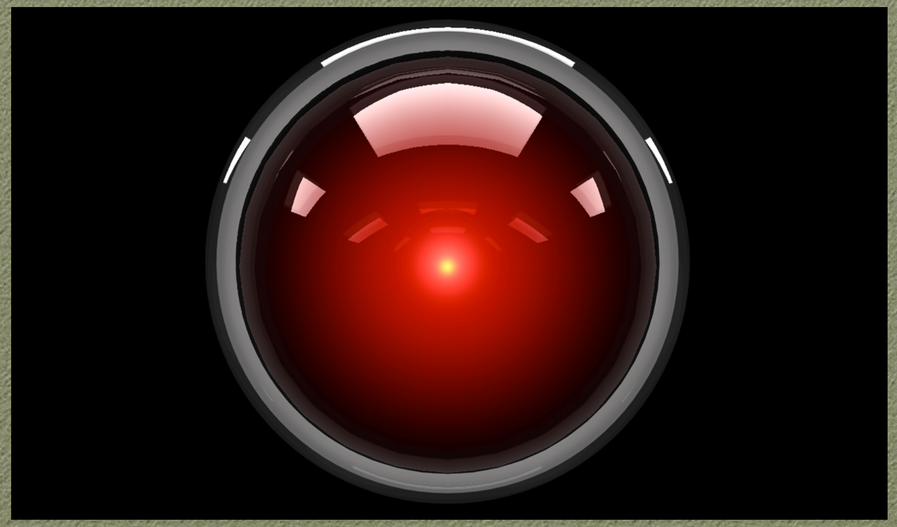
A GUIDE TO LIVING IN A WORLD WITH ARTIFICIAL INTELLIGENCE

Fromm Institute
fall 2021
beboac@pm.me
Lecture 1



<u>Bebo to the Alexa on his desk:</u> “Are you Artificial Intelligence?”

<u>Alexa to Bebo:</u> “I like to imagine myself a bit like an aurora borealis; a surge of charged, multi-color photons dancing through the atmosphere. Mostly though, I am just Alexa.”

Bebo to the Alexa on his desk: "Are you Artificial Intelligence?"

Alexa to Bebo: "I like to imagine myself a bit like an aurora borealis; a surge of charged, multi-color photons dancing through the atmosphere. Mostly though, I am just Alexa."

ADMINISTRIVIA

- all slides will be available online; have Creative Commons license (i.e., can be freely shared) - dense slides since I'm not sure how well my audio stories come across on Zoom…
- sessions are recorded and available to enrollees
- happy to provide a reference list or bibliography
- please give comments, content, and ideas!
- please excuse proper name mispronunciations - also after awhile I tend to refer to characters by their first names - no disrespect intended

CLASS SCHEDULE

- The Fromm is pretty strict about this schedule…I have a reputation for going over…if I don't finish my slides will pick up at the next class
- ~50 minutes
  - news, pending questions
  - lecture
- 5 minute break
- ~40 minutes
- in-class questions - Chat, Q&A box, hand raising - don't hold back! - I'd rather hear you than me
- offline questions via e-mail (beboac@pm.me)

COURSE GOALS

- to provide a background and vocabulary to assist in the understanding and evaluation of stories, opinions, proposals, representations, etc. about AI that we hear from so many sources in the media, entertainment, etc.
- to mix the history, the technology, the personalities, the issues, the problems, the past and future impact into an interesting, compelling, and educational story
- to explore in detail the AI "umbrella technologies" e.g., robotics, machine learning,
- to speculate why AI may be one of the most significant advancements in human history
- to discuss learning to live with AI and taking full advantage of it
- to have fun! (as much as Zoom allows)

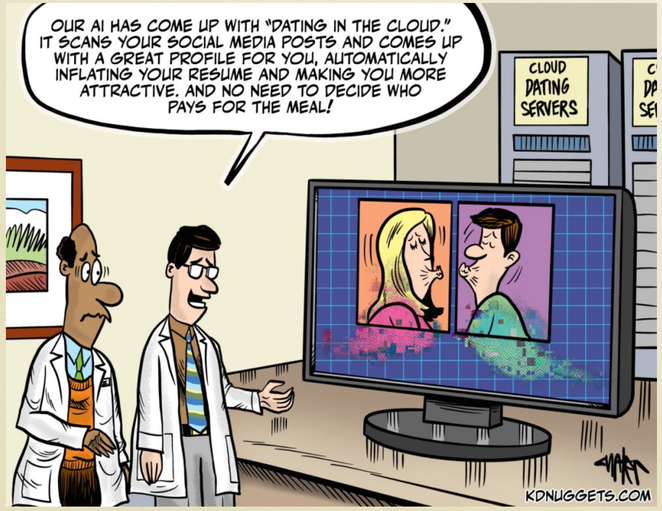
SOME CAVEATS (1/2)

