PETER ZENG

✉ peterateast@gmail.com | ⎇ peterzeng

EDUCATION

Stony Brook University *August 2023 - Expected May 2027*
Ph.D. in Computer Science Advisor: Owen Rambow
Stony Brook University *August 2017 - May 2022*
BS/MS in Computer Science

EXPERIENCE

Research Assistant *June 2025 - Present*
BIAS-NRT
- Evaluating and improving the ability of VLMs to describe/match specific objects the way a human can.

Research Assistant *August 2023 - May 2025*
IARPA HIATUS
- Introduce interpretability into existing authorship attribution deep-learning models by training them to learn from the residuals of cosine similarities between pairs of interpretable feature vectors of documents.
- This research improves interpretable linguistic based NLP systems by ∼20-30% in accuracy and equal error rate on the task of authorship attribution, all while retaining interpretability.

Software Engineer *June 2022 - August 2023*
Amazon Web Services
- Generated ∼2 billion dollars of yearly revenue for AWS as a member of ∼100 people in the Elastic Containers Service team.
- Wrote and deployed alarm-based rollback deployments, enabling 2 million customer deployments a day to customers in regions around the world.

RESEARCH

Residualized Similarity for Interpretability in Authorship Verification *Spring 2025*
EMNLP Findings 2025, First Author
- We introduce *residualized similarity*, a method of improving performance of cosine similarity-based systems using interpretable features by training a neural network to predict a similarity residual, i.e. the error in the similarity predicted by the interpretable system.

Synthetic Audio Helps for Cognitive State Tasks *Fall 2024*
NAACL Findings 2025, Co-author
- We present Synthetic Audio Data fine-tuning (SAD), showing tasks related to cognitive state modeling benefit from multimodal training on both text and zero-shot synthetic audio data from an off-the-shelf TTS system.
- We show improvement over the text-only modality when adding synthetic audio data to text-only corpora. Furthermore, on tasks and corpora that do contain gold audio, we show our SAD framework achieves competitive performance using text and synthetic audio compared to text and gold audio.

Gram2Vec: An Interpretable Document Vectorizer *Spring 2024*
*Arxiv **Peter Zeng**, Eric Sclafani, Owen Rambow*
- We present Gram2Vec, a grammatical style embedding algorithm that embeds documents into a higher dimensional space by extracting the normalized relative frequencies of grammatical features present in the text. Compared to neural approaches, Gram2Vec offers inherent interpretability based on how the feature vectors are generated. **Views**

Are My Own, But Also Yours: Benchmarking Theory of Mind using Common Ground *Fall 2023*
ACL Findings 2024, Co-author
- We introduce a new theory of mind corpus focused on naturally occurring dialogs, Common-ToM, and show through zero-shot and fine-tuning experiments that LLMs struggle to track beliefs. We then demonstrate that integrating a simple, explicit representation of beliefs improves performance on Common-ToM.

Re-Examining FactBank: Predicting the Author's Presentation of Factuality *Spring 2022*
COLING 2022, Co-author
- We correct Stavnosky's representation of the FactBank (Sauri and Pustojovsky 2009) corpus, invalidating several years of previously published results, and obtain SOTA results for predicting the author's presentation of factuality on four corpora including FactBank.