# Timothy Liu

**Davis, CA**

**(408)-674-4861 | timo.yixin.liu@gmail.com**
**https://github.com/timo-liu | https://www.linkedin.com/in/timothy-liu-866534208/**

Machine Learning researcher with a strong background in NLP, LLMs, and computational linguistics, with hands-on experience in model training, evaluation, and deployment using PyTorch and GCP.

## Skills

Languages: Python, R | Frameworks: PyTorch, TensorFlow, Hugging Face

Cloud: GCP (VertexAI, Model Garden) | Tools: Pandas, Numpy, Arduino, BioSemi

## Research Experience

SEP 2024 - JUN 2025

### UC Davis Computational Linguistics Lab - *Graduate Student Researcher*

- Investigating the impact of tokenization on downstream LLM tasks
    - Tools: PyTorch, NLTK
- Designed and implemented a TensorFlow model for the syllabification of English words
    - Tools: TensorFlow, NLTK, Pandas

JAN 2023 - JUN 2025

### UC Davis Cognitive Neurolinguistics Lab - *Graduate Student Researcher*

- Analyzed orthographic priming effects in real-time speech processing through auditory lexical decision task, finding minimal influences in a heteronymic priming condition
    - Tools: Python, BERT, Neurobs Presentation, R
- Engineered custom hardware and software for a pilot experiment investigating the impact of haptic stimulation on N400 surprisal in an oddball paradigm
    - Tools: Python, Arduino, BioSemi, Neurobs Presentation

## Projects

### eng-syl | https://github.com/timo-liu/eng-syl

An English word and pseudoword syllabifier for use in psycholinguistic research. Provides additional grapheme decomposition and grapheme to phoneme capabilities.

- Created a Python package with TensorFlow for English syllabification and G2P
- Achieved an increase in F1 score (0.91) over Hyphenate (0.88)

### SYLGPT

- Pretrained and evaluated a GPT-2 model trained on syllable and BPE tokenization paradigm and 2 billion tokens of text from the RadioTalk Corpus

- Evaluated both models on PhonologyBench, finding no improvement of syllable tokenization over BPE tokenization

**Heteronyms Honors Thesis**

- Used BERT word embeddings to generate psycholinguistic stimuli investigating how humans process heteronyms
- Ran 22 participants and performed statistical analysis in R revealing negligible priming effects

## Education

SEP 2024 - JUN 2029

**University of California, Davis** - *PhD Linguistics*

SEP 2024 - JUN 2026

**University of California, Davis** - *MA Linguistics*

SEPTEMBER 2022 - JUNE 2024

**University of California, Davis** - *BS Cognitive Science, Minor Computer Science*

## Coursework

Design & Analysis of Algorithms (ECS 222A; 122A)

Artificial Intelligence (ECS 170)

Machine Learning & Discovery (ECS 271)

Advanced Computer Architecture (ECS 201A)

## Awards

2025
Google Cloud Academic Research Grant ($1000)
Provost's Fellowship in the Arts, Humanities and Social Sciences ($60,000)
2024
Glushko Prize ($500)
Departmental Citation in Cognitive Science