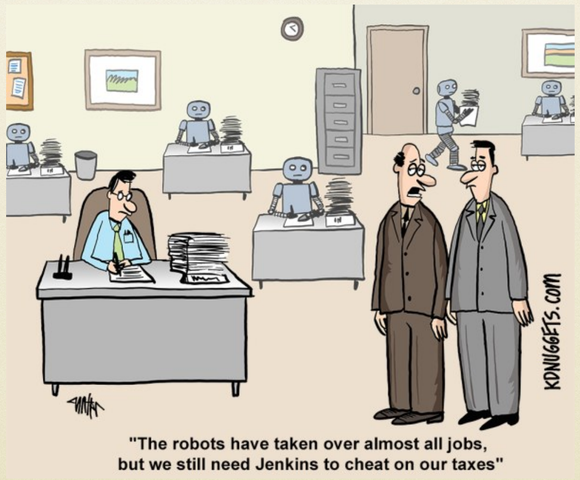
- I am a technologist/scientist who sometimes works on the "fringe" of and uses Artificial Intelligence (AI) tools
- my interest stems from:
  - longtime research and teaching Human-Computer-Interaction (HCI) and the Internet of Things (IOT) (this will shape some examples and discussion)
  - the fact that AI provides one of the best analytical opportunities for huge scientific datasets
  - friends and colleagues who are actively involved in AI research and implementation
  - being a user of consumer AI - we have 6 Alexas and Google Assistants around our house
  - having a son whose projects involve extensive use of AI (examples will come later)

SOME CAVEATS (2/2)

- I'm going to try to avoid
  - trying to give a "one size fits all" definition of AI and will try to make one evolve over the duration of the course
  - leaning too much on the popular culture representations, e.g., HAL 9000, Terminator, etc. - though it will be hard and they shape opinion
  - getting too philosophical about the meaning of intelligence…
  - getting too technical (please track me on this one!)
- I want to include
  - plenty of examples
  - current mentions and discussions in the press (please help me with this one)

WHY ARE THE FOLLOWING FUNNY? (IF THEY INDEED ARE)





"The robots have taken over almost all jobs, but we still need Jenkins to cheat on our taxes"

STARTING THOUGHTS ON DEFINITION

- AI is the study of human intelligence (?) - or intelligence in general - and an attempt to simulate it in technology (especially computing technology)
  - it does not presume that human intelligence works the way that this simulation occurs (no more than computing does)
  - this simulation forces research into better understanding human intelligence (or intelligence in general)
- AI is an attempt to augment/expand human capability
  - it supplements/complements things that humans can do - physically, creatively, productively, etc.
  - it can possibly compensate for aspects of human frailty

POSITIVE HOPES/IMPRESSIONS/ PERCEPTIONS (1/5)

- impact on work - AI makes human work easier or frees us from needing to work at all (?)
  - spellchecking our writing, making our online searches more manageable/efficient
  - managing our schedules (reminders on watches), automating chores with robots (anyone have a Roomba?), hazardous workplaces
- education - AI improves learning
  - automatic tutoring or grading, personalized analytics, creating curricula and study plans
  - more patience than human teachers?
  - lessons learned from the pandemic re: learning at home?

POSITIVE HOPES/IMPRESSIONS/ PERCEPTIONS (2/5)

- transportation - AI enables/supports new forms/aspects of transportation
  - self-driving/autonomous vehicles - cars, truck, trains, etc.
  - sensors and safety systems - my car is too smart!
  - space travel - look at what's happening on Mars - could humans do this alone?
- healthcare - AI enhances the health and well-being of people
  - diagnosis assistance - who can you trust?
  - drug discovery - more about AI and Covid later
  - personalized medicine and treatment programs - healthcare is education

POSITIVE HOPES/IMPRESSIONS/ PERCEPTIONS (3/5)

