Zhixin (Liam) Zeng

Burnaby, BC, Canada | zeng.zhix@northeastern.edu | +1 236-966-0425
github.com/Crimson725 | Google Scholar

Education

Northeastern University — Vancouver, BC, Canada
M.S. in Computer Science; GPA: 3.9/4.0 — Sep. 2023 – Dec. 2024

Sichuan University — Chengdu, China
B.S. in Cybersecurity; GPA: 3.8/4.0 — Sep. 2019 – Jun. 2023

Publications and Preprints

[1] **Zhixin Zeng** and Jiahao Zhong. Cassette-Editor: An Agent Harness for Editing Video the Way You Do. *Preprint*, arXiv:2610.37690, 2026. Latest version on Zenodo.

[2] Wenbin Zhao, Peisong He, **Zhixin Zeng**, and Xiong Xu. Fake News Detection Based on Knowledge-Guided Semantic Analysis. *Electronics*, 13(2):259, 2024.

[3] Yaxuan Wang, **Zhixin Zeng**, and Qijun Zhao. Evaluating the Perceived Safety of Urban City via Maximum Entropy Deep Inverse Reinforcement Learning. In *Proceedings of the Asian Conference on Machine Learning (ACML 2022)*, PMLR 189:1085–1100, 2023.

[4] Yaxuan Wang*, **Zhixin Zeng***, Qiushan Li, and Yingrui Deng. A Complete Reinforcement-Learning-Based Framework for Urban-Safety Perception. *ISPRS International Journal of Geo-Information*, 11(9):465, 2022. (*equal contribution)

Research Experience

Cassette-Editor: An Agent Harness for Video Editing — Jan. 2025 – Present
Lead author and system designer; co-founder of Cassette — Paper [1]

Coding agents get located faults from compilers and a revisable history from version control; a rendered video offers neither. Cassette recasts video editing as reading and writing a document that encodes the timeline's semantics, so a general model can edit video with the abilities it already has as a coding agent.

- Made **the timeline document the action language**: instead of a catalog of per-operation tools, which covers only anticipated edits and composes unpredictably, the agent reads through one tool and writes typed changes to the editor's document through another, so the interface grows with the schema, not a list of tools.
- Gave the agent and the user **one editing path**: a manual gesture and an agent patch become the same typed write, which one executor applies, deriving its consequences for later clips, attached titles, and transitions; one rulebook judges the result and one undo history records it. A write commits atomically or changes nothing.
- Built **the editor as a compiler** for agent edits: a total front end reads any input as the edit it can mean and reports what it left out; the rulebook applies a fix that is unique and keeps the intent, and rejects other blocking violations at their entity and field, with typed repair options and the reads that would supply missing facts. Agent-written motion-graphics programs and shaders compile at commit or are refused at the source line.
- Built a **multimodal evidence layer that follows the edit**: shot descriptions, speech, sound events, and beats, analyzed once in source time, reach edited frames through retiming, reversal, and nested clips using the code shared by validation, preview, and rendering; a subject the agent marks is tracked into the final picture.
- Guaranteed **manual–agent editing parity** by construction: the agent's write format is derived mechanically from the document schema, and each change translates to one canonical change and back, so the agent can express any content change a user can make. Main-storyline gestures are checked against OpenTimelineIO.
- Evaluating each design decision (in progress) through controlled ablations against a basic-tools baseline.
- Built and deployed Cassette (https://app.trycassette.online) as primary developer: browser editor, macOS app with local GPU export, agent runtime, cloud rendering, MCP service, and Final Cut Pro exchange. Released **oh-my-cassette**, an open-source plugin connecting coding agents to the editor.

Research Experience (continued)

Knowledge-Guided Fake News Detection Oct. 2022 – May 2023
Undergraduate Researcher, Sichuan University; B.S. thesis Paper [2]
Advisor: Prof. Peisong He

Investigated how agreement between news claims and external knowledge can improve fake news detection and generalization beyond the training domain.

- Extracted subject–relation–object triplets with REBEL and proposed **fuzzy-matching alignment** to the CommonSense Knowledge Graph, accommodating variations in entity and relation wording without manual labeling or entity linking.
- Independently developed a **knowledge-guided dual-branch model**: aggregated TransE triplet embeddings to guide a BERT text branch, scored claim plausibility in knowledge-embedding space, and fused the branches with learned interaction weights.
- Implemented the full pipeline and conducted all experiments. Training on LUN and testing on the 360-article SLN set yielded **94.45% accuracy and 94.38% macro-F1**, compared with 89.12% and 89.29% for SMHN. On a separate 200-document timing benchmark, average processing time was 0.325 s per document versus 1.257 s for SMHN.
- Established the contribution of external knowledge through ablations: removing the knowledge branch reduced accuracy to **86.28%**; separate ablations confirmed the contributions of the aggregation and scoring modules.

Urban Safety Perception as Human-Like Decision Making Oct. 2021 – Aug. 2022
Undergraduate Researcher, Sichuan University Papers [3, 4]
Advisors: Prof. Qijun Zhao, Dr. Qiushan Li

Modeled perceived urban safety as a sequential decision problem, using crowdsourced judgments to learn interpretable visual preferences from a small dataset.

- Built the **SmallCity dataset**: selected 522 geographically balanced street views from 45,412 Baidu images, stitched four camera headings, and collected balanced pairwise safety judgments through a crowdsourcing website. Converted judgments to scores using TrueSkill and LambdaRank.
- Designed an **interpretable scene representation** with SegFormer after comparing four segmentation models, encoding greenery, sky and building fields of view, wire poles, car count, and visual entropy.
- Formulated safety judgment as an **MDP** in a custom Gym environment and trained PPO, A2C, SAC, and D3QN with Ray RLlib. With an expert-designed reward, D3QN achieved **0.686 AUC**, compared with 0.617 for SVM and 0.550 for a five-layer CNN [4].
- Extended the framework with **maximum-entropy deep inverse reinforcement learning** to recover rewards from crowd judgments. IRL+D3QN achieved **69.3% F1 and 0.709 AUC**, compared with 66.7% and 0.618 for ViT; ablations identified greenery as the strongest factor [3].