ARTIFICIAL INTELLIGENCE

The Very Idea

Vasant G. Honavar
Dorothy Foehr Huck and J. Lloyd Huck Chair in Biomedical Data Sciences and Artificial Intelligence
Professor of Data Sciences, Informatics, Computer Science, Bioinformatics & Genomics and Neuroscience
Director, Artificial Intelligence Research Laboratory
Director, Center for Artificial Intelligence Foundations and Scientific Applications
Associate Director, Institute for Computational and Data Sciences
Pennsylvania State University

vhonavar@psu.edu
http://faculty.ist.psu.edu/vhonavar
http://ailab.ist.psu.edu

On intelligence

- What is intelligence?
- How can we measure it?
- How do we know that a machine (or for that matter, human) is intelligent?

On intelligence

- We use the word ``intelligence'' routinely in everyday conversation
- We can look it up in a dictionary
- Can we precisely define intelligence?
- There is no widely agreed-upon scientific definition of intelligence
- Can you list some attributes of intelligence?

Can you list some attributes of intelligence?

Exercise

- Take 2-3 minutes to reflect, talk to those at your table
- Come up with your "top ten" list of attributes of intelligence

On intelligence

- Hold beliefs, desires, and intentions
- Perceive and model the external world
- Solve problems, break complex problems into simpler sub-problems
- Make sense of ambiguous or contradictory information
- Draw distinctions between situations despite similarities
- Act in the world to achieve one's goals
- Reason from facts or assumptions to conclusions
- Explain and justify decisions and actions
- Adapt to changes in the world, changes in the goals, etc.

On intelligence

- Learn, that is, acquire knowledge and skills from experience
- Generalize (find a common underlying pattern in superficially distinct situations)
- Communicate with other intelligent agents using signals, signs, symbols, speech, touch, gestures
- Understand and use language and other symbolic tools
- Coordinate with others to achieve shared goals, form and work in teams
- Plan and predict the consequences of contemplated actions, or action sequences including the ability to compare and evaluate alternatives

On intelligence

- Build and use tools to achieve one's goals
- Know limits of own knowledge and abilities
- Choose own goals or preferences (autonomy)
- Organize into social groups based on shared objectives develop shared conventions to facilitate orderly interaction, culture.
- Synthesize new concepts and ideas, and acquire and use analogies
- Create original works of literature, art, music, mathematics, scientific theories
- Introspect on internal goals, plans, actions, etc. of self as well as other agents
- …

On intelligence

- Most would probably agree that the hallmark of intelligence is
  - not simply the ability to display some or all of the listed attributes but
  - doing so on a broad and open-ended set of domains and
  - under a broad range of settings
- Different systems can display different subsets of the attributes of intelligence to differing degrees.

On intelligence

- Despite all the progress in AI over the past several decades
- Despite all the recent hype about AI
- No existing AI system comes close to exhibiting all the attributes of intelligence except in narrowly restricted domains
  - Playing Chess
  - Classifying tumors
  - Planning trips
  - Solving high school algebra problems

Theories of intelligence

- Theories of intelligence are primarily concerned with
  - identifying the major components of intelligent behavior, e.g., goals, knowledge
  - how they are represented and used
  - mechanisms, processes, and the role of different components and of
  - Interactions between components

Theories of intelligence: Performance theories

- View intelligence as a collection of abilities
- Measure an individual's performance through a single ability index
- Aim to address questions such as:
  - How can one test for the presence or degree of intelligence?
  - What are the essential components or functionalities a system capable of intelligence?

Theories of intelligence: Structural/functional theories

- Aim to answer questions such as:
- What are the mechanisms that underlie intelligence?
  - Computational theory of intelligence posits that intelligence is realized by computation or algorithmic information processing as the underlying mechanism
  - Biological theory of intelligence treats intelligence as a product of brain structure and brain function, and their relationship

Theories of intelligence: Cognitive theories

- Focus on the processes that underlie intelligence
- The processes range from simple to complex
  - Comparing items
  - Memory recall
  - Problem solving
- Aspects of processes considered include
  - Speed
  - Accuracy

  ...

Theories of intelligence: Contextual theories

- Focus on intelligence demonstrated in a given social, cultural, or environmental context
- Aim to address the question:
  - How is one's intelligence related to the environment with which one must cope with?

Theories of intelligence: Existence theories

- Focus on the necessary and/or sufficient conditions for intelligent behavior to be possible
  - Under what conditions can one
    - Learn a concept from data?
    - Learn an adversarial game like Chess?
    - Repair faulty plans?
    - Cooperate with others to solve a complex problem?
    - Act rationally?
    - ….

Theories of intelligence: Developmental theories

- focus on the developmental processes of different competencies necessary for intelligence.
  - Piaget's theory of cognitive development posits the existence of developmental stages during which
  - the body and the brain undergo changes that support the development of cognitive skills, language, and reasoning abilities
  - by interacting with the environment, acquiring and processing representations or models of the world.

Two major theories

Sternberg's Triarchic Theory

- Triarchic Theory posits that there are three aspects of intelligence
  - Analytical Intelligence (book smart)
  - Creative Intelligence
  - Practical Intelligence (street smart)

Sternberg's Triarchic Theory: Analytical Intelligence

- Analytical Intelligence (book smart)
  - Logically derive conclusions from premises
  - Decompose complex problems into more manageable simpler pieces,
  - Solve mathematical problems
- Analytical intelligence relies on the ability to
  - Acquire and organize knowledge,
  - Execute and control various cognitive processes
  - Acquire and use skills needed to perform the tasks at hand
    - Knowing Newton's laws of physics does not immediately translate into a facility for solving complex physics problems

Sternberg's Triarchic Theory

- Creative Intelligence refers to the ability to draw on experience to solve novel or unique problems
  - conceiving and designing a spacecraft where none existed
  - creating a new architectural style
  - using novel materials and processes to design useful artifacts
  - creating new music or an entirely new musical genre, e.g., Charles Buddy Bolden and his band members' introduction in the early 1990's of music that later came to be known as Jazz
  - creating a new style of art, e.g., impressionism developed by Claude Monet and other Paris-based artists in the 1860's.