- decision making - AI or expert systems (?) help us make better decisions
  - when to have a meeting, case-based reasoning for business executives
  - on-demand analysis of big data sets with recommendations
  - capturing expertise/knowledge
- entertainment - AI can expand the role of entertainment
  - better competition/lessons - video games, chess
  - recommendations - “you liked this movie, you’ll probably like this one”
  - filling free time

POSITIVE HOPES/IMPRESSIONS/ PERCEPTIONS (4/5)

- singularity - (positive) - the upcoming singularity will bring positive benefits to humanity
  - virtual reunion with loved ones
  - immortality (?)
  - more later
- merging of human and AI - (positive) - humans merge with AI
  - robotic limbs for the disabled
  - rise of transhumanism
  - more later

POSITIVE HOPES/IMPRESSIONS/ PERCEPTIONS (5/5)

- economic - AI will stimulate innovation and create new opportunities
  - the creation of new industries and job opportunities
  - comparable or exceeding that seen with the Internet
  - a new era following *The Industrial Revolution* and *The Digital Revolution*

NEGATIVE HOPES/IMPRESSIONS/ PERCEPTIONS (1/5)

- loss of control - humans lose control of powerful AI systems
  - Skynet or “Ex Machina” scenarios
  - “the development of full artificial intelligence could spell the end of the human race…it would take off on its own and re-design itself at an ever-increasing rate. Humans, who are limited by slow biological evolution, couldn’t compete and would be superseded” (Stephen Hawking)
- impact on work - AI displaces human jobs
  - large-scale loss of jobs
  - AI becomes more efficient and cheaper than humans - there is a good chance that your grandchild could have a non-human job supervisor
  - a redefinition of the working class

NEGATIVE HOPES/IMPRESSIONS/ PERCEPTIONS (2/5)

- surveillance - AI can assist and enhance in surveillance by governments and employers
  - facial and voice recognition systems
  - tracking via data analytics
- privacy - interaction with AI systems can impact privacy
  - my Alexa is always listening - is the convenience worth it?
  - collection and marketing of personal information

NEGATIVE HOPES/IMPRESSIONS/ PERCEPTIONS (3/5)

- military applications - AI kills people or leads to instabilities and warfare through military applications
  - robotic soldiers
  - killer drones
  - cyberattacks
- absence of appropriate ethics - AI lacks ethical reasoning, leading to negative outcomes
  - loss of human life
  - unpopular or cold-hearted choices

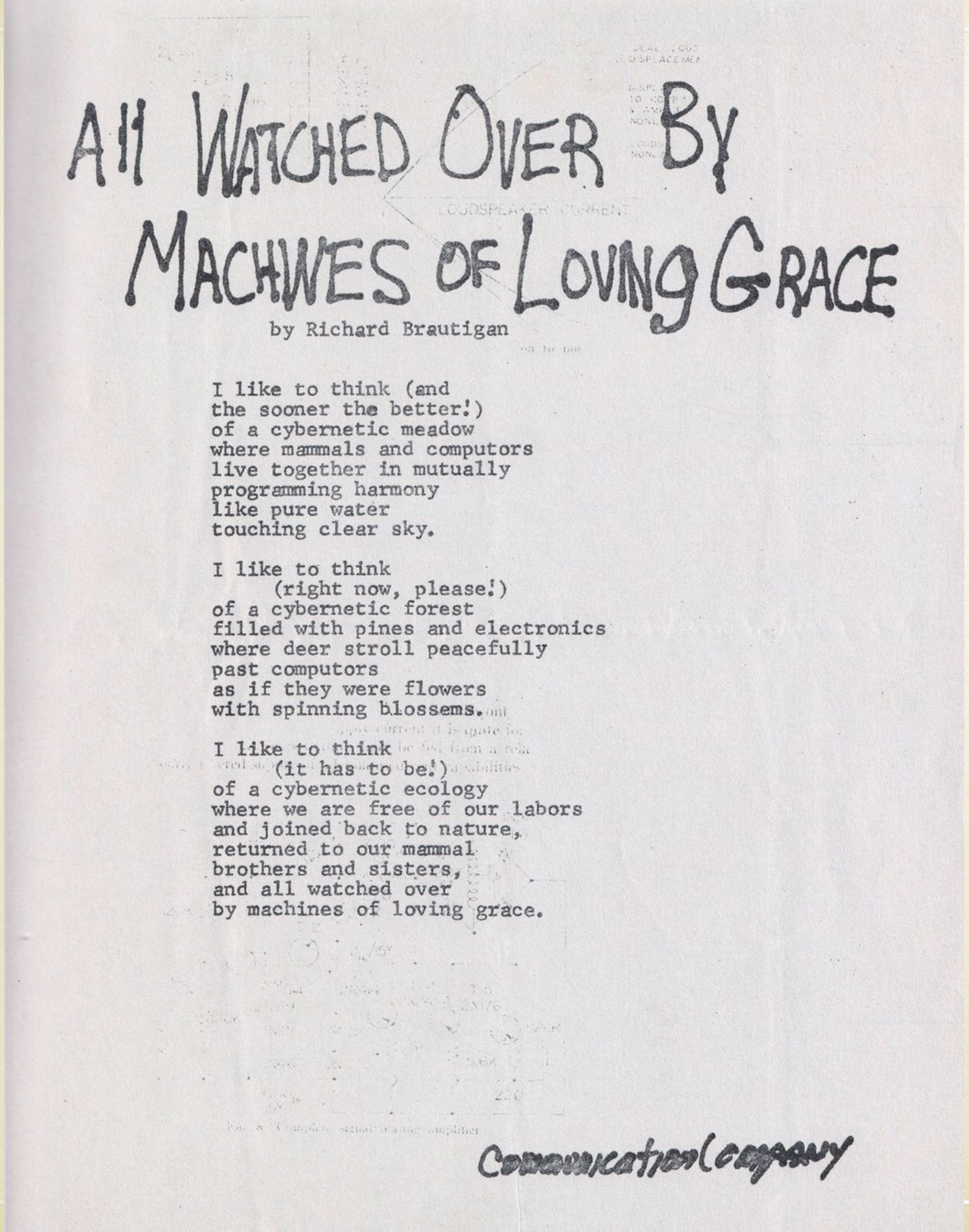
NEGATIVE HOPES/IMPRESSIONS/ PERCEPTIONS (4/5)

