+ peer reviews at IEEE Transactions & AAAI.

Experience

Maket.AI Aug 2025 – Present

Generative AI Engineer *— Multimodal Pipelines, Vision–Language, 3D Vision* *Montreal, Canada*

- **Production multimodal generation pipelines** — multi-stage LLM orchestration over floor-plan and design-brief inputs (vision + language) driving design reasoning, 3D visualization, and iterative AI-chat design loops.
- **Real-time multimodal AI chat** — context-aware vision & language responses driving iterative design refinement; prompt orchestration ensuring reliable outputs across multi-stage generation workflows.
- **3D vision in production** — SAM3D segmentation; Gaussian Splatting / NeRF reconstruction from phone video for in-app 3D scene capture and visualization.

Concordia University May 2020 – June 2025

Doctoral Researcher *— Video/Multimodal Perceptual Metrics, Eval Infrastructure & Adversarial Eval* *Montreal, Canada*

- **Novel perceptual metrics for video transformers (STAA)** — first single-pass spatio-temporal attention attribution method, defining *faithfulness* and *monotonicity* measurements at <150 ms latency (97% reduction vs. baseline). Validated on Kinetics-400 and Something-Something V2; enables eval-as-CI inside training loops. *IEEE Access 2025.*
- **Scalable automated multimodal evaluation infrastructure (XAIport / XAIpipeline)** — microservice framework turning evaluation from manual notebook into automated MLOps stage; OpenAPI 3.0 contracts; cloud-agnostic orchestration of heterogeneous vision/language backends; regression testing, dashboards, drift monitoring, alerting. Validated at scale of 500+ evaluation pipelines. *ICSE 2024; IEEE SSE 2025.*
- **Multi-cloud vision/multimodal benchmarking** — cross-platform comparative evaluation across Azure Cognitive Services, GCP Vertex AI, and AWS Rekognition behind unified REST contracts; containerized microservices; reproducible benchmark harness reporting. *IEEE TCC 2024.*
- **Adversarial red-teaming & regression harness** — first joint spatio-temporal adversarial attack on video transformers via XAI-guided gradient perturbation: 84.6% attack success rate; statistical evaluation across 20,000 videos; reversed into adversarial-aware training cutting model vulnerability by >50%. *ACM TOMM 2025.*
- **Teaching Assistant** — Software Engineering (COEN 6311), Programming on Cloud (COEN 6313), Distributed Software Systems (COEN 6731), Programming Methodology (COEN 244); 30–100+ students per term.

Quantum Apr 2023 – Aug 2025

AI Engineer *— LLM Research Pipelines, Investment Diligence* *Montreal, Canada*

- **AI-orchestrated research pipeline** — built an LLM workflow combining web search, multi-source cross-verification, and structured bilingual report generation with inline citations.
- **Tech-stack-anchored startup evaluation** — scored each company's defensibility along 5 axes (data moat, representation quality, retrieval architecture, evaluation rigor, online monitoring).

Université de Montréal Sep 2019 – Mar 2021

Research Associate *— Numerical Simulation & Dynamical Systems Modeling* *Montreal, Canada*

- **Numerical simulations & PDE solvers** — fluid transport, multi-physics coupling, and dynamical systems in Python / MATLAB / COMSOL; finite-volume, finite-element, and spectral solvers validated against analytical benchmarks.

Education

Concordia University 2020 – 2025
Ph.D., Computer Engineering *— perceptual metrics & adversarial robustness for video transformers* *Montreal, Canada*

Technical University Dortmund 2014 – 2018
M.Sc., Process System Engineering — modeling, PDE, control theory, computational simulation *Dortmund, Germany*

China University of Mining and Technology 2010 – 2014
B.Sc., Chemical Engineering *Xuzhou, China*

Publications

Wang, Z., Liu, Y. "STAA: Spatio-Temporal Attention Attribution for Real-Time Interpreting Transformer-Based Video Models." *IEEE Access*, vol. 13, pp. 101647–101661, 2025.

Wang, Z., Liu, Y. "Joint Spatio-temporal Adversarial Attacks on Video Transformer Models Through XAI-guided Gradient Perturbation." *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 2025.

Kondal, A.S., Ghataura, R.S., Liu, Y., **Wang, Z.** "VideoXAI: A Hybrid Architecture for Explainable AI Pipelines of Robust Video Classification." *Proc. IEEE International Conference on Big Data*, 2025.

Wang, Z., Liu, Y., Huang, J. "An Open API Architecture to Discover the Trustworthy Explanation of Cloud AI Services." *IEEE Transactions on Cloud Computing*, vol. 12, no. 2, pp. 762–776, 2024.

Wang, Z., Liu, Y., Thiruselvi, A.A., Hamou-Lhadj, W. "XAIport: A Service Framework for the Early Adoption of XAI in AI Model Development." *Proc. 46th IEEE/ACM International Conference on Software Engineering (ICSE)*, Lisbon, Portugal, pp. 67–71, 2024.

Wang, Z., Liu, Y. "Cloud-Based XAI Services for Assessing Open Repository Models Under Adversarial Attacks." *Proc. IEEE International Conference on Software Services Engineering (SSE)*, pp. 141–152, 2024.

Wang, Z., Liu, Y. "XAIpipeline: Automated Orchestration of Explainable AI Service for Cloud AI and Open-Source Models." *Proc. IEEE International Conference on Software Services Engineering (SSE)*, 2025.

Wang, Z., Liu, Y. "Spatio-temporal Explanation for Adversarial-Aware Cloud Vision Services." *Proc. IEEE Computer Software and Applications Conference (COMPSAC)*, 2025.

Wang, Z., Liu, Y. "The Role of Provenance Modeling in Tracing and Reproducing Explainable AI Pipelines." *Proc. International Conference on Computational Science and Computational Intelligence (CSCE)*, 2025.

Huang, J.*, **Wang, Z.***, Li, D., Liu, Y. "The Analysis and Development of an XAI Process on Feature Contribution Explanation." *Proc. IEEE International Conference on Big Data*, Osaka, Japan, pp. 5039–5048, 2022.

Li, D., Liu, Y., Huang, J., **Wang, Z.** "A Trustworthy View on Explainable Artificial Intelligence Method Evaluation." *IEEE Computer*, vol. 56, no. 4, pp. 50–60, 2023.

Neghawi, E., **Wang, Z.**, Huang, J., Liu, Y. "Linking Team-level and Organization-level Governance in Machine Learning Operations through Explainable AI and Responsible AI Connector." *Proc. IEEE Computer Software and Applications Conference (COMPSAC)*, pp. 1223–1230, 2023.