Asocial Design Yields Antisocial Agents

Agentic technologies are increasingly emerging in intrinsically social settings; yet, the social capacities of such systems remain critically deficient. Lacking the ability to navigate antisocial dynamics in particular, artificial agents have profound potential to cause harm. For example, Microsoft’s chatbot Tay - having spent one day on Twitter - morphed from the cheery teenage girl persona with which it was deployed into an unabashed proponent of White and male supremacy. While Tay’s radicalization stemmed from the toxic disinhibition of the bot’s interlocutors, it was Tay’s inability to recognize aggression that enabled the transformation.

In this talk, I will discuss my recent empirical work to highlight several ways in which design that is agnostic or inattentive to human social dynamics can foster antisociality in human-agent interactions. In particular, I will discuss the consequences of hyper-anthropomorphization - from aversion reflected by the uncanny valley toanimosity present in people’s objectification of humanlike robots - and broader considerations, both negative (e.g., capacity of agentic technologies to exacerbate social marginalization) and positive (e.g., capacity to promote prosocial dynamics), toward improving the sociality of future systems.

Dr. Megan Strait is an Assistant Professor in the Department of Computer Science at the University of Texas Rio Grande Valley where she leads the Social Systems (SOSY) Lab. Drawing upon an interdisciplinary background in cognitive science, computer science, and psychology, her research addresses the less positivistic side of human-agent interaction dynamics to bear out both aesthetic and architectural design guidelines for socially oriented technologies. Her research has won multiple Best Paper and Best Paper Honorable Mention awards at the ACM/IEEE HRI and IEEE RO-MAN conferences, as well as a dissertation award from Tufts University where she earned her Ph.D. and M.Sc. (in Cognitive Science and Computer Science, respectively) developing both computational and experimental models of emotion in human-robot interaction.

www.meganstrait.com