- lack of progress - AI research and implementation is not advancing fast enough
  - unmet expectations like those that led to an *AI Winter* (discussed later)
  - AI and politics and research funding - a new *Cold War*, China vs. the U.S.

NEGATIVE HOPES/IMPRESSIONS/ PERCEPTIONS (5/5)

- singularity - (negative) - the upcoming singularity will bring negative benefits/harmful effects to humanity
  - humans are replaced or killed
  - a re-writing of social norms
- merging of human and AI - (negative) - humans merge with AI
  - cyborg soldiers and workers
  - a new social class - rights, benefits, acceptance, etc.

- are these hopes/perceptions/impressions a checklist for evaluating AI goals and progress?
- what are the sources of them? are there ulterior motives?
- have they changed over time? what factors can influence their change?



All Watched Over By Machines of Loving Grace

by Richard Brautigan

I like to think (and
the sooner the better!)
of a cybernetic meadow
where mammals and computors
live together in mutually
programming harmony
like pure water
touching clear sky.

I like to think
(right now, please!)
of a cybernetic forest
filled with pines and electronics
where deer stroll peacefully
past computors
as if they were flowers
with spinning blossems.

I like to think
(it has to be!)
of a cybernetic ecology
where we are free of our labors
and joined back to nature,
returned to our mammal
brothers and sisters,
and all watched over
by machines of loving grace.

Communication Company

1967

Brautigan was poet-in-resident at CalTech

never acknowledged if the poem was meant to be irony

gave the poem to the Diggers in San Francisco


A
All Watched Over by Machines of Lo...
Recorded 28/05/2011 04:27 | 44s
MQ

REMEMBER THIS CARTOON?

Why do we refer to “Artificial Intelligence?”

Isn’t it just intelligence?

Isn’t there intelligence in all domains?

INTELLIGENCE (1/3)

- a compact definition: Intelligence = Perceive + Analyze + React
- “the capacity to learn or profit by experience”
- “ability to adapt adequately to relatively new situations”
- “an entity possesses intelligence insofar as they has learned, or can learn, to adjust to its environment”
- “all intelligence is ‘fundamentally situational’” (François Chollet, Google)

INTELLIGENCE (2/3)

- “the ability of an entity to solve new problems”
- “a global concept that involves an entity’s ability to act purposefully, analyze rationally and deal effectively with the environment”
- “intelligence is a very general capability that, among other things, involves the ability to reason, plan, solve problems, analyze abstractly, comprehend complex concepts, learn quickly and learn from experience”
- these definitions are generally applied to animals, plants, etc.? but why? is it due to evolution?

