"Cogito, Ergo Machina?": Topics in Machine Thought

MIT Splash! 2015

Instructor: Arnav Sood, New York University, asood@nyu.edu. A fuller biography can be found at https://esp.mit.edu/learn/teachers/renoir/bio.html.

Seminar Description: This seminar is an exploration of fundamental problems in machine thought, using the practical lens of IBM's "Watson!," and some classical thought experiments.

Readings: There are two types of reading for this class; some articles on each philosophical topic, and some IBM papers on Watson. Both are entirely optional, but I'll give you all chocolate if you skim the Watson ones[1]. All the readings can be found online at http://princeton.edu/~asood/Splash_2014.

Seminar Structure: First, we will clarify our intuitions, drawing from the following topics:

- Topic A: The Turing Test. What is it, how does it work, and does passing it even signify anything? We will play with some chatbots, like ELIZA and Jabberwacky, and discuss results from the first implementation of the Turing Test (the Loebner Prize).
- Topic B: Functionalism and Behaviorism. What are these? If something was functionally isomorphic to the brain, what would we know about it? For example, need it exhibit intelligent behavior, think, or be conscious?

  Guided by these questions, we'll talk about the concept of qualia, and meet some beings that have none (zombies and Blockheads). We'll also examine some more peculiar implementations of brains, such as AND gates built of cats and mice, the China Brain, Hofstadter's Einsteinian Book, and maybe artificial neural networks.
- Topic C: Intentionality and Intensionality. What are they, and to what extent are they significant to intelligence? We will discuss Davidson's Swampman, derived versus intrinsic intentionality, and the idea of an intensional context.
- Topic D: Symbols, Syntax, and Semantics. What are these, and what does it mean to intelligently process language? We will discuss the conception of the brain as a syntactic engine driving a semantic engine, syntactic versus semantic computation, Searle's Chinese Room, and the language of thought hypothesis as it pertains to Watson's internal processing.

Then, we will use these ideas to assess Watson's intelligence. In particular, we will focus on the following features of Watson's language engine:

[1] To learn more about whether or not you should do the readings, visit http://en.wikipedia.org/wiki/Free_rider_problem

- **Syntactic-Semantic Graph**: Watson places identified objects at nodes, and models constructed relationships as the edges. For example, Watson might use the relation `AuthorOf(Author, Composition)` to connect the node `Harper Lee` to `To Kill a Mockingbird,` based upon their association by viable verbs, such as `write` or `produce.`
- **Categorical Agglomeration**: Once Watson identifies the syntax for a thing (for example, `Harper Lee`), it creates an abstraction ("Harper Lee,") which is a matrix of that thing's properties. For example, Watson might add things like `AuthorOf: To Kill A Mockingbird` or `Born: Monroeville, AZ` to flesh out Harper Lee's matrix of properties.

Acknowledgements: This class was inspired by Professor Ned Block's[2] course, Minds and Machines.

[2] `http://www.nyu.edu/gsas/dept/philo/faculty/block`