Personal Statement: Ethan Perez

I want to develop algorithms that understand language and reason about the concepts language describes: abstract ideas, the visual world, actions and goals, and more. In this way, I hope to enable beneficial AI applications such as intelligent web and image search, chatbot assistants, and robots that understand verbal instructions. To tackle these traditional natural language processing tasks, I am drawn to algorithms whose performance scales with the world's growing quantity of data: machine learning and, more specifically, *deep learning* and *reinforcement learning*.

I first became interested in language and reasoning last summer while researching in visual reasoning at the University of Montreal with Dr. Aaron Courville. Dr. Courville and I wanted to build a model that learned to visually reason purely from images, questions, and answers, unlike prior work at the time. To this end, we chose CLEVR, a dataset from FAIR consisting of complicated, multi-step, image-based questions such as "Are there more blocks in front of the large purple cube than yellow metal blocks?" I investigated various methods, alternating among exploring existing literature, bouncing ideas with colleagues, and testing the ideas I found most promising.

Through this process, I found a surprising result. With a general-purpose conditioning layer, a standard convolutional network could learn state-of-the-art visual reasoning, going against the trend of reasoning-specific models from DeepMind, FAIR, and Berkeley. I subsequently wrote up the results with a few other students into a *first author workshop paper at the International Conference on Machine Learning (ICML) 2017 [1]* and later a *first author paper at the Association for the Advancement of AI (AAAI) 2018 Conference [2]*.

Publishing and presenting this work opened the door for me to get to know exceptional AI researchers. Besides ICML, I received the chance to give invited talks in Montreal, and I enjoy regular emails with researchers excited about extending this work. DeepMind worked to reimplement my model, and Facebook AI Research used my method in later work [3]. The community's interest brought me into conversations with researchers whose work I admired and critics whose perspectives were eye-opening. I thus learned both how rewarding it is to be a part of a dynamic intellectual community and how much I have to learn, motivating my interest in a career in research.

Excited about avenues for future work in language understanding, I took a semester off from my undergrad to continue my internship and pursue my ideas with Dr. Aaron Courville and Dr. Hugo Larochelle, a research scientist at Google Brain. In particular, I intuited that my findings in visual reasoning might translate to language instruction-following in reinforcement learning (RL), as both benefit from learning language grounded in visual features. Thus, I helped develop HoME, a **Ho**usehold **M**ultimodal **E**nvironment to build instruction-following agents ("Go to the large leather couch" or "Find the wooden sink in the kitchen") in a richer context and vocabulary than prior work. This ongoing work resulted in a *second author workshop paper at the Neural Information Processing Systems (NIPS) 2017 Conference [4]*. With this more challenging and realistic testbed, I plan to develop RL methods for grounding language and reasoning via interaction.

Extending my internship has given me the opportunity to assist others building on my work. I am helping with a Distill literature review tying together numerous deep learning methods, including the conditioning layer I introduced. A few other students and I are applying insights from my work to machine translation and visual (image-based) dialogue. These broader experiences have shown me how fascinating I find machine learning, language understanding, and reasoning, motivating me to take a FAIR internship next summer and to pursue a Ph.D. to gain long-term, in-depth expertise in these areas.

What led me to pursue research in Montreal was my earlier research at Rice with Dr. Ankit Patel. Here, I helped develop a state-of-the-art semi-supervised deep model for image classification, leading to a *third author paper to be submitted to ICML 2018 [5]*. Outside academia, I enjoyed industry internships which sparked my original interest in machine learning. At Google Maps, I researched various models for accurately localizing users via Bluetooth signals, and at Uber, I developed models to detect fraudsters attempting to take rides on other users' accounts. I thus learned how to develop and iterate on ideas cleanly and quickly, accelerating my research, and I grew interested in more deeply exploring machine learning via research.

[**School-specific paragraphs below.**]

Studying a Ph.D. at [Stanford], I could learn how to advance machine comprehension and reasoning from its leaders. Faculty such as *Professors Manning, Liang, Jurafsky, and Li* develop ground-breaking datasets like SNLI, SQuAD, SCONE, and CLEVR and techniques for sentence summarization (pointer-generator networks with coverage), using knowledge graphs (via compositional training), dialogue (via deep RL), and language grounding (iteration-based scene graph generation), for instance. Working with these experts, whose listed work has shaped my research interests, I would equip myself to advance the cutting-edge in these areas throughout my Ph.D. and lifetime.