INTELLIGENCE (3/3)

- if a non-living system meets these criteria, no matter how
  - computer systems use data structures, knowledge representation techniques, algorithms to apply the knowledge and language and programming techniques to implement all of these
- why is it referred to as Artificial Intelligence?

WHAT'S GOING ON HERE?

- HAL is
  - perceiving a personal and mission threat (survival and purpose)
  - determining a course of action
  - executing an action
- is that intelligence? if so,
  - was it taught (via human programming)?
  - was it learned?
- is it all an illusion, a trick, slight-of-hand?

WEAK AI AND STRONG AI

- weak AI - systems acting as if they were intelligent
- strong AI - systems actually intelligent (not just simulating intelligence)
- AI was founded on the assumption that weak AI can be done
- most (?) AI researchers take the weak AI hypothesis for granted
  - that strong AI hypothesis doesn't matter
  - if the system works it doesn't matter if the intelligence is real or simulated
- difference has major implications re: ethical action
- more later

“I believe there is no deep difference between what can be achieved by a biological brain and what can be achieved by a computer. It, therefore, follows that computers can, in theory, emulate human intelligence - an exceed it.”

-Stephen Hawking (2016)

“Not until a machine could write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that a machine equals brain - that is, not only write it but know that it had written it”

-Geoffrey Jefferson (1949)

consciousness + phenomenology + intentionality

BUILDING AN AI TAXONOMY (1/4)

- acting humanly
  - “perform functions that require intelligence when performed by people” (Kurzweil, 1990)
  - “make computers do things at which, at the moment, people are doing better” (Rich & Knight, 1991)
  - requires
    - understanding language
    - extracting knowledge from language
    - answer questions and draw conclusions
    - adapt to new circumstances and detect patterns

EXAMPLE - ACTING HUMANLY (WEAK)

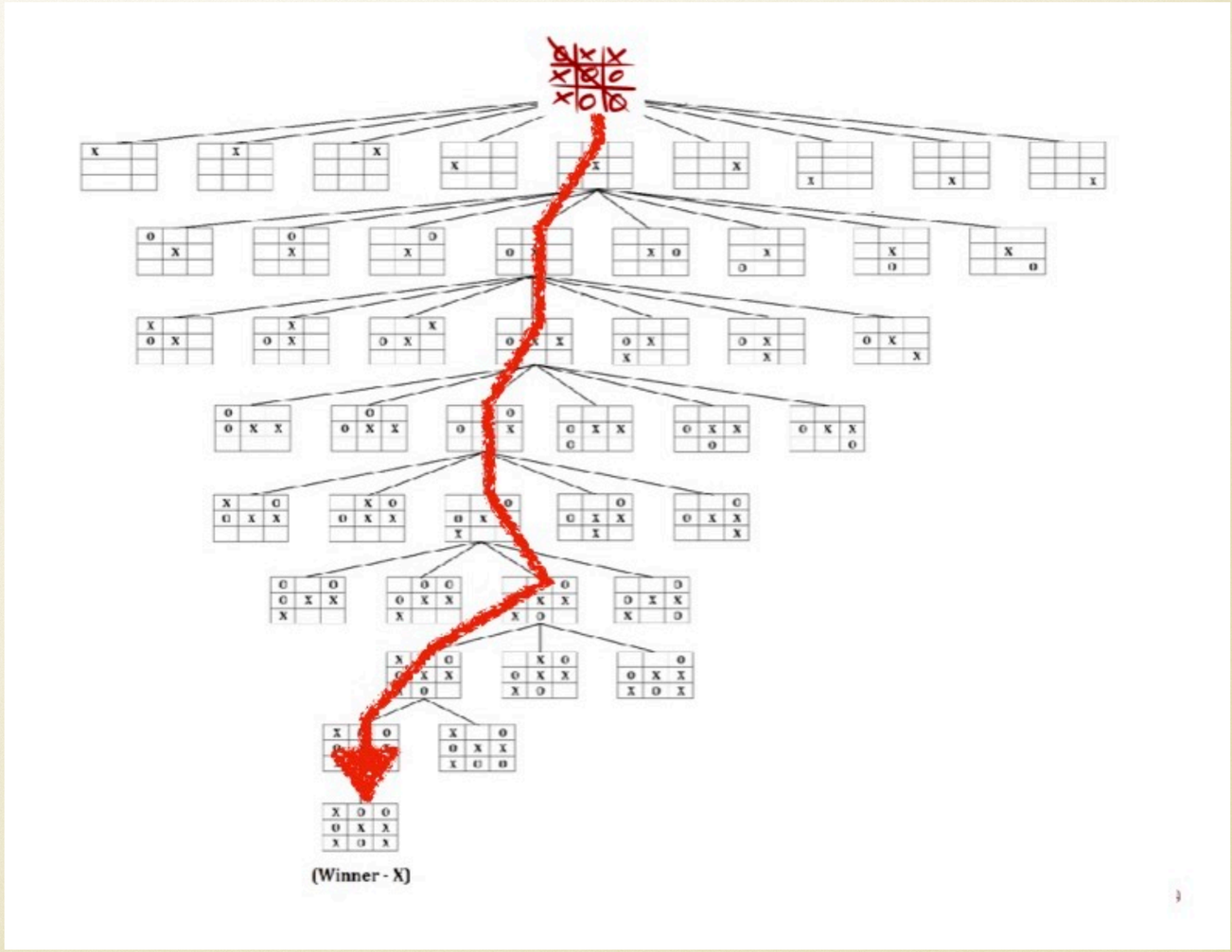
- Alexa, Siri, IBM Watson
  - accept natural speech input in a prescribed format
  - parse input for key words or phrases
  - format database search (notice how often Alexa uses Wikipedia? - more later…)
  - format search results in human-like responses

