Zehan Li

(858) 999-7585 | lizehan@gmail.com | San Diego, CA

EDUCATION

University of California, San Diego CA, USA

- Cumulative GPA: 3.61 / 4 Major GPA: 3.77 / 4

Bachelor of Science in Appplied Math *June 2025(Expected)*

Dual degree: *Bachelor of Science in Cognitive Science specializing in Machine Learning*

- ***Honors Program in Cognitive Science,*** *with* ***Highest Distinction*** *expected upon completion*
- **Area of Specialization:** Computational Linguistics (NLP), Machine Learning, Machine Reasoning, Text Mining, Computational Social Science, Human-Computer Interaction

Minor *in History*

PUBLICATION

- **Zehan Li**, Ruhua Pan, Xinyu Pi, Beyond LLMs: A Linguistic Approach to Causal Graph Generation from Narrative Texts (Accepted to NAACL 2025 Workshop on Narrative Understanding)

RESEARCH EXPERIENCES

Linguistic Approach to Causal Graph Generation from Narrative Texts | UCSD | Under Prof. Zhiting Hu | **02/2024 – 09/2024**

- Utilized linguistic properties, such as time boundness, analyzing causal relationships, improve causal inference accuracy
- ***Independently*** conceptualized and developed an open-source Python-based model for generating causal graphs, achieving 82% interpretability, which is a **9% improvement** over GPT-based directly prompting models (73%)
- Developed a linguistics-based feature extraction tool for causal link identification using *spaCy* and *nltk*, optimizing computational efficiency with **15% increase** in accuracy compared to LLM prompts (based on human evaluation)
- ***Independently*** authored and published a peer-reviewed research paper detailing methods for causal graph generation, focusing on efficient and interpretable causal inference in narrative texts

Quantifying Conceptual Density and vacuity in Text | Honors Thesis | Under Prof. Zhuowen Tu | **08/2024 – Present**

- Project Goal: Develop a quantifiable metric system for evaluating text specificity and vacuity in written communication
- Developing a theoretical model for quantifying conceptual density using linguistics properties as primary dimensions
- Benchmarking outputs from LLMs like to compare their conceptual density and informational richness (In Process)
- Incorporate human cognitive metrics by estimating daily concept and category exposure to understand ties between text specificity and cognitive engagement

Estimating Dinosaur Mass through Machine Learning and Skeletal Analysis | Online | Twoples Project | **05/2023 – 07/2023**

- Combined paleontological methods with modern biological insights to estimate organism mass, advancing the efficiency of dinosaur research. Contributed to interdisciplinary understanding by bridging paleontology and data science
- ***Independently*** created a dataset of 2,000 paleontological specimens by integrating skeletal data from six major museum fossil collections, focusing on vertebral and femoral dimensions as predictive features
- Designed a machine learning model using *RoBerta-Based* Model that achieves 84% prediction accuracy**,** meeting the quality benchmarks of existing 3D modeling-based methods without requiring computationally expensive 3D modelling

SELECTED RESEARCH PROJECTS

Causal Enhancement in Game-Based Reinforcement Learning | UCSD | Under Prof. Zhiting Hu | **03/2024 – 06/2024**

- Integrated causal reasoning and counterfactual thinking into reinforcement learning (RL) algorithms, developing a unified framework for causal discovery, identification, and utilization in policy optimization.
- Focused on gaming environments such as a ***taxi-pickup simulator*** and ***StarCraft II demos*** to evaluate model performance across diverse scenarios, set the basic State Variables and discover the variable change over time
- Achieved a 15% improvement in average rewards compared to ***standard Q-learning*** approaches, and 8% improvement of ***Greedy Exploration Scheme***, with rewards continuing to increase over time due to improved causal generalization

Building QA Model in TOEFL Comprehension Via BERT | UCSD | Under Prof. Zhuowen Tu| **01/2024 – 05/2024**

- Fine-tuned BERT for question-answering tasks, focusing on TOEFL reading comprehension questions to evaluate model

performance in understanding complex queries, balancing high accuracy with reduced computational demands

