Personal, Background, and Future Goals Statement

Humans rely on vast amounts of background knowledge to communicate about their environment, goals, and actions. By utilizing a wealth of priors, humans are able to resolve references to the objects in their environments and perform higher-level logical reasoning. Recent advances developing massive neural network architectures have yielded success across a breadth of natural language processing benchmarks. While these models excel in emulating structural and stylistic patterns in language, models learned in language-only settings struggle to capture the physical properties and spatial relationships of real world situations. As a result, substantial performance gaps remain on complex tasks that require models to resolve and understand broad world knowledge - such as in commonsense reasoning and language-based navigation. To inform models with the concepts and relationships missing from text-based approaches, I will explore techniques for grounding language in the world knowledge that comes from perception and other modalities.

With the support of the NSF GRFP, I will apply my research to build new technologies to increase the accessibility of artificial intelligence applications for both practitioners and the general public. As an active member of the research community I will: (1) design efficient training procedures to reduce the computational and financial overhead of machine learning , (2) enrich the reasoning capabilities of machine learning models to enable new accessibility applications at the human-computer interface, and (3) engage students of all backgrounds in discussions about careers in artificial intelligence.

Intellectual Merit My interest in the computational efficiency of multimodal models is motivated by two key challenges that I encountered during my undergraduate research. From my research on commonsense reasoning, I confronted the lack of world knowledge in text and the resulting inability of models to learn higher level relationships and properties. Moreover, while investigating efficient training algorithms for neural language models, I gained an awareness for the enormous computational costs that accompany model training and inference. Together, these challenges pose impressive barriers that prevent natural language systems from being able to perform computations in real world settings. Motivated by these problems, I developed my interest in computationally efficient techniques for multimodal machine learning for language grounding. As a graduate student, I will leverage the intuition that I developed from these projects, alongside the machine learning fundamentals that I acquired from my other experiences in natural language processing (NLP) and computer vision to tackle these problems.

My interests in language grounding emerged largely from my recent projects exploring the acquisition of commonsense reasoning capabilities by neural network language models with Professor Doug Downey at Northwestern University. In these projects, I gained insights into the challenges of imbuing models with the necessary information to make predictions and reason about real world settings. In particular, commonsense reasoning poses a difficult problem for machine learning models because it requires models to resolve long term textual dependencies and then utilize priors about commonsense properties and behaviors. To explore the limitations of state-of-the-art common sense question-answering systems, I constructed the CODAH commonsense challenge dataset and performed evaluations of transformer-based neural models. The CODAH dataset consists of questions adversarially authored by expert human annotators educated about state-of-the-art models and directed to write questions that could fool these models, even after extended retraining. Although state-of-the-art BERT models have achieved superhuman performance on the benchmark SWAG dataset, we found that even with added fine-tuning, CODAH still poses a challenge to the these models with performance lagging human-level by over 30%. Among the questions in the CODAH dataset, we identified a few notably challenging areas of reasoning, namely quantitative reasoning and negation. The CODAH dataset and our accompanying analysis were published as a paper at the *RepEval Workshop* [3].

While the difficulty of CODAH indicates that commonsense reasoning is still an unsolved task, development of models for this task is further complicated by the limited amounts of training data for commonsense. In comparison to other unsupervised NLP tasks, sourcing commonsense training data is especially difficult and costly because commonsense knowledge is largely assumed and rarely stated explicitly in text or web data. To provide models with additional training data, I developed the G-DAug technique to generate informative and diverse sequences for augmenting commonsense datasets. In G-DAug, multiple choice completion questions are generated by two pretrained OpenAI GPT-2 models with one one outputting the correct sequence and the second generating incorrect distractor answers. Generated questions are then filtered for diversity and quality by sampling questions according to their number of unique unigrams and their expected utility as predicted by an influence function. In our evaluations, we found that supplementing model training with the data generated by G-DAug increases performance across nearly all commonsense benchmarks. Our work on G-DAug was recently published in *Findings of the ACL: EMNLP* [4]. Between my work on CODAH and G-DAug, my experiences highlighted the potential of machine learning systems to gain higher level reasoning abilities. The excitement for such systems is the catalyst for my interest in building models with complex world representations, and motivates my current interests in jointly learning these models over both language and vision as a means to overcome the limitations I encountered in the text-only setting.

In addition to sparking my interest in language grounding, my undergraduate research with Professor Downey aroused my interest in the computational efficiency of these algorithms. While neural network models frequently achieve state-of-the-art performance, these models suffer from heavy computational costs that often require GPUs or other specialized hardware. To reduce the training costs of recurrent neural network language models while seeking to avoid losses in performance, I developed several weighted importance sampling distributions for distilling large training corpora to a core-set of highly informative sequences. These distributions were designed to encourage sequence diversity and utility by sampling sequences 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. The results of this work culminated in a first-author paper, which I presented at the *ACL: Student Research Workshop* [2]. Working to design efficient training algorithms, I am prepared to account for the computational considerations of machine learning development and deployment and will ensure that my work can be applied in practical settings.

