Artificial Intelligence

Chris Gregg, based on slides by Eric Roberts
CS 208E
November 15, 2018
The Imitation Game

- How to know a human from a machine?
- Is the computation model of the brain more than a complex Turing machine?
- Tests and Modern *Imaginaries*.
“Communities are to be distinguished, not by their falsity or genuineness, but in the style in which they are imagined.”

Human or Machine Intelligence?

- Fixed task in a narrow field (e.g.: chess, arithmetic.)
- Unforeseen paths to get to a victory
- Ability to look at every single possibility
- Part way answers and ambiguity
- Subtlety
- Consciousness of such subtlety
Artificial Intelligence and the Turing Test

• In 1950, Alan Turing posed a thought experiment to address a question that underlies the field of Artificial Intelligence.

Computing Machinery and Intelligence
A. M. Turing

1 The Imitation Game
I propose to consider the question, ‘Can machines think?’ . . .

[A revised form of the question] can be described in terms of a game which we call the ‘imitation game’. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. . . .

We now ask the question, ‘What will happen when a machine takes the part of A in this game?’
A Dialogue in the Imitation Game

In his paper, Turing suggests the following as a possible dialogue between the interrogator and the unknown subject:

Please write me a sonnet on the subject of the Forth Bridge.
Count me out on this one. I never could write poetry.
Add 34957 to 70764.
105621.
Do you play chess?
Yes.
I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?
R-R8 mate.

Turing notes that a computer trying to fool the interrogator might wait several seconds before answering the arithmetic problem, but does not point out that the answer is incorrect.
A Q(q)uestion of Ambiguity

In the English notation, the squares are designated by an abbreviation of the piece in the first row. Thus, “I have a K at my K1” is Ke1 in algebraic notation (assuming that the interrogator plays white.)

- I have K at K1.
- You have only K at K6 and R at R1.
Ambiguity in Turing’s Dialogue

The Chess Example in Turing’s Mind Paper is Really about Ambiguity

Figure 3: Four possible interpretations of Turing’s chess problem
More on Chess and Game AI

See https://www.chess.com/article/view/computers-and-chess---a-history


and http://users.ox.ac.uk/~uzdm0006/scans/1kchess/
Eliza

In the mid-1960s, MIT Professor Joe Weizenbaum wrote a now famous program called *Eliza*, which would mimic—at least in minor formulaic ways—the style of a Rogerian psychotherapist.

Note: The Eliza program is built into GNU Emacs; to run it, type ESC-X Doctor RET.
Eliza

> what do you think?
Oh... think?
> yes
I see.
> what do you see?
You're not really talking about me, are you?
> yes
I understand.
> nonsense
OK... "what do me see". Tell me more.

Source: http://psych.fullerton.edu/mbirnbaum/psych101/Eliza.htm
This is the source code for the version of Doctor distributed with Gnu Emacs 20.2.2.

;;; doctor.el --- psychological help for frustrated users.
;; Maintainer: FSF
;; Keywords: games
;; This file is part of GNU Emacs.

...
(make-local-variable 'want)
(setq want '((want) (desire) (wish) (want) (hope)))
(make-local-variable 'shortlst)
(setq shortlst
  '((can you elaborate on that ?)
    (($ please) continue .)
    (go on\, don\'t be afraid .)
    (i need a little more detail please .)
    (you\'re being a bit brief\, ($ please) go into detail .)
    (can you be more explicit ?)
    (and ?)
    (($ please) go into more detail ?)
    (you aren\'t being very talkative today!)
    (is that all there is to it ?)
    (why must you respond so briefly ?)))
make-local-variable 'replist
(setq replist
  '(((i . (you))
    (my . (your))
    (me . (you))
    (you . (me))
    (your . (my))
    (mine . (yours))
    (yours . (mine))
    (our . (your))
    (ours . (yours))
    (we . (you))
    (dunno . (do not know))
  ;;
    (yes . ()))
    (no\, . ()))
    (yes\, . ()))
    (ya . (i))
    (aint . (am not))
    (wanna . (want to))
    (gimme . (give me))
    (gotta . (have to))
    (gonna . (going to))
    (never . (not ever)))
...
Parry

- PARRY was tested in the early 1970s using a variation of the Turing Test. A group of experienced psychiatrists analyzed a combination of real patients and computers running PARRY through teleprinter.

- Another group of 33 psychiatrists were shown transcripts of the conversations.

- The two groups were then asked to identify which of the “patients” were human and which were computer programs. The psychiatrists were able to make the correct identification only 48 percent of the time—a figure consistent with random guessing.

Watson

- **Watson** is a question answering computer system capable of answering questions posed in natural language, developed in IBM’s DeepQA project by a research team led by principal investigator David Ferrucci. Watson was named after IBM’s first CEO and industrialist Thomas J. Watson.

- The computer system was specifically developed to answer questions on the quiz show *Jeopardy!*. In 2011, Watson competed on *Jeopardy!* against former winners Brad Rutter and Ken Jennings. Watson received the first place prize of $1 million.

  https://www.youtube.com/watch?v=WFR3lOm_xhE
Strong AI vs. Weak AI

**Strong AI (Searle’s approach)**
“The computer is not merely a tool in the study of the mind, rather the appropriately programmed computer really *is* a mind in the sense that computers given the right programs can be literally said to understand and have other cognitive states.”

