

Cheng Chen

Tongji University | (+86) 13587919792
2351445@tongji.edu.cn | chenceason10@gmail.com
Homepage | GitHub | Google Scholar

Education

Institute of Automation, Chinese Academy of Sciences Sept. 2027 (expected)
Incoming Ph.D. Student, Beijing, China. Supervisor: **Prof. Kaiqi Huang**.

Tongji University Sept. 2023 - Jun. 2027 (expected)
B.Eng. in Communication Engineering, Shanghai, China.
Average score: **92.02/100** GPA: **4.70** Major rank: **2/85**

Main interests

I explore simple, elegant ways to make AI models more capable. My interests span multimodal large language models, image/video generation, and embodied AI.

Publications

* Equal contribution; † corresponding author.

- 1) **If You Hear It, Help Find It: Orthogonal Knowledge Distillation for Open-Vocabulary Audio-Visual Event Localization.**
Cheng Chen*, Yi Xu*, Wenzhuo Lei.
ACM MM 2026, accepted. [PDF]
Separates supervision from visual and audio teachers for open-vocabulary audio-visual event localization.
- 2) **OrthKD: Extracting Generalized Clinical Knowledge from Heterogeneous Teachers for Lightweight Deployment.**
Yi Xu*, **Cheng Chen***, Mufan Cao.
IJCAI 2026, accepted. [PDF]
Distills complementary clinical knowledge from heterogeneous teachers into a lightweight retinal-image classifier.
- 3) **Enabling Visual Reasoning in Text-Only LLMs via Symbolic Scene Descriptions.**
Wenzhuo Lei, Mufan Cao, Chaorong Ye, Yi Xu, **Cheng Chen†**.
IJCNN 2026, accepted. [PDF]
Links object detection to text-only language models using compact symbolic scene descriptions.

Preprints

- 4) From Scores to Checks: A Human-Annotated Rubric Benchmark for AI-Generated Image Faithfulness. *Preprint, 2026*.
Uses human-annotated rubric checks to evaluate whether generated images satisfy individual prompt requirements.
- 5) What Can Activation-Patching Evidence Establish? *Preprint, 2026*.
Examines the scope and limitations of the functional claims that activation-patching evidence supports.
- 6) Follow Me If You Can: Auditing Understanding-to-Generation Transfer. *Preprint, 2026*.
Audits transfer from visual-understanding training to image generation through controlled comparisons.
- 7) Allocation Entropy for Single-Pass Risk Triage in Neural Operators. *Preprint, 2026*.
Studies single-pass screening of unreliable neural-operator predictions without altering the predicted physical field.

Research internships

Research Intern
Hong Kong University of Science and Technology (Guangzhou) Sept. 2026 - Present
Zhejiang University Jun. 2026 - Sept. 2026
Tongji University Oct. 2025 - Jun. 2026

Academic service

Reviewer: ICLR, PRCV, ICME, IJCNN, AAAI.

Research experience

If You Hear It, Help Find It: Orthogonal Knowledge Distillation for Open-Vocabulary Audio-Visual Event Localization *ACM MM 2026; co-first author*

- Routed strong visual supervision to the localization path and weaker audio supervision to an orthogonal auxiliary path; used text prototypes to support unseen event categories.
- Reached **0.816 segment average precision** on OV-AVEBench. F1@0.5 improved over the official fine-tuning baseline by **2.7 points** overall and **3.4 points** on unseen categories.

OrthKD: Extracting Generalized Clinical Knowledge from Heterogeneous Teachers for Lightweight Deployment *IJCAI 2026; co-first author*

- Combined full supervision from a stronger teacher with representation-level transfer from a weaker teacher; orthogonal student projections preserved complementary clinical information.
- Used a combined pool of **132,049** retinal images. A **5.4M**-parameter student reached **0.885 QWK** (quadratic weighted kappa) on EyePACS and improved zero-shot Messidor-2 QWK from **0.507 to 0.728**.

Enabling Visual Reasoning in Text-Only LLMs via Symbolic Scene Descriptions *IJCNN 2026; corresponding author*

- Proposed the initial idea and worked with a junior student, who led implementation and experiments, to connect an object detector with a text-only language model through compact scene reports.
- The pipeline reached **42.3%** visual question answering accuracy on a stratified **5,000-image VQAv2 subset**, versus **38.2%** for scene-graph serialization, using **1/2.5 as many prompt tokens as that baseline**.

From Scores to Checks: A Human-Annotated Rubric Benchmark for AI-Generated Image Faithfulness *2026*

- Studied whether generated images satisfy individual requirements in text prompts, using human annotations to separate different types of faithfulness errors.
- Compared lightweight evaluators under matched conditions; examined confidence calibration, supporting image regions, and transfer across generators and datasets.

What Can Activation-Patching Evidence Establish? *2026*

- Studied which functional claims observations from neural-network interventions can support, including cases where matching observed outputs leaves other behavior unresolved.
- Combined mathematical analysis with controlled language-model tests to examine additional measurements and failures on previously unmeasured responses.

Follow Me If You Can: Auditing Understanding-to-Generation Transfer *2026*

- Audited whether training for visual understanding changes image generation in unified multimodal models, focusing on counting and spatial relations.
- Used paired model states and controlled comparisons to distinguish gains in answering questions, effects on generation, and transfer of the requested visual attribute.

Allocation Entropy for Single-Pass Risk Triage in Neural Operators *2026*

- Studied risk indicators available within one prediction pass of a neural operator, preserving the predicted physical field and avoiding a separate learned risk head.
- Tested how the scoring representation affects error ranking and acceptance decisions across physical tasks; compared screening quality and computational cost with alternative indicators.

Awards and honors

- National First Prize, China Robot and Artificial Intelligence Competition (2026); Team Captain.
- National Second Prize, Global Campus Artificial Intelligence Algorithm Elite Competition (2025); Team Captain.
- National Third Prize, National Undergraduate Embedded Chip and System Design Competition, FPGA Track (2025); Team Captain.
- Provincial Third Prize, Contemporary Undergraduate Mathematical Contest in Modeling (2025); Team Captain.
- Provincial Third Prize, National College Student Statistical Modeling Competition (2025); Team Captain.
- Provincial Second Prize, The Chinese Mathematics Competitions (2024–2025).
- Tongji University Outstanding Student Scholarship, Second Class (2023–2024 and 2024–2025).

Technical skills

Programming: Python, C/C++, Verilog. **Machine learning:** PyTorch.

Tools: Git, Linux, LaTeX, OpenAI Codex, Claude Code.

Foundations: Machine learning, algorithms, linear algebra, probability, optimization, and signal processing.