Zhenhao Chen

+971 58 523 9375 | zhenhao.chen@mbzuai.ac.ae
github.com/viewsetting | scholar.google.com/citations

Education

Mohamed bin Zayed University of Artificial Intelligence
Doctor of Philosophy in Machine Learning Supervisor: Prof. Kun Zhang & Dr. Mingming Gong *Aug. 2023 - Present*

Mohamed bin Zayed University of Artificial Intelligence
Master of Science in Machine Learning, GPA:3.95/4.0 *Aug. 2021 - June. 2023*

Northeastern University
Bachelor of Engineering in Computer Science and Technology, GPA:88/100 *Sep. 2016 - Jun. 2020*

Research Focus

Full-time PhD Student Aug. 2021 - Present
Mohamed bin Zayed University of Artificial Intelligence

- **LLM Agents:** Built and evaluated LLM agents for scientific discovery and active experimentation; developed interactive benchmarks (e.g., *CausalGame*) to assess causal thinking under biases and confounders.
- **LLM Inference-time Reasoning:** Pioneered a novel decoding strategy, *Reflection-Window Decoding*, which introduces a "look-ahead" refinement mechanism to address the inherent sub-optimality of autoregressive generation. Demonstrated significant gains on complex reasoning benchmarks.
- **LLM Self-Correction & Alignment:** Investigated the intrinsic self-correction capabilities of LLMs without external oracles. Proposed the *If-or-Else (IoE)* framework, identifying model *confidence* as a critical latent factor for effective hallucination detection and correction.
- **Causal Representation Learning:** Explored the intersection of Causal Inference and Generative Models. Developed a generic paradigm for temporal causal representation, enhancing model robustness and interpretability.

Industry Experience

Research Intern May 2026 - Present
TEG, Tencent

- **Copilot Agent for Game:** Building and training a copilot agent for Honors of Kings series game. Exploring LLM policy and enhanced reasoning for strategy game copilot.

Selected Publications

CausalGame: Benchmarking Causal Thinking of LLM Agents in Games
*Accepted to **ICML 2026 (Oral)** & **FMs for Science, ICLR 2026 Workshop***

- Introduced **CausalGame**, an interactive benchmark for evaluating causal thinking of LLM agents via active experimentation.
- Designed 14 game settings with selection bias, noisy measurements, and hidden confounders to emulate realistic scientific discovery.
- Lead author & tech lead. Project Website

Reflection-Window Decoding: Text Generation with Selective Refinement
*Accepted to **ICML 2025***

- Proposed a novel **Reflection-Window** decoding strategy to address sub-optimality of autoregressive generation.
- Introduced a pausing criterion that allows the model to selectively refine past tokens during generation.
- Co-first author. Paper Link

Confidence Matters: Revisiting Intrinsic Self-Correction Capabilities of LLMs
arXiv Preprint

- Identified **confidence** as a key latent factor in self-correction and proposed the **If-or-Else (IoE)** framework.
- Achieved consistent accuracy gains in intrinsic self-correction without external oracles.
- Co-first author. Paper Link

CaRiNG: Learning Temporal Causal Representation under Non-Invertible Generation
*Accepted to **ICML 2024***

- Built a generic paradigm of video understanding from the view of Causal Representation Learning.
- Co-first author. Paper Link

Unsupervised Sampling Promoting for Stochastic Human Trajectory Prediction
*Accepted to **CVPR 2023***

- Proposed BOSampler to adaptively mine potential paths with Bayesian optimization in an unsupervised manner.
- Co-first author. Paper Link

Academic Services

Journal Reviewer: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *Pattern Recognition* and *ACM Computing Survey*.
Conference Reviewer: NeurIPS (2024 -), ICLR (2024 -), ICML (2025 -), AISTATS (2024 -)
Student Leader: Student leader of Center for Integrative Artificial Intelligence (CIAI) of MBZUAI (2022 -), organized the weekly seminar and maintain the official web page.

