

Michael Xu

<https://michaelxu.dev>

Email: mxu1925@gmail.com

Linkedin: [in/yunhuixu](https://www.linkedin.com/in/yunhuixu)

EXPERIENCE

Senior Research Engineer

May 2022 – Present

Microsoft Research

San Francisco, CA

- **Delivered executive-level demonstrations** of gaming AI innovations, effectively communicating technical concepts to drive strategic decision-making at Microsoft Research and Xbox.
- **Provided technical leadership** for LLM integration across Xbox game studios, implementing LLM-driven systems that enhanced narrative complexity and created rich generative player interactions in multiple in-development game titles.
- **Cross-functional collaboration** with Microsoft AI on the Gaming Copilot Companion initiative, integrating on-device multimodal vision models with real-time gameplay analysis to provide contextual assistance across multiple game genres, resulting in improved player experience and accessibility.
- **Designed and implemented a scalable data collection platform** on Xbox Insider, leveraging gamified interactions to process 50,000+ daily human preference data points, aimed at increasing AI creativity capabilities through RLHF.
- Architected a goal-oriented conversational prompting framework to improve LLM alignment that improved quest scenario completion rates by 25% ([paper](#)).
- Conducted research on a dynamic branching narrative framework using LLMs, enabling procedurally generated dialogue trees that increased story variation while maintaining narrative coherence across storylines ([paper](#)).

Software Engineer

Feb. 2021 – May 2022

Microsoft

Redmond, WA

- **Migrated a mission-critical content management system** to the cloud, handling 27,000+ requests per second, implementing distributed caching and load balancing to maintain 99.99% uptime during high-traffic events. Actively hosting 200+ sites across Microsoft.com including Software Download and Investor Release.
- **Constructed a multi-region deployment architecture** using Azure Pipelines and Kubernetes, reducing deployment times by 60% while eliminating service interruptions through rolling updates.
- Created automated systems for log aggregation and SSL certificate management, reducing manual operations by 80% and eliminating certificate-related incidents.

PROJECTS

Craft an Iron Sword | *Minecraft, Conversational Agent, RAG* | [GitHub](#)

Aug. 2021 – Present

- Developed an advanced conversational AI system combining LLMs with RAG methodology for enhanced natural language understanding and collaborative task completion in Minecraft ([paper](#)).
- Earned recognition from Microsoft CTO Kevin Scott and won Executive Hack category at Microsoft Hackathon 2022, demonstrating innovative applications of AI in gaming.
- Enhanced the system with multimodal capabilities including real-time video and voice interaction, leading to a showcase demonstration by CEO Satya Nadella at Build 2024 ([video](#)).

Cryptopoly | *Typescript, Vue, Three.js, WebRTC, Solidity* | [GitHub](#)

Apr. 2019 – Jun. 2023

- Designed and implemented a blockchain-based gaming protocol using Ethereum and State Channels, reducing transaction costs by 90% while maintaining decentralized gameplay integrity.
- Developed a full-featured Monopoly implementation as proof-of-concept, demonstrating transparent game mechanics and trustless multiplayer interactions on the blockchain.
- Engineered a scalable serving architecture handling concurrent gameplay for thousands of users, with built-in DDoS protection and automatic load balancing.

EDUCATION

University of Illinois at Urbana-Champaign

Champaign, IL

M.S. in Computer Science, Data Science Concentration - 3.91 GPA

Jan. 2022 – Dec. 2023

Arizona State University

Tempe, AZ

B.S. in Computer Science, Minor in Statistics - 4.0 GPA

Aug. 2017 – Dec. 2020

TECHNICAL EXPERTISE

Research Areas: Conversational AI & Natural Language Processing, Large Language Models & RLHF, Model Optimization & Deployment, Multimodal AI Systems

Languages: Python, TypeScript / JavaScript, PHP, SQL, Java, Triton, C++

Frameworks: PyTorch, vLLM, Transformers, TensorRT, DeepSpeed, Kubernetes, Docker