Yaowen Shi

18938684938 | ethan_ywshi@foxmail.com | ethanywshi.github.io

Education

South China University of Technology (SCUT)

B.S. in Big Data Management and Application *Sep. 2023 – Jun. 2027 (Expected)*

- Academic standing: GPA: **3.95/4.00**; average score: **93.7**; major rank: **3/161**
- Selected courses: Calculus I and II (**99, 100**); Linear Algebra and Analytic Geometry (**100**); Operations Research (**99**); Python Data Analysis (**97**); Probability Theory and Mathematical Statistics (**96**); Large Language Models and GenAI (**96**); Econometrics (**95**); Data Structures (**95**); ... **Ranked 1st in 11 courses**.

Research Interests

I am interested in **multimodal agent systems** that connect reasoning, planning, memory, tool use, and multimodal interaction with real creation workflows. My research focuses on **LLM-based agents for visual and multimodal generation**: agents that understand human intent, decompose open-ended tasks, coordinate specialized models or tools, and improve outputs through **planning, feedback, and self-reflection**. I am broadly interested in image generation, video generation, 3D/4D generation, visual reasoning, and autonomous task execution, with the long-term goal of making foundation models more capable, controllable, and useful for complex creative tasks.

Experience

Artist-Created 3D Mesh Generation Mar. 2025 – Apr. 2025

Collaborator: Jinxiu Liu, Research Scientist in Nex-AGI

- Studied efficient mesh compression and tokenization for autoregressive artist-created mesh generation: explored VQ-VAE-based mesh vocabularies and adjacency-aware tokenization to shorten mesh token sequences, increase the number of faces models can generate, and preserve geometric structure; also investigated human-preference alignment methods such as RLHF/DPO to improve perceptual quality and user preference.

Image Generation Apr. 2025 – Jun. 2025

Collaborator: Jinxiu Liu, Research Scientist in Nex-AGI

- Prototyped self-reflection and reinforcement learning for flow matching text-to-image models (e.g., FLUX): turned intermediate samples into self-critique signals to define rewards, and learned an adaptive conditioning/guidance schedule under a fixed sampling budget; improved prompt adherence, text rendering, and subjective preference while reducing trajectory drift.

Controllable Video Generation Jul. 2025 – Aug. 2025

Collaborator: Jinxiu Liu, Research Scientist in Nex-AGI

- Studied training-free video motion transfer methods with Wan 2.1 and DiT-based video models such as CogVideoX (5B/2B), using DiTFlow as a baseline: extracted patch-wise motion cues from source videos via cross-frame attention / Attention Motion Flow (AMF), guided latent denoising or positional embeddings with motion losses at inference, and compared how motion fidelity trades off with prompt adherence, identity, and appearance under subject or scene changes.

Omni Multimodal Generation Agent System Sep. 2025 – Jan. 2026

Collaborator: Jinxiu Liu, Research Scientist in Nex-AGI

- Built an omni multimodal generation system with NexAU as the agent-universe orchestration layer and ComfyUI as the workflow execution backend: defined tool-using agents, skills, and workflow primitives to plan and execute image/video generation, image/video editing, and related multimodal creation tasks; inspired by omni systems such as Nano Banana and BAGEL, and achieved state-of-the-art on ComfyBench.

Vision as Inverse Graphics in Blender Feb. 2026 – Present

Collaborator: Jinxiu Liu, Research Scientist in Nex-AGI

- Designed an efficient multi-agent collaboration system for vision-as-inverse-graphics generation: used Native Skills to coordinate GPT-5, Qwen3-VL-8B, visual reasoning agents, and symbolic-program agents for 2D, 3D, and 4D multimodal generation; achieved competitive scores on BlenderGym, BlenderBench, and SlideBench compared with the VIGA baseline.

Projects

OpenDinq | *Evidence-backed AI-native profiles, card workspaces, and explainable people discovery* GitHub

- Built an open-source product alpha for evidence-backed AI-native profiles: source ingestion from GitHub, websites, OpenAlex, arXiv, ORCID, manual links, and notes, then normalized, deduped, and quality-scored claims.
- Implemented card workspaces for profile review and curation, public profile publishing, and explainable people discovery with matched claims, artifacts, evidence snippets, and score breakdowns.

Awards

Meritorious Winner (Top 7%), Mathematical Contest in Modeling (MCM) / Interdisciplinary Contest in Modeling (ICM) 2025

Second Prize (Top 15%), Contemporary Undergraduate Mathematical Contest in Modeling (CUMCM), Guangdong Division 2024

Technical Skills

Languages: C++, Python, SQL, JavaScript, HTML, CSS
ML: PyTorch, Transformers, diffusers
Developer Tools: Git, Docker
English: CET-6