Can machines think? Can machines experience psychological suffering? If yes, how?

Presentation/abstract

In his classic 1950 paper, Alan Turing posed the question “Can machines think?” - a question that has haunted the field of Artificial Intelligence for close to 70 years. In the paper, Turing argued that the question was not helpful as “thinking” could not be defined clearly enough. For that reason, he proposed the “Imitation game” (commonly referred to as the “Turing test”) as a mean for establishing criteria for success in the field of AI.

Contextual behavioral science is a broad research field emerging from behavioral psychology. Relational Frame Theory (RFT) is a contextual behavioral account of language and cognition. The fundamental thesis of RFT is that language and cognition are all instances of arbitrarily applicable relational responding (AARR). Hence, the question above from an RFT standpoint becomes “Can machines do AARR?”. Furthermore, as RFT has modeled various aspects of different clinical phenomena, a natural next question from the point of clinical psychology is: “Can machines experience psychological suffering?”

The subfield of AI research that aims to build machines that can think is commonly referred to as Artificial General Intelligence (AGI). That is, building general-purpose systems that could perform at human-level on various tasks across different domains. One approach to AGI is Pei Wang’s non-axiomatic logic, that aims to formalize principles for general intelligence within a logical framework and its software implementation NARS (Non-Axiomatic Reasoning System).

In the talk, we will propose a step-wise process toward exploring whether machines can think, investigated through the question “Can NARS do AARR?”. Furthermore, we will explore possible roads towards answering the question “Can NARS experience psychological suffering?”