Abstract

The Foundations of Robosociology: Values and The Aggregate Behaviors of Synthetic Intelligences Adam Alonzi, Interdisciplinary, Analyst at EthicsNet

Adam Alonzi

Abstract:

Outcomes on the macro level often cannot be accurately extrapolated from the microbehaviors of individual agents. The interdependence of complex system's components makes simulation a viable option for exploring cause and effect relationships within it (Miller and Page, 2009). Chaos theory emphasizes the sensitivity of such networks to starting conditions (Boccaletti, 2000), which strongly suggests thought should be put into the architecture of an AGI "society" before it begins to take shape. Protocols for emergency interventions should certainly be in place, but the network itself should be robust enough from the beginning to handle sudden deviations from basic ethical precepts by one or more of its members. Outside of its context, and without any information about the parts to which it is connected, a cell or leaf or animal can be studied, but not understood in a meaningful way (Mitchell, 2009). Creating moral agents in a hyperconnected world will involve modeling their interactions with entities like and unlike themselves in the face of both predictable and unforeseen events. This will be helpful as groups can behave differently than their individual parts (Schelling, 1969) Keeping AI friendly does not end with giving each AI a set of maxims before letting them loose, but satisfactorily explicating upon the emergent phenomena that arise from the interactions of similarly or differently "educated" machines. Because of the near certainty that synthetic intelligences will communicate rapidly and regularly, it is imperative that thought leaders in AI safety begin thinking about how groups of artificially intelligent agents will behave.

Biography:

I, Abhishek Gupta, PhD, have completed PhD at the age of 32 years from King George's Medical University, Lucknow, India and postdoctoral studies from Canada and Sweden Universities. I am presently the Principal Scientist, Preclinical Biology, Drug Discovery and Development Division at Patanjali Research Institute, Patanjali, a FMCG company of India. I have published more than 22+ articles and 45+ short papers in well reputed journals and has been serving as scientific review board, board member and professional affiliations of repute.

Publication of speakers:

1. Boccaletti, Stefanos, et al. "The control of chaos: theory and applications." Physics reports 329.3 (2000): 103-197.

2. Miller, John H., and Scott E. Page. Complex adaptive systems: an introduction to computational models of social life: an introduction to computational models of social life. Princeton university press, 2009.

3. Page, Scott E. Diversity and complexity. Princeton University Press, 2010.

4. Mitchell, Melanie. Complexity: A guided tour. Oxford University Press, 2009.

5. Schelling, Thomas C. "Models of segregation." The American Economic Review 59.2 (1969): 488-493.