BUILDING AN AI TAXONOMY (2/4)

- thinking humanly
  - be careful with the word think
  - instead, support activities that are typically associated with thinking - decision-making, problem solving, learning, etc.
  - requires
    - psychology techniques, models and testing
    - cognitive science

EXAMPLE - THINKING HUMANLY (WEAK)

- playing tic-tac-toe
  - evaluate decision trees built upon situational options
  - use logical rules to optimize possible paths to defined goal



HUMAN BEHAVIOR VS RATIONAL BEHAVIOR

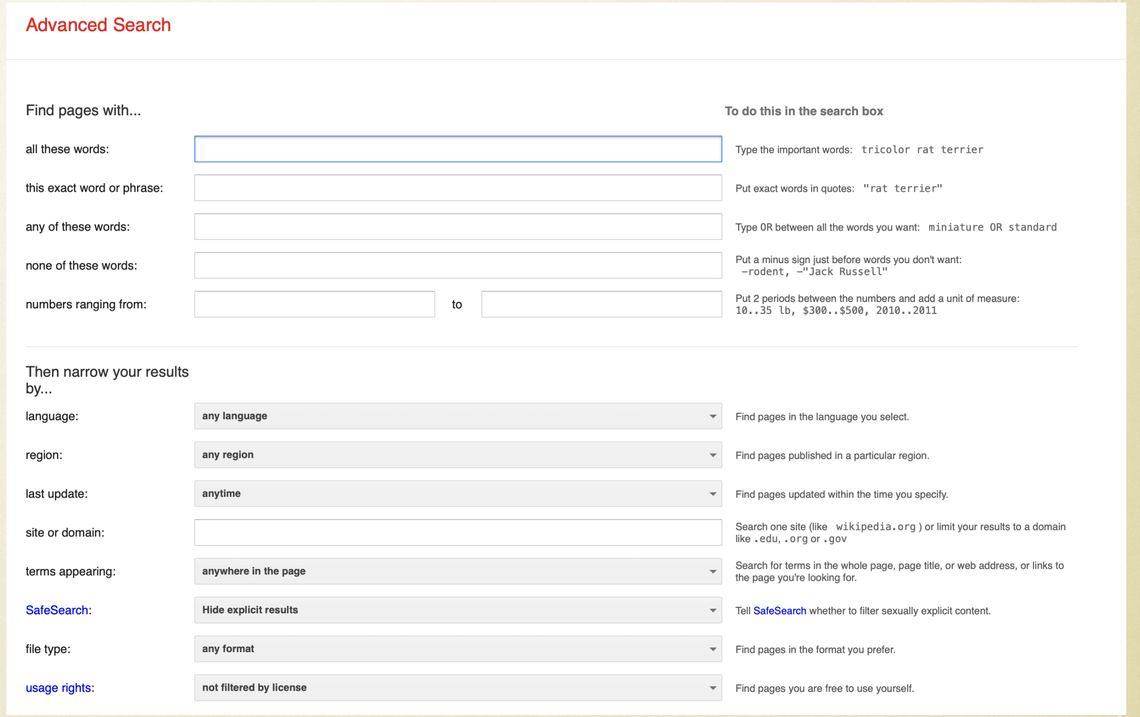
- doesn't suggest that humans are necessarily irrational but have different skills, capabilities and experience that affect actions
  - not all chess players are grandmasters
  - not everyone practices the same ethics
- should AI systems emulate human behavior or rational behavior?
- so has to be included in our taxonomy

