Statement of Purpose

Zhengxuan Wu, Symbolic System M.S. Internal Applicant, Fall 2020

Motivations

I am pursuing a masters degree in Symbolic System at Stanford University to further broaden my knowledge in computer science as well as cognitive psychology before starting my next academic journey in a more advanced degree in cognitive articial intelligence (AI). From Google Home to Tesla Autopilot, AI is progressing rapidly. To build the next generation of AI, I believe that it is essential for AIs to have humanlike cognitive awareness of surroundings. In this way, AIs can collaborate with humans; i.e., they can understand our intentions and reward systems as individuals and our ability to attribute mental states to others (as in theory of mind[1–3]). To achieve this goal, I am interested in learning to build cognitively inspired AI models, and interpret AI models to infer knowledge about the cognition process of our brain. The Symbolic System program is especially appealing to me given Stanfords reputation for having highly collaborative Symbolic System departments, which is critical, given the interdisciplinary nature of research in cognitive AI systems.

Experience in Fundamentals

Bolstered by my past industrial and research experiences in cognitive psychology and computer science, I am fortunate to possess solid foundations needed to approach problems in elds bridging cognitive science and AI. During my current masters degree at Management Science & Engineering, I was fortunate to work with researchers in psychology, cognitive science, social science, and computer science. Within my broad research spectrum, all of my experiences centered on reasoning about humans using computational models that maps well to the department goals of Symbolic System, i.e., bridging human and machine.

Reasoning of Intuitive Psychology Under the guidance of Dr. Desmond Ong and Dr. Jamil Zaki at Stanford, my current research focuses on enabling AI to have humanlike cognition of emotion and inferring human cognition process of emotions from trained AI models. Our topics include (1) predicting emotional states of others with multimodal inputs, (2) reverse engineering how humans intuitively reason about other people from trained parameters and comparing that with brain activities, and (3) codifying such reasoning via probabilistic modeling: a humanlike approach that involves both symbolic encoding for knowledge representations and hierarchical reasoning using Bayesian inferences. This extended journey has culminated in the publication of our Stanford Emotional Narratives Dataset on IEEE Transaction on Affective Computing, as well as the building of a Transformer-based memory fusion network model and variational neural networks to accurately predict emotional states of humans, at the Affective Computing Intelligent Interaction Conference with me as the rst author.[4,5] Currently, we are expanding our research in understanding how people and computer models attend to different cues (i.e., language, visual, acoustic) in understanding emotional states of others. By understanding human cognition, I am condent that researchers can build AI agents capable of building robots that understand the internal states of others and that are fair and safe while interacting with humans and other agents.[6,7]

Reasoning of Human Behaviors Through computational models with proper design and training through persuasive systems, we can better understand the human cognition process, which can further be used to inuence human perception and behavior collaboratively.[8,9] With this in mind, I was fortunate to work on HabitLab, a project led by Dr. Geza Kovacs and Dr. Michael Bernstein at Stanford. HabitLab is a chrome browser extension that contains a variety of productivity interventions aiming to reduce the time spent on user-specied websites or applications. Leveraging in-the-wild experiences with online interventions, we used this platform to study and inuence user behaviors in a naturalistic way. I supported these efforts by helping with outlining the interplay between efficacy and intervention attrition rates. To optimize efficacy, I also helped in building adaptive interventions that are optimized for individuals using a multiarmed bandit algorithm. We also looked at the conservation of procrastination across multiple devices. Specically, we investigated whether productivity interventions can actually help users to save time or just redistribute it across devices. Additionally, we investigated changes in user motivation over time, as observed through the lens of intervention difficulty levels, and submitted our paper to CHI2020. Through this study, I have contributed to two papers that were published in CSCW 2018 and CHI 2019 as full papers.[10,11] Through this valuable experience, I learned how AIs can not only better understand our behaviors and meet our needs but also cooperate with humans to augment our intelligence.

Reasoning of Social Psychology To discover how AIs and computational models in general can help humans understand cognition beyond the individual level, I was fortunate to work with Dr. Michal Kosinski in the eld of computational social science, endeavoring to understand social cognition using digital footprints. Digital footprints can predict social traits, including sexual orientation.[12] Our study has focused on the differences in social traits between heterosexuals and homosexuals, using the data mining and deep learning methods on the Facebook datasets from the myPersonality website. We concluded that masculinity-femininity scores can predict sexual orientation in males, but not in females. We illustrated how our work aligns with previous psychology experiments that elucidate the power of digital footprints in studying social cognition13 and that are close to submission of paper to a psychology journal paper.[13] From these studies, I not only discovered the power of digital footprints in understanding human social traits but also became aware that this power comes at the cost of privacy.12 It is with this philosophy in mind that I would like to build intelligent and safe AI systems.

Besides these experiences, I also participated in projects related to probing semantics of emotions with word embedding,[14] modeling interpersonal emotion differentiation,[15] modeling facial movements that imply emotions,[16] reasoning of morality, and analyzing social networks of Chinese politicians,[17] which all led to publications and manuscripts in preparation. Each of these challenging yet rewarding projects has broadened my research spectrum, but also affirmed that I want to work on building AI systems with human-like cognition and ability to learn.

Interests

If admitted to the program, I would be keen on expanding my current interdisciplinary research project in affective computing with Prof. Jamil Zaki at Stanford University and Prof. Desmond C. Ong at National University of Singapore. I will be receiving advices from both professors in terms of cognitive psychology and computer science.

My proposed project area includes two main sub-domains, and each of them complements one another. One of the sub-domains is building cognitive psychology experiments and computational models to understand how people reason about emotions and mental states of others. For instance, how people integrate different cues (e.g., voices, languages, facial expressions) during communications with others to infer underlying emotions? The other sub-domain is building interpretable AI algorithms to help us gain a better understanding of how the brain works in interpreting emotions of others. Furthermore, leveraging with those commonalities found in AI algorithms and brains, I would be keen on building cognitively inspired AI models. For example, I would like to continue my current project in understanding how attention mechanism works both in AI algorithms and brains in emotional cognition with different cues. Additionally, I hope to connect the dots between AI algorithms and brains with comparing cues that they attend to. Combining advantages from both domains and the interdisciplinary nature of this program, I would be thrilled to develop innovative paradigms to understand how the brain works in understanding emotional states of others and to enable AI systems to have humanlike cognition of emotion.

In conclusion, I believe that my education and experiences so far have prepared me well to take the next step in my career, by completing Stanfords Symbolic System masters program. It is my goal to undertake research that promises to make a safe, robust, and reliable difference for our foreseeable future in AI. I deeply believe that, to deliver such difference, we need to enable AI systems to have humanlike cognition of surroundings. I came across this program while reading a personal story about this program on Quora from a recent graduate Lucy Li (now studying Ph.D. at University of California, Berkeley). Her words, The interdisciplinary nature of this program in cognitive science and computer science opened up many new doors and research ideas for her in building AI systems, sparked my interest in joining the program in this AI era to gain knowledge in different elds. The curriculum provided by Symbolic System would enable me to master the needed skills for conducting my own research in the future. My robust interdisciplinary perspective gained in cognitive psychology and computer science prepares me to contribute to the community in the program. In turn, I believe the rich research community at Stanford will provide me with invaluable training in both cognitive science and computer science, along with a cohort of exceptional peers. Thank you for your consideration.

References

[1] Henry M Wellman. *The child's theory of mind*. The MIT Press, 1992.

[2] Josef Perner. *Understanding the representational mind*. The MIT Press, 1991.

[3] David Premack and Guy Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, 1(4):515–526, 1978.

[4] Desmond C Ong, Zhengxuan Wu, Tan Zhi-Xuan, Marianne Reddan, Isabella Kahhale, Alison Mattek, and Jamil Zaki. Modeling emotion in complex stories: the Stanford Emotional Narratives Dataset. *IEEE Transactions on Affective Computing*, to appear.

[5] Zhengxuan Wu, Xiyu Zhang, Tan Zhi-Xuan, Jamil Zaki, and Desmond C. Ong. Attending to emotional narratives. *IEEE Affective Computing and Intelligent Interaction (ACII)*, 2019.

[6] Andrea Bajcsy, Dylan P Losey, Marcia K O'Malley, and Anca D Dragan. Learning from physical human corrections, one feature at a time. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 141–149. ACM, 2018.

[7] Andrea Bajcsy, Dylan P Losey, Marcia K OMalley, and Anca D Dragan. Learning robot objectives from physical human interaction. *Proceedings of Machine Learning Research*, 78:217–226, 2017.

[8] Brian J Fogg. Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002(December):5, 2002.

[9] Harri Oinas-Kukkonen and Marja Harjumaa. Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems*, 24(1):28, 2009.

[10] Geza Kovacs, Zhengxuan Wu, and Michael S Bernstein. Rotating online behavior change interventions increases effectiveness but also increases attrition. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW), November 2018.

[11] Geza Kovacs, Drew Mylander Gregory, Zilin Ma, Zhengxuan Wu, Golrokh Emami, Jacob Ray, and Michael S Bernstein. Conservation of procrastination: Do productivity interventions save time or just redistribute it? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 330. ACM, 2019.

[12] Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.

[13] Zhengxuan Wu and Michal Kosinski. Homosexual women are not masculine, 2019.

[14] Zhengxuan Wu and Yueyi Jiang. Disentangling latent emotions of word embeddings on complex emotional narratives. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 587–595. Springer, 2019.

[15] Erik Nook, Christina Chwyl, Isabella Kahhale, Zhengxuan Wu, and Jamil Zaki. Interpersonal emotion differentiation, 2019.

[16] Alison Mattek, Michael Smith, Zhengxuan Wu, Isabella Kahhale, Marianne Reddan, Desmond Ong, and Jamil Zaki. Modeling facial movements that track emotion inference, 2019.

[17] Yueyi Jiang, Zhengxuan Wu, Arseny Ryazanov, and Piotr Winkielman. Social influence shifts gamble preferences for monetary and moral decisions, 2019.