Selena Kexin Song

Tokyo, Japan SelenaSong08@gmail.com +81 (080)-4120-1780

Education

University of California, San Diego 2025.06 – Present
Visiting Graduate Student, **Advisor:** Biwei Huang San Diego, CA, USA

University of Tokyo 2023.10 – 2025.10
Master of Engineering, **Advisor:** Yutaka Matsuo Tokyo, Japan
· **Core Coursework:** Deep Learning, Reinforcement Learning, Machine Learning

Fudan University 2019.09 – 2023.06
Bachelor of Science in Physics, **Advisor:** Jiping Huang Shanghai, China
· **Core Coursework:** C Programming, Linear Algebra, Probability & Statistics, Quantum Mechanics

Publications

** Denotes equal contribution.*

1. **MMA: Benchmarking Multi-Modal Large Language Model in Ambiguity Contexts**
   Selena Song[*], Ru Wang[*], Liang Ding, Mingming Gong, Yusuke Iwasawa, Yutaka Matsuo, Jiaxian Guo
   Accepted by ICLR 2025 Workshop on Navigating and Addressing Data Problems for Foundation Models

2. **Learning Plug-and-play Memory for Guiding Video Diffusion Models**
   Selena Song [*], Ziming Xu [*], Zijun Zhang, Kun Zhou, Jiaxian Guo, Lianhui Qin, Biwei Huang
   Submitted to CVPR 2026

3. **Beyond In-Distribution Success: Scaling Curves of CoT Granularity for Language Model Generalization**
   Ru Wang[*], Wei Huang[*], **Selena Song**, Haoyu Zhang, Yusuke Iwasawa, Yutaka Matsuo, Jiaxian Guo

4. **Beyond Universal Attacks: Automated Diagnosis of Model-Specific Reasoning Vulnerabilities**
   Ru Wang[*], Qi Cao[*], Song Guo[*], **Selena Song**, Yusuke Iwasawa, Yutaka Matsuo, Jiaxian Guo

Research Experience

Learning Plug-and-play Memory for Guiding Video Diffusion Models 2025.05 – Present

Supervised by Professor Biwei Huang of UC San Diego

- Architected DiT-Mem, a retrieval-augmented framework that injects physical-world priors into frozen Diffusion Transformers using a parameter-efficient memory module of 150 million parameters, thereby eliminating the need to fine-tune the backbone generator.
- Formulated a novel frequency-domain analysis on hidden states to theoretically validate the disentanglement of high-frequency physical dynamics from low-frequency visual appearance, leveraging this insight to design a Shared Attention mechanism that guides motion fidelity without compromising semantic coherence.
- Achieved state-of-the-art physical consistency on PhyGenBench by demonstrating substantial reductions in hallucinations across complex scenarios such as fluid dynamics and rigid-body collisions through effective retrieval and encoding of reference video contexts.

Automated Diagnosis of Model-Specific Reasoning Vulnerabilities 2025.03 – 2025.07

- Developed a teacher–student–verifier framework that automatically derives answer-preserving, semantic adversarial paraphrases from the student's own reasoning traces, revealing model-specific shortcut behaviors on math and code benchmarks.
- Built adversarial training pipelines and compared post-training strategies, showing that reinforcement learning on discovered adversarial samples markedly improves the robustness of open-source models on both in-domain and OOD benchmarks, while standard supervised fine-tuning yields little gain.

Scaling CoT Granularity for Language Model Generalization 2024.12 – 2025.03

- Designed a controlled benchmark on three compound reasoning tasks to compare Q-A vs. Chain-of-Thought supervision under distribution shift, revealing that Q-A–only training can reach high in-distribution accuracy yet catastrophically fails out of distribution.
- Developed theoretical and empirical analyses of shortcut learning in transformers, showing that fine-grained CoT and positional-embedding recap substantially narrow the ID–OOD generalization gap with far fewer training examples.

Explore New Knowledge Discovery Ability With LLMs by Entity Decomposition 2024.10 – 2025.03

- Proposed a novel benchmark to evaluate the ability of LLMs to identify and adapt to new knowledge by simulating novel entities and relationships in structured knowledge graphs.
- Introduced methods combining knowledge representation, reasoning techniques, and novel entity construction to advance dynamic knowledge discovery, creating cost-effective data synthesis for evolving benchmarks.

Benchmarking Multi-Modal LLMs in Ambiguity Context 2024.03 – 2024.06

- Designed and constructed the MMA benchmark, the first to systematically test multimodal ambiguity resolution, covering lexical, syntactic, and semantic ambiguities. The benchmark features a unique paired-image design that matches one question with two different images to force reliance on visual context.
- Conducted a comprehensive zero-shot evaluation of 25 MLLMs, revealing a stark performance gap between the average model and a human baseline. The analysis identified critical limitations, including a strong textual bias where models ignore visual context and a significant weakness in resolving syntactic ambiguity.

Evaluating Many-Shot Causal Reasoning and Generalization in LLMs 2024.02 – 2024.05

- Evaluated the many-shot in-context learning and generalization capabilities of LLMs for complex causal reasoning, leveraging the Gemini model's large context window with the CLEAR benchmark dataset.
- The key evaluation tested if the model could generalize knowledge from numerous simple in-context examples to accurately reason about new, more complex causal scenarios.

Enhancing Formal Causal Reasoning in Large Language Models 2023.11 – 2024.02

- **Event Causality:** Evaluated LLMs' ability to identify causal links in stories using the COPES dataset, uncovering significant model hallucinations and issues with dataset annotations.
- **Pure Causal Reasoning:** Demonstrated that LLMs struggle to correctly deduce causal relationships from statistical correlations, with experiments on the CORR2CAUSE dataset showing very poor performance from models.

Agent Model Based Bitcoin Price Prediction and Simulated Annealing Algorithm 2023.02 – 2023.06
Supervised by Professor Jiping Huang of Fudan University

- Designed an agent-based model to simulate individual behaviors and interactions in the market, enabling the prediction of short-term Bitcoin price fluctuations.
- Utilized a simulated annealing algorithm to optimize agents' trading strategies, incorporating factors such as market sentiment and risk-return ratios for improved predictive accuracy.

Option Pricing of Shanghai Composite Index by Bouchaud-Sornette Method 2022.09 – 2022.12
Supervised by Professor Yu Chen of University of Tokyo

- Conducted an option pricing experiment on Shanghai Composite Index options using the Bouchaud-Sornette Method to deduce implied volatility and optimize hedging strategies.
- Compared results with those from the traditional Black-Scholes theory, and provided solutions for practical problems involving non-zero transaction costs.

Skills

Programming & ML:	Python, C/C++, SQL, PyTorch, NumPy, Pandas, Matplotlib, Jupyter Notebook
Tools & Platforms:	Linux, Git, Docker, COMSOL, OriginLab, HTML, LaTeX
Languages:	Mandarin, English, Japanese