- Achieved over 70% accuracy on TOEFL-style reading questions, surpassing the average performance of non-native English-speaking high school students, implemented innovative methodologies to enhance BERT's capabilities

Quantifying and Modulating Emotional Expression in Text | UCSD | Under Prof. Leon Bergen | **09/2024 – 12/2024**

- Research on emotion processing, integrating classification, intensity estimation, and text generation into a unified pipeline
- Developed a regression model for emotion intensity estimation, achieving error margin under 0.1 in normalized predictions
- Designed and implemented a ***T5-based*** text generation model for emotion transfer, addressing challenges in controllability and intensity modulation with human evaluations scoring outputs at 5–7 out of 10
- Highlighted key limitations in ***representation engineering*** for emotion transformation, proposing strategies for refining intensity alignment and improving semantic consistency in text generation

Eigenvalue Analysis & Dimensionality Reduction in Handwritten Data | UCSD | Under Prof. Ioana Dumitriu**| 03/2024 – 06/2024**

- Applied *Marčenko-Pastur law* to analyze eigenvalue distributions, identifying noise and signal components
- Conducted *PCA* to reduce data dimensionality, uncovering principal components with clustering patterns
- Visualized eigenvalue spectra and explained variance with cumulative and scree plots to optimize components
- Utilized *PCA* and *SVD* to identify the significant variance contributors, supporting dimensionality reduction decisions

PROFESSIONAL EXPERIENCES

Supplemental Instructor (SI): Calculus and Analytic Geometry | UCSD Math Department**|** **01/2023 – 04/2023**

- Created Lesson plans and supplementary study materials for students in Vector Calculus weekly by weekly.
- Led one 80-minute sessions weekly for 10 weeks, engaging up to 20 students in active problem-solving.
- Created over 50 practice problems to reinforce class concepts, which helped increase student grades by 20 percent

Instructional Assistant: Supervised Machine Learning | UCSD Cogs Department **|** Under Prof. Zhuowen Tu**| 09/2024 – 01/2025**

- Conducted weekly office hours and discussion sessions to deepen student understanding of regression, classification, and neural network concepts, led 80-minute sessions weekly for 10 weeks, engaging over 100 students in problem-solving.
- Provided individualized guidance on complex topics, offered debugging support, and clarified assignments.
- Graded assignments and exams with prompt, detailed feedback, improving student comprehension and learning outcomes

Instructional Assistant: Deep Learning | UCSD Cogs Department **|** Under Prof. Zhuowen Tu**|** **01/2025 – Present**

- Conducted weekly office hours and discussion sessions to deepen student understanding of GPT, Prompt, and deep learning network concepts, led 80-minute sessions weekly for 10 weeks, engaging over 100 students in problem-solving.
- Provided individualized guidance on complex topics, offered debugging support, and clarified assignments.
- Graded assignments and exams with prompt, detailed feedback, improving student comprehension and learning outcomes

HONORS AND AWARDS

Triton Research & Experiential Learning Scholars Award | UCSD Council of Provosts | **Summer 2023, Spring 2024**

Provost Honors | UCSD | **Fall 2021, Spring 2023, Summer 2023, Winter 2024, Spring 2024, Fall 2024**

Finalist, 2023 America Case Competition, Sponsored by Capital One (Over 200 Teams) **– Top 5%**

4th Place, 2023 Student Medicover Case Competition - **awarded $1000**

3rd Place, 2024 BioTechathalon Case Competition (Over 30 teams)

EXTRACURRICULAR ACTIVITIES

President of Media Department | UCSD Chinese Engineering Society | **09/2022 – Present**

*Oversaw content creation and social media strategy; led a team of 15 members, reaching **30,000+** annual views across platforms.*

SKILLS

Programming: Python, Java, CSS, C++, HTML, C, JavaScript, Markdown, R, Latex, Scikit-learn,

Languages: Chinese Mandarin (native), English (proficient), French (Intermediate)

Leadership: Project Management, Team Collaboration, Decision-Making and Problem-Solving, Agile Product management