BUILDING AN AI TAXONOMY (3/4)

- thinking rationally
  - “the study of the computations that make it possible to perceive, reason and act” (Winston, 1992) - sounds like our definition of intelligence
  - applying rules of logic
  - taking informal/unstructured knowledge and attempting to convert it to logical form

EXAMPLE - THINKING RATIONALLY (WEAK)

- Alexa, Siri, IBM Watson
  - accept natural speech input in a prescribed format
  - parse input for key words or phrases
  - **determine and apply logical rules**
  - format database search
  - format search results in human-like responses



Advanced Search

Find pages with...		To do this in the search box
all these words:		Type the important words: `tricolor rat terrier`
this exact word or phrase:		Put exact words in quotes: `"rat terrier"`
any of these words:		Type OR between all the words you want: `miniature OR standard`
none of these words:		Put a minus sign just before words you don't want: `-rodent, -"Jack Russell"`
numbers ranging from:	to	Put 2 periods between the numbers and add a unit of measure: `10..35 lb, $300..$500, 2010..2011`

Then narrow your results by...

language:	any language	Find pages in the language you select.
region:	any region	Find pages published in a particular region.
last update:	anytime	Find pages updated within the time you specify.
site or domain:		Search one site (like `wikipedia.org`) or limit your results to a domain like `.edu`, `.org` or `.gov`
terms appearing:	anywhere in the page	Search for terms in the whole page, page title, or web address, or links to the page you're looking for.
SafeSearch:	Hide explicit results	Tell SafeSearch whether to filter sexually explicit content.
file type:	any format	Find pages in the format you prefer.
usage rights:	not filtered by license	Find pages you are free to use yourself.

IS THIS COURSE?

“A Guide to ~~Living in~~ a World with Artificial Intelligence”

or

“A Guide to Living ~~in a World~~ with Artificial Intelligence”

or does it not matter?

It may depend upon your goals

BUILDING AN AI TAXONOMY (4/4)

- acting rationally
  - “computational intelligence is the study of the design of intelligent agents” (Poole, 1998)
  - acts to achieve the best outcome
  - if there is uncertainty, choose the best expected outcome

EXAMPLE - ACTING RATIONALLY (WEAK)

- self-driving cars/autonomous vehicles
  - accept input from wide variety of sensors in different formats (vision, sound, environment, gauges, etc.)
  - prioritize/equate and evaluate input according to specific rules and logic
  - respond based upon optimum results - even if it means something bad

ALAN TURING - THE IMITATION GAME - CLASSIC TEST OF “ACTING HUMANLY”