**Weak AI**
“Computers just *simulate* thought, their seeming understanding isn’t real understanding (just as-if), their seeming calculation is only as-if calculation, etc. Nevertheless, computer simulation is useful for *studying* the mind (as for studying the weather and other things).”

In 1980, the philosopher John Searle posed an interesting thought experiment. Suppose that a person who speaks no Chinese is placed in a room with a huge rule book that contains Chinese answers to any conceivable question. This person’s job is to look at questions on an input screen, match the pattern of characters in the rule book, and then copy down the answer. Searle’s question is whether the room understands Chinese.
To Be or Not To Be (Conscious)
That is the Question
Functionalism

If we could replicate all the functions which the human brain performs in a machine, we could (essentially) replicate the brain and we would have whatever experiences the brain might have.

— Daniel Dennett

• Daniel C. Dennett is University Professor and Austin B. Fletcher Professor of Philosophy, and Co-Director of the Center for Cognitive Studies at Tufts University.

Behavioral Level

• Behavior-based robots (BBR) usually show more biological-appearing actions than their computing-intensive counterparts, which are very deliberate in their actions. A BBR often makes mistakes, repeats actions, and appears confused, but can also show the anthropomorphomorphic quality of tenacity. Comparisons between BBRs and insects are frequent because of these actions.

• BBRs are sometimes considered examples of weak artificial intelligence, although some have claimed they are models of all intelligence.

• The term was introduced by Rodney Brooks and colleagues at MIT in 1986. Brooks is widely influential in autonomous robotics and elsewhere in real-time AI.
Tests and More Tests

• The Loebner Prize in Artificial Intelligence
• MIT’s algorithm passes Turing tests
The Loebner Prize

- Hugh Loebner in 1990, in conjunction with the Cambridge Center for Behavioral Studies, underwrites a contest to implement the Turing Test, which was inaugurated at the Computer Museum (Boston, USA) in 1991.

- $100K and a gold medal for first computer that gave the same responses.

- Looking for “the most human-like computer.”

- Each entry provided with a set of 20 questions in English with at least two Winograd style questions.

- Winner is the best entry relative to other entries that year, not to the absolute pool.

- The 2016 contest was held at Bletchley Park on 17th September. First prize went to Mitsuku created by Steve Worswick.
Superintelligence

• In 2014, philosopher Nick Bostrom wrote a book titled, *Superintelligence*. In it, he argues that if (when) machines surpass humans in intelligence, humans may just replace humans as the dominant “life form” on earth.

• Bostrom cautions that AI researchers and designers need to be extremely careful to keep a superintelligent computer under tight control, lest it “escape” and wreak havoc.

• If a superintelligent computer has a notion of freedom (and why wouldn’t it?), it might do everything it can to break free of its creators.
The Paperclip Maximizer

• Another concern Bostrom brings up in *Superintelligence* is demonstrated with a thought experiment about a *paperclip maximizer*.

• Suppose that a machine with human-level intelligence has been programmed with a single goal: to maximize the number of paperclips in its collection.

• This seems like a benign goal, and it is hard to think of how a machine programmed as such could be malicious.

• However, consider this: the machine might collect paperclips, it might earn money to buy paperclips, or it might start a manufacturing plant to produce them.

• Additionally, though, it might also work to improve its own intelligence, possibly in an *intelligence explosion*.

https://wiki.lesswrong.com/wiki/Paperclip_maximizer
The Paperclip Maximizer

• The machine could eventually innovate to produce more paperclips (remember, this is its ultimate goal!). It will continue to improve its intelligence, too, so it can better meet its goal.

• Eventually, in support of the goal, it may decide that it needed to use all of Earth’s resources to produce more paperclips, in the process destroying biological life on earth to do so (something, something, *Soylent Green*).

• If it remained unchecked, the machine would continue to produce paperclips until the entire universe is made up of paperclips, until the heat death of the universe itself.

• So – we must create machines that have human values, in order to combat superintelligent machines.

https://wiki.lesswrong.com/wiki/Paperclip_maximizer
MIT’s AI Passes Key Turing Tests

• At CSAIL (Computer Science and Artificial Intelligence Lab) a team has created a deep-learning algorithm which passes the Turing Test for sound.

• Able to predict the sounds of a video by looking at the sound properties of each frame and stitching bits of audio which match similar sounds within a database.

• AI was able to simulate “thud,” staccato taps, rustles.

http://www.zdnet.com/article/mits-artificial-intelligence-passes-key-turing-test/
MIT Program Writes Characters

- The MIT system also passes a “visual” Turing Test by writing characters.

In Our Minds
Lingering Questions

http://www.bfi.org.uk/are-you-a-replicant
Contrary Views on the Question

- Alan Turing’s paper offers the following possible objections:
  1. The theological objection
  2. The “head in the sands” objection
  3. The mathematical objection
  4. The argument from consciousness
  5. Arguments from various disabilities
  6. Lady Lovelace’s objection
  7. Arguments from continuity in the nervous system
  8. The argument from informality of behaviour
  9. The argument from extra-sensory perception
Thomas Kuhn’s Perception Experiment

In a moment, I’m going to turn over these cards. Look at them and write down what you see.

What were the five cards?
Most people don’t notice that the diamond is black instead of red.
Perceptions Can Be Influenced
Perceptions Can Be Influenced
Subjective Contours
Subjective Transparency
Adding Color
Mental vs. Physical Transparency
The Dangers of AI

Why the Future Doesn’t Need Us
By Bill Joy
Playing God and “The Singularity”
The End