Statement of Purpose: Carnegie Mellon University
Jared Fernandez (User ID: fern.jared@gmail.com)

I wish to pursue a PhD in natural language processing and machine learning with the goal of developing intelligent agents able to understand and interact with the world. In particular, I am interested in studying the knowledge representations learned by language models and investigating how they can be grounded to better capture abstract and contextual properties.

As an undergraduate at Northwestern, my research spanned topics in natural language processing and human perception, culminating in three publications including two as a first-author. My research has investigated: properties of word embeddings, representations of visual texture, training techniques for language models, and failures modes of commonsense reasoning. Working on these problems, I was fascinated by the capacity of models to learn both textual and visual semantics. In my doctoral program, I hope to further explore these topics while building machine learning systems with human-level intelligence, able to address tasks such as commonsense reasoning or the understanding of human intents.

I first developed my interest in language representations while analyzing the properties of word embeddings with Professor Douglas Downey. Neural models frequently use pretrained embeddings to improve generalizability and reduce training time. However, downstream performance is often dependent on how well the word representations contained in these embeddings match those in the task's corpus. In a first-author paper at EMNLP, I developed the VecShare framework to enable easy selection of embeddings through rapid similarity comparisons between embeddings and a task's raw text corpus [1]. I introduced embedding signatures which are compact representations for full sets of word embeddings, consisting of a combination of embedding properties (i.e. word similarities) and corpus-level statistics (i.e. word frequencies). By computing similarities between embedding signatures and a corresponding signature constructed from the task's corpus, we can efficiently determine the most relevant embedding for a particular downstream task. In my experiments on document classification, convolutional neural networks trained with embeddings selected by the VecShare framework yielded significant increases in model accuracy. Alongside my experimental results, I released an open-source library for selecting word embeddings online.

I also studied the representations used to model human perception while designing similarity metrics for image texture with Professor Thrasyvoulos Pappas. To examine the latent space of image textures, I developed a novel Structural Texture Similarity Metric (STSIM) using features extracted from an image's steerable filter decomposition. My metric computes the Mahalanobis distance between STSIM features by utilizing a novel formulation of the covariance matrix that is robust to both intraclass variance and correlations. In experiments on texture classification, my method is competitive with state-of-the-art neural approaches and obtains near full accuracy [4]. Furthermore, my proposed metric provides increased interpretability, reduced training time and requires significantly fewer parameters than comparable deep models. Together, my experiences with both word embeddings and texture features have cultivated my intuition for developing both learned and designed features for machine learning models.

In addition to studying the properties of various representations, my research has also examined active learning-based data selection techniques for efficiently learning those features. Recurrent neural network language models (RNNLMs) observe strong performance on sequence modeling tasks but suffer from high computational costs. To determine whether RNNLM's training costs can be reduced without losses in performance, I designed several weighted importance sampling distributions to distill large training corpora to a core-set of high information sequences. Informative sequences were sampled with increased likelihood based on the sequence's perplexity, as measured by an *n*-gram language model trained on starter data. The *n*-gram models's rapid training and inference times allowed it to serve as an efficient proxy for computing RNNLM perplexity, which enabled us to approximate loss-based sampling rapidly. Models trained on sequences sampled with my proposed distributions observed substantial perplexity reductions in comparison to both *n*-gram and neural baselines. I published these results in a first-author paper at the Association for Computational Linguistics: Student Research Workshop [2]. From this project, I gained invaluable knowledge training large networks and an understanding of how they learn complex distributions.

More recently, I am excited by the challenge posed by commonsense reasoning, an especially difficult task since commonsense knowledge is rarely explicitly stated in text. In work published at the RepEval Workshop [3], I examined failure modes of neural common sense question-answering systems through the construction of CODAH, a challenge dataset for commonsense. The CODAH dataset consists of questions adversarially authored by human annotators educated about QA models and directed to write questions that could fool state-of-the-art models, even after extended retraining. While state-of-the-art BERT models have achieved superhuman performance on the benchmark SWAG dataset, we found that commonsense still poses challenges to these large transformer-based models. We identified a few notably challenging areas of reasoning in CODAH, namely quantitative reasoning and negation. Presently, I am contributing to ongoing work in automatically generating questions for commonsense utilizing transformer-based models rather than relying on human annotators. In the future, I hope to explore how representations can be grounded with contextual information and implied relationships to address some of these outstanding challenges.

Since graduating from Northwestern, I have continued to work on embedding techniques for natural language while developing scalable models as a Software Engineer at Google. With the Dialogflow NLU team, I worked on enterprise chatbot agents and developed domain adaptation techniques for the neural models used to produce sentence-level embeddings. These fine tuned representations were better able to handle conversational text's short form and syntax, yielding substantial increases in intent recognition performance. Seeing the effect of better language representations on practical applications has further invigorated my desire to pursue a career in research and my excitement to improve machine learning models for natural language.

In my doctoral program, I intend to investigate how we can develop agents capable of reasoning about complex scenarios, whether human behaviors or real world environments. My experience analyzing both linguistic and perceptual features, along with my current work in commonsense, provide both a strong background and immediate intuition for tackling these problems. Consequently, I am interested in the work of Professor Yonatan Bisk on both physical and visual commonsense reasoning. Likewise, I am excited by Professor Louis-Philippe Morency's research on multimodal models to incorporate the wide array of non-textual signals into the reasoning process. Furthermore, I am fascinated by Professor Ruslan Salakhutdinov's work to improve learned representations through new model and layer architectures. At Carnegie Mellon, I see a clear alignment of my interests and previous research experiences that will enable me to make strong contributions to these groups.

References

1. **Jared Fernandez**, Zhaocheng Yu, Doug Downey. *VecShare: A Framework for Sharing Word Representation Vectors*. In Empirical Methods in Natural Language Processing (EMNLP), 2017.
2. **Jared Fernandez**, Doug Downey. *Sampling Informative Training Data for RNN Language Models.* Association for Computational Linguistics: Student Research Workshop (ACL-SRW), 2018.
3. Michael Chen, Mike D'Arcy, Alisa Liu, **Jared Fernandez**, and Doug Downey. *CODAH: An Adversarially Authored Question-Answer Dataset for Common Sense*. Proceedings of the 3rd Workshop on Evaluating Vector Space Representations for NLP (RepEval), 2019.
4. **Jared Fernandez**, Thrasyvoulos Pappas. "Structural Texture Similarity Metric Based on Intraclass Covariances". Under Preparation.