Studying a Ph.D. at [NYU], I could learn how to advance language understanding and reasoning from its leaders. Faculty such as *Professors Kyunghyun Cho, Jason Weston, Rob Fergus, and Sam Bowman* develop groundbreaking language modeling techniques such as Gated Recurrent Units, TransE, memory networks, and Variational Autoencoders for natural language generation. Furthermore, these faculty push machine learning to solve qualitatively different challenges, introducing tasks in emergent translation, interactive learning (Mastering the Dungeon and Learning by Asking Questions), reasoning (bAbI), and natural language inference (SNLI and MultiNLI). Working with these experts, whose listed work has significantly influenced my research interests, I would equip myself to advance the cutting-edge in these areas throughout my Ph.D. and lifetime.

Studying a Ph.D. at [MIT BCS], I could learn to advance machine comprehension and reasoning from the leaders of a distinct, invaluable perspective: human self-examination. Faculty such as Professors Tenenbaum, Poggio, and Levy bring ideas from instrinsic motivation to hierarchical RL, from child learning to machine learning, from the visual cortex to recurrent neural networks, and language context to grammar correction, for instance. Working with these advisors to understand how humans learn language and reasoning and translating these findings to machines, I would learn to bring qualitatively different insights to AI throughout my Ph.D. and lifetime.

Studying a Ph.D. at [MIT EECS], I could learn how to advance language understanding and reasoning from leaders in these areas such as Professors Barzilay, Jaakkola, and Torralba. Reading these professors' work, I have found them keen on pushing what is qualitatively possible with machine learning, such as improving reinforcement learning game-play by learning from game manuals, transferring linguistic style using non-parallel corpora alone, and grounding language in cross-modal representations without paired data. Working with these experts, I would equip myself to advance the cutting-edge in language understanding and reasoning throughout my Ph.D. and lifetime.

Studying a Ph.D. at [CMU], I could learn how to advance machine comprehension and reasoning from its leaders. Faculty such as *Professors Ruslan Salakhutdinov, William Cohen, and Graham Neubig* have developed novel architectures for grounding language in visual percepts and actions, for multi-hop reading comprehension, for enhancing language understanding via knowledge graphs and external knowledge, and for adversarial representation learning. Working with these experts, I would equip myself to advance the cutting-edge in language understanding and reasoning throughout my Ph.D. and lifetime.

Studying a Ph.D. at [UW], I could learn how to advance machine comprehension and reasoning from its leaders. Faculty such as Professors Luke Zettlemoyer and Ali Farhadi have driven breakthroughs in grounding language in vision and interaction via instruction-following, interactive question-answering, and object detection. These faculty, along with Professor Noah Smith, have also spurred advances in more traditional language research areas such as machine comprehension (BiDAF, TriviaQA) and word representations (multilingual, hierarchical, and sparse embeddings). UW's further close ties with the Allen Institute for AI, whose research goals in machine reading and reasoning directly align with my own, will open up further access for me to collaborate with knowledgeable field experts. Working in this environment, I would equip myself to advance the cutting-edge in language understanding and reasoning throughout my Ph.D. and lifetime.

Studying a Ph.D. at [Berkeley], I could learn how to advance language understanding and reasoning from leaders of diverse aspects of these areas. *Professors Klein, Levine, and Darrell* have led exciting breakthroughs relevant to my interests, introducing module networks , language into latent representations , and improved techniques for grounding language in vision and translating multi-agent communication . More broadly, these faculty each have expertise in machine learning and deep learning, which, with the world's growing data and compute, are promising and scalable long-term approaches to machine intelligence.

References

[1] **E. Perez**, H. de Vries, F. Strub et al. *ICML 2017 Workshop*. Learning Visual Reasoning Without Strong Priors.

[2] **E. Perez**, F. Strub, H. de Vries et al. *AAAI 2018*. FiLM: Visual Reasoning with a General Conditioning Layer.

[3] I. Misra, R. Girshick, R. Fergus et al. *arXiv 2017*. Learning by Asking Questions.

[4] S. Brodeur, **E. Perez**, A. Anand et al. *NIPS 2017 Workshop*. HoME: a Household Multimodal Environment.

[5] T. Nguyen, W. Liu, **E. Perez** et al. *arXiv 2017*. Semi-Supervised Learning with the Deep Rendering Mixture Model.

[6] R. Jia, P. Liang. *EMNLP 2017*. Adversarial Examples for Evaluating Reading Comprehension Systems.