Other Publications

CausalEvolve: Towards Open-Ended Discovery with Causal Scratchpad
*Accepted to **ICLR 2026 Workshop on AI with Recursive Self-Improvement (Spotlight)***

- Proposed *CausalEvolve*, an evolutionary AI Scientist framework equipped with a causal scratchpad that identifies outcome-level and procedure-level factors to guide the evolution process.
- Demonstrated improved evolution efficiency and state-of-the-art results across 4 open-ended scientific discovery tasks (Hadamard Matrix, Autocorrelation Inequality, Circle Packing, AIME).

Maniplvm-r1: Reinforcement Learning for Reasoning in Embodied Manipulation with Large Vision-Language Models
*Accepted to **AAAI 2026***

- Developed an embodied LMM capable of understanding complex instructions and generating precise object-centric manipulation trajectories.
- Fine-tuned LMMs with robotic data, bridging the gap between high-level reasoning and low-level control.

PersonaX: Multimodal Datasets with LLM-Inferred Behavior Traits
*Accepted to **NeurIPS 2025 Multimodal Algorithmic Reasoning Workshop***

- Constructed a large-scale multimodal dataset with behavioral traits inferred by high-performing *LLMs*.
- Introduced a Causal Representation Learning (CRL) framework to analyze the relationship between visual attributes and LLM-inferred personality traits.
- Access dataset through Huggingface.

Hazards in Daily Life? Enabling Robots to Proactively Detect and Resolve Anomalies
*Accepted to **NAACL 2025 (Long Paper)***

- Proposed *AnomalyGen*, a generative framework where *LLM-based agents* collaborate to brainstorm and simulate household hazard scenarios.
- Enabled robots to proactively detect anomalies by leveraging foundational models to bridge the gap between language descriptions and 3D environments.

MMAC-Copilot: Multi-modal Agent Collaboration Operating System Copilot
arXiv Preprint

- Developed a *GUI Multi-Agent* framework that leverages multi-modal LLMs to interact with complex operating system environments (e.g., office, gaming).
- Proposed a collaborative agent team (”Librarian”, ”Programmer”, ”Viewer”) to reduce hallucination and enhance planning.
- Achieved state-of-the-art results on GAIA and Visual Interaction Benchmark (VIBench).

BenchLMM: Benchmarking Cross-style Visual Capability of Large Multimodal Models
*Accepted to **ICCV 2024***

- Constructed a comprehensive benchmark to evaluate the robustness of **Large Multimodal Models (LMMs)** (e.g., GPT-4V, LLaVA) across diverse artistic styles.

- Analyzed the reasoning gaps in current LMMs when facing out-of-distribution visual inputs.
- Project Homepage

Empowering Graph Invariance Learning with Deep Spurious Infomax

Accepted to ***ICML 2024***

- Developed a method to enhance Graph Neural Networks generalization on out-of-distribution (OOD) data.
- Utilized the invariance principle to mitigate the impact of spurious correlations in graph structured data.

Technical Skills

Core Research: LLM Post-training & Alignment, Inference-time Reasoning, Agents, Causal Representation Learning, Multi-modal Learning, Prompt Engineering
Languages: Python, C/C++, Rust, Shell, SQL
Frameworks & Libraries: PyTorch, Transformers, DeepSpeed, vLLM, Accelerate, Pandas, NumPy, Scikit-learn
Developer Tools: Docker, Git, Linux/Bash, WandB, VS Code, Raspberry Pi

References

Prof. Kun Zhang (Carnegie Mellon University & MBZUAI)

Title: Professor & Associate Chair of the Machine Learning Department, MBZUAI; Professor, Carnegie Mellon University
Email: kunz1@cmu.edu
Tel: +1(412)268-8573
Address: Baker Hall 161B, 5000 Forbes Ave, Pittsburgh, PA 15213