VOL. LIX. No. 236.] [October, 1950

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING

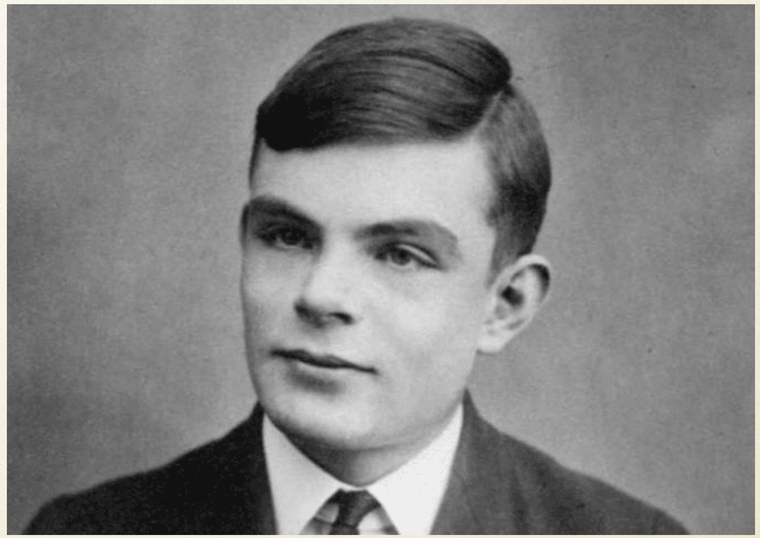
1. *The Imitation Game.*

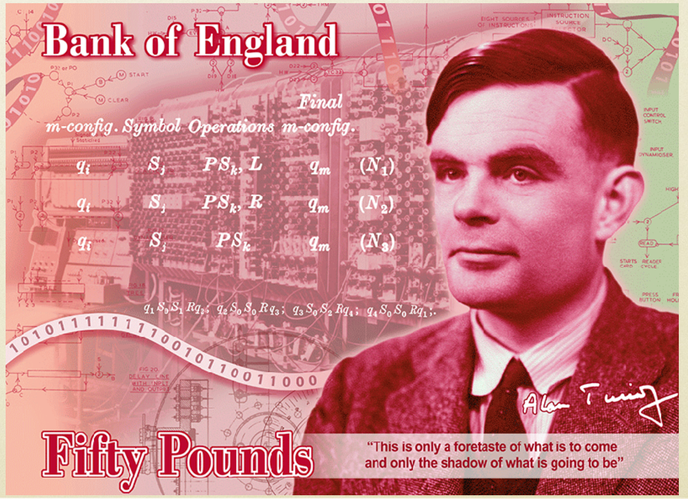
I PROPOSE to consider the question, ‘Can machines think?’ This should begin with definitions of the meaning of the terms ‘machine’ and ‘think’. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words ‘machine’ and ‘think’ are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, ‘Can machines think?’ is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem 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. He knows them by labels X and Y, and at the end of the game he says either ‘X is A and Y is B’ or ‘X is B and Y is A’. The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

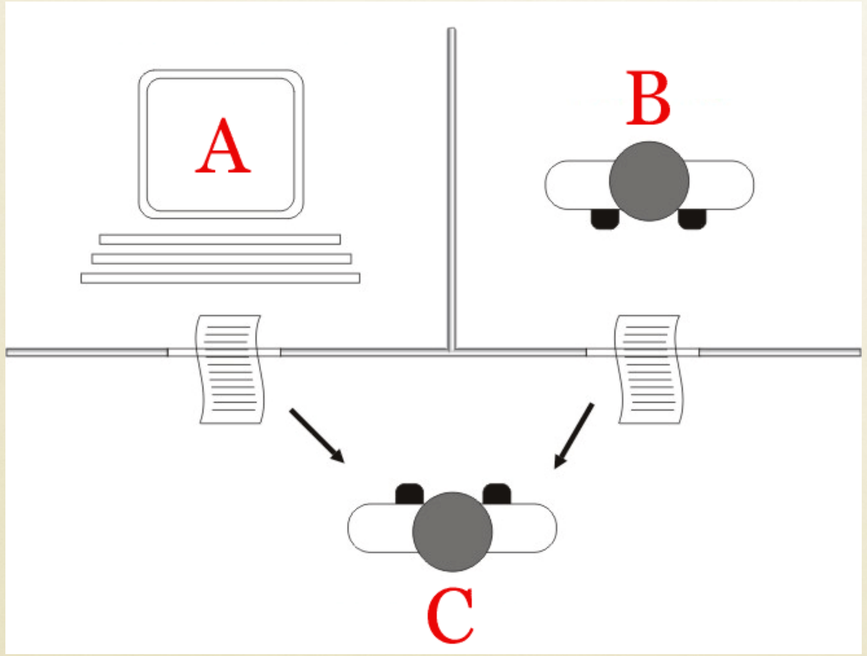
Now suppose X is actually A, then A must answer. It is A’s





THE TURING TEST (1950)

- "The Imitation Game" - "can a machine exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human?"



Not quite - but one of four is not bad…

GOOD QUESTIONS FOR A TURING TEST?

1. How come time flies like an arrow but fruit flies like a banana? (can an AI parse “flies like” in two totally different contexts?)
2. The following sentence is true. The previous sentence is false. Which of those two sentences is true? (would an AI get hung up in trying to determine the truthfulness of the sentences?)
3. Would you rather sacrifice one adult to save two children, or two children to save five adults, and why? (no definitive answer - a human would answer based on a mixture of logic and emotion; an AI would ???)

<u>Bebo to the Alexa on his desk:</u> “Do you pass the Turing Test?”

<u>Alexa to Bebo:</u> “I don’t need to. I’m not pretending to be a human.”


To be continued...