Towards Real-World Social AI

Paul Pu Liang

As Artificial Intelligence (AI) increasingly blends into our everyday lives, building systems that display social intelligence has become one of the next grand challenges. **Socially intelligent AI** should comprehend human social cues, intents, and affective states, engage in social conversation, and understand social norms and common-sense in order to maintain a rich level of interaction with humans. Humans naturally communicate using a coordinated structure of multimodal signals spanning language, visual, and audio. While AI has shown tremendous promise in solving various tasks, designing social AI with the capability to communicate with humans, by incorporating all involved modalities, is a fundamental challenge.

My research builds towards social AI with the long-term goal of engaging people through social and physical interactions [7], monitoring human behavior to understand and predict the types of help people need [1, 3], and offering assistance in schools, hospitals, and the workplace [5]. While prior research has made impressive strides in affective computing and dialog systems, my research specifically focuses on bridging the gap towards real-world deployment by improving robustness, fairness, and interpretability. These advances will improve the accessibility of social AI, particularly for under-represented groups.

As steps towards real-world social AI, I have outlined three major milestones:

1. **Multimodal perception of human communication** through understanding social cues, intents, affective states, personalities, and references to the broader environment.

I have worked towards addressing the lack of multimodal resources by releasing the largest dataset of multimodal sentiment and emotion recognition enabling generalizable studies of human communication [12, 2]. On the modeling side, my work in characterizing the desiderata for multimodal representations beyond discriminative performance [11], data and compute-efficient multimodal learning [4], and modeling both person-independent and person-dependent signals for affect analysis [3] have substantially improved the state-of-the-art in the field.

2. Modeling the **long-term interactive loop between social perception and action** by **communicating** with humans and **acting** in an environment over a long-term horizon involving many human and AI agents.

It remains a core technical challenge to model the high-dimensional action spaces as well as generation quality of multimodal outputs. I am currently working on 1) action abstractions in multimodal reinforcement learning which will enable agents to communicate and act over long-term horizons, and 2) a new paradigm to improve the sample-efficiency and consistency of cross-modal generation: generating semantically meaningful data in a target modality given a source modality which is useful for modeling multimodal interaction between humans and AI.

3. Closing the gap towards real-world deployment via 1) **robustness** to noisy and missing modalities, 2) **fair** representation learning from human-centric data, and 3) **interpretable** modeling of social commonsense.

I have worked towards robustness to noisy modalities via tensor representations [6] and missing modalities via cross-modal translation [10]. I have also developed methods to mitigate social biases in sentence representations [8]. Current, I am extending this work to mitigate biases in social AI systems that learn from multimodal experience.

Moving forward, I have identified **3 long-term goals** to benchmark, improve, and deploy real-world social AI.

<u>5.1 Benchmarking interactive intelligence</u>: Given the challenges in modeling and evaluating long-term interactive social intelligence, I am building new evaluation tasks leveraging human-in-the-loop and active learning to provide useful human labels in an interactive setting.

<u>5.2 Modeling interpretable social commonsense</u>: I am designing interpretable priors of social commonsense by combining research in psychology with data-driven approaches. This allows for flexibility in defining the structured knowledge present, fine-grained analysis of a model's decision-making process, and commonsense facts to model social intelligence.

<u>5.3 Formalizing the tradeoffs in multimodal learning</u>: Finally, existing approaches optimize for performance without quantifying potential drawbacks involving increased complexity, decreased robustness from imperfect modalities,

and unfair learning from biased modalities when deployed. I aim to quantify these tradeoffs to determine the overall contribution of a modality given its potential risks.

I am also particularly passionate about using social AI in the early detection and treatment of mental health disorders. Social AI holds promise as a support tool to assist clinicians in early detection of mental illness via 1) active recording of interactions between AI and humans and 2) passive monitoring via smartphones, both of which could be deployed at scale across at-risk populations. I hope to design AI that analyzes subtle verbal and nonverbal behaviors from humans before sending this information to clinicians for specialized diagnosis, while at the same time ensuring fair learning from human-centric data and privacy of personal and protected attributes [5, 9].

References

[1] Paul Pu Liang, Ziyin Liu, AmirAli Bagher Zadeh, and Louis-Philippe Morency. Multimodal language analysis with recurrent multistage fusion. In *EMNLP*, 2018.

[2] Paul Pu Liang, Ruslan Salakhutdinov, and Louis-Philippe Morency. Computational modeling of human multimodal language: The mosei dataset and interpretable dynamic fusion. *Carnegie Mellon University*, 2018.

[3] Paul Pu Liang, Amir Zadeh, and Louis-Philippe Morency. Multimodal local-global ranking fusion for emotion recognition. In *ICMI*, 2018.

[4] Paul Pu Liang, Yao Chong Lim, Yao-Hung Hubert Tsai, Ruslan Salakhutdinov, and Louis-Philippe Morency. Strong and simple baselines for multimodal utterance embeddings. In *NAACL-HLT*, 2019.

[5] Paul Pu Liang, Terrance Liu, Liu Ziyin, Nicholas B. Allen, Randy P. Auerbach, David Brent, Ruslan Salakhutdinov, and Louis-Philippe Morency. Towards personalized federated learning: Theoretical analysis and applications. *NeurIPS Workshop on Federated Learning*, 2019.

[6] Paul Pu Liang, Zhun Liu, Yao-Hung Hubert Tsai, Qibin Zhao, Ruslan Salakhutdinov, and Louis-Philippe Morency. Learning representations from imperfect time series data via tensor rank regularization. In *ACL*, 2019.

[7] Paul Pu Liang, Jeffrey Chen, Ruslan Salakhutdinov, Louis-Philippe Morency, and Satwik Kottur. On emergent communication in competitive multi-agent teams. In *AAMAS*, 2020.

[8] Paul Pu Liang, Irene Li, Emily Zheng, Yao Chong Lim, Ruslan Salakhutdinov, and Louis-Philippe Morency. Towards debiasing sentence representations. In *ACL*, 2020.

[9] Terrance Liu, Paul Pu Liang, Michal Muszynski, David Brent, Nicholas Allen Randy Auerbach, and Louis-Philippe Morency. Learning multimodal privacy-preserving markers of mood from mobile data. *under review*, 2020.

[10] Hai Pham*, Paul Pu Liang*, Thomas Manzini, Louis-Philippe Morency, and Barnabás Póczos. Found in translation: Learning robust joint representations by cyclic translations between modalities. In *AAAI*, 2019.

[11] Yao-Hung Hubert Tsai*, Paul Pu Liang*, Amir Zadeh, Louis-Philippe Morency, and Ruslan Salakhutdinov. Learning factorized multimodal representations. In *ICLR*, 2019.

[12] AmirAli Bagher Zadeh, Paul Pu Liang, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. Multimodal language analysis in the wild: Cmu-mosei dataset and interpretable dynamic fusion graph. In *ACL*, 2018.