Additionally, as both a researcher and industry engineer, I studied the effectiveness of using pretrained word embeddings to improve the performance of downstream NLP systems. In work published at *EMNLP*, I found that by aligning the domain of the pretrained word embeddings to match that of a target downstream document classification task improved performance by an average of 2.60% [1]. To take advantage of these gains, I developed techniques for efficiently computing the similarity of a word embedding set and a raw text corpus to rapidly determine the most relevant set of embeddings. Although it is possible to learn domain-specific embeddings directly, the models used to generate embeddings are trained on large amounts of general web text which is often inappropriate to the downstream task. As a software engineer at Google, I developed new domain adaptation techniques to fine tune our embedding representations for use in conversational dialogue systems. By performing extended pretraining on spoken word queries, the resulting embeddings were better able to handle conversational text's short form and syntax and resulted in substantial increases in intent recognition performance. In future research, I will employ the intuition I gained working on embeddings to inform my modeling decisions through better choice in features and training objectives.

Complementing my experiences in natural language processing, I also developed a background in working on feature representations used in computer vision and human perception. In work with Professor Thrasyvoulos Pappas, I studied the representations and similarity metrics for image texture, I developed a novel Structural Texture Similarity Metric (STSIM) based on a the Mahalanobis distance between the features extracted from an image's steerable filter decomposition. In experiments on texture classification, my method is competitive with state-of-the-art neural approaches and obtains near full accuracy. As these features were extracted using signal-based processing techniques, they provide increased interpretability, reduced training time and orders of magnitude fewer parameters in comparison with similar neural models.

In my upcoming work on multimodal models, I will utilize the familiarity I gained working in the visual domain to supplement my background in NLP in determining the best methods for learning across these areas.

Broader Impacts In my research, I intend to improve the capacity of machine learning models to perform grounded language understanding so that humans will be able to communicate with intelligent devices. Although machine learning has led to the advent of a wide range of new technologies, they largely remain inaccessible to the general public due to high technical barriers to entry. Efficiency improvements for language models will enable the proliferation of these technologies on edge devices, such as mobile phones or robots, that can be used by any human. Furthermore, improvements in the grounded language understanding capabilities of these models will enable humans to interact directly with these systems about real world situations without needing specialized technical knowledge. By constructing a natural way for humans to communicate with machines and reducing the hardware requirements for the underlying models, I will make artificial intelligence applications more accessible for the general public.

To increase the accessibility of artificial intelligence research and its applications, I have consistently engaged in outreach and direct mentorship to underrepresented populations. Although artificial intelligence as a field has recently observed rapid growth, traditionally underrepresented groups continue to be marginalized through lack of resources to access professional opportunities. Without the input of people from all backgrounds, future algorithms and applications will remain biased towards the needs of those developing them and will not reflect the needs of all people. As a member of the Eta Kappa Nu Honor Society at Northwestern University and more recently with the AI Undergraduate Mentorship Program at Carnegie Mellon University, I have worked with students from underrepresented groups to connect them with resources and to build their network. To help provide students with the necessary technical background, I led the redesign of the IE3 Technical Program as a member of the Executive Board of Northwestern University's IEEE Student Chapter. In the IE3 Technical Program, we paired undergraduate mentors with students of all backgrounds to develop a software project from start to finish in a small group setting. In conjunction with the program, I assisted in the organization of our annual project showcase to provide an avenue for students to build their professional network and receive recognition for their work from industry representatives. Over the course of my three years managing the IE3 Program and project showcase, we observed a three-fold increase in student involvement with over 50 students participating in the most recent iteration. In graduate school and beyond, I will continue to build a more diverse research community with the goal of building technologies that better address the problems of people across the world.

Future Goals With the support of the NSF GRFP, I will have freedom to pursue a career in machine learning research and explore exciting new directions of artificial intelligence research that will lead to far reaching impacts in machine learning accessibility and applications. I will impart a positive impact on society by developing new systems at the interface of language and vision that is able to both interpret and converse about their environments. Furthermore, by focusing on both the training and inference-time efficiency of these models, I will ensure that such systems are accessible to the majority of artificial intelligence and machine learning practitioners. In addition to affecting positive change to the scientific community with my research, I will also continue my efforts to improve on the overall accessibility of scientific research by organizing initiatives for outreach and informing the general public of pathways to get involved in research.

[1] J. Fernandez et al. "Vecshare: A framework for sharing word representation vectors". In: *EMNLP*. 2017.

[2] J. Fernandez et al. "Sampling Informative Training Data for RNN Language Models". In: *ACLSRW*. 2018.

[3] M. Chen et al. "Codah: An adversarially-authored question answering dataset for common sense". In: *Workshop on Evaluating Vector Space Representations for NLP*. 2019.

[4] Y. Yang et al. "Generative Data Augmentation for Commonsense Reasoning". In: *ArXiv*. 2020.