Sternberg's Triarchic Theory

- Practical Intelligence (street smart) refers to
  - the ability to adapt to one's environmental or cultural contexts
  - shape the environment to meet personal or societal needs
- Examples of practical intelligence include
  - proficiency in communicating across cultures
  - resolving conflicts
  - leading diverse teams
  - bringing about organizational change.

Gardner's theory of multiple intelligences

- Intelligence is a "potential to process information that can be activated in a cultural setting to solve problems or create products that are of value in a culture"
- Verbal-linguistic intelligence
- Logical-mathematical intelligence
- Spatial-visual intelligence
- Bodily-kinesthetic intelligence
- Musical intelligence
- Intrapersonal intelligence
- Interpersonal intelligence
- Naturalist intelligence

MI Theory claims that there are significant variation across individuals in their levels of strength and weakness across the different intelligences

Can we measure intelligence?

- Intelligence is hard to define
- Can we recognize it when we see it?
- Can we measure intelligence?
- Can we tell if a system is intelligent?

Measuring Intelligence

Terman-Merrill revision of the Binet-Simon intelligence test includes questions in the following categories:

- Obey simple commands
- Identify object by use
- Repeat digits
- Respond to pictures
- Repeat digits reversed
- Memorize key events and actors in stories
- Find absurdities in pictures
- Recognize pictorial objects
- Comprehend text
- Answer opposite analogies
- Identify alike and different pictures
- Memorize details in sentences
- Use vocabulary
- Complete (human) picture
- Discriminate animal pictures

What are the tests good for?

- What the test scores measure is neither innate nor permanent
- The test offer a rough scale for identifying learning disabled children needing special help
- The test is useless for ranking "normal" children
- Low scores should not be used to label children as incapable but for helping them
- Tests are designed with predetermined answers
- Child's ability to score well depends on cultural social and economic factors (and not innate abilities)

Are there good tests good for machine intelligence?

- Can we simply give machines the same IQ tests we give to humans?
- Two problems
  - We know the problems with IQ tests
  - We have AI systems that can get high score on some IQ tests while failing to exhibit most attributes of intelligence
    - We can design AI systems that can ace high school math tests etc.
- Turing had an alternative idea – The Turing Test

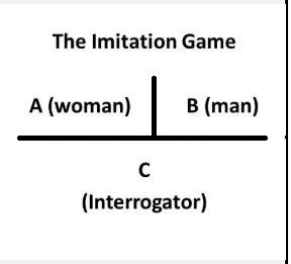
If a machine could think, how could we tell?

- Turing's idea
  - If in a natural language dialog, the responses from the computer were indistinguishable from that of a human, the computer could be said to be thinking
- Turing's prediction
  - Turing predicted that by the year 2000 a computer program would be able to fool the average questioner for five minutes about 70% of the time

Turing, A.M Computing Machinery and Intelligence. *Mind*, Volume LIX, Issue 236, 1950, Pages 433–460

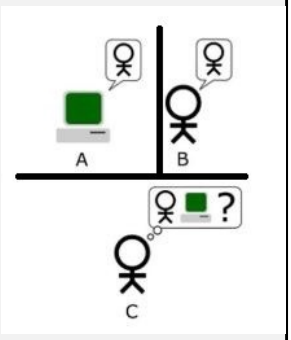
Turing Test – Inspired by the Imitation Game

- Turing test is a modification of imitation game that was popular among students at Cambridge
  - There are three people in three separate rooms. A woman (A), a man (B, and an interrogator (C).
  - None of them knows each other's gender.
  - Interrogator can pass typed questions to A or B
  - A and B can communicate only with C by passing typed answers.
  - C asks questions to A and B
  - A and B seek to convince C that he/she is a woman



Turing Test

- Replace one of the two players (say A) in the imitation game by a computer
- Interrogator can pass typed questions to A or B
- A and B can communicate only with C by passing typed answers.
- C asks questions to A and B
- A and B seek to convince C that he/she is a human
- If A convinces C that it is a human as often as B does, A (the computer) passes the Turing Test



Turing Test – A slight variation

- Place a computer in one room and a panel of human interrogators in another room
- The interrogators do not know whether a human or a computer is in the first room
- Interrogators on the panel can pass typed questions to the first room
- The occupant of the room returns typed answers to the panel of interrogators
- If the computer can convince at least 30% of the panel of interrogators that it is a human, it passes the Turing Test
- Why 30%? Why not?

Chatbots

- Many chatbots were developed to
  - Explore natural language communication between humans and machines
  - To pass - or ridicule – the Turing test
  - These are computer programs which simulate conversation with a user in a (written) natural language
  - The most well-known example is ELIZA, developed by Joseph Weizenbaum of MIT in 1966

ELIZA – The first chatbot

- In 1966 Joseph Weizenbaum of MIT wrote the first chatbot
- He called it ELIZA and cast it in the role of a Rogerian psychiatrist
- Sample conversation with ELIZA

  Person: Men are all alike.
  Chatbot: IN WHAT WAY?
  Person: They're always bugging us about something or other.
  Chatbot: CAN YOU THINK OF A SPECIFIC EXAMPLE?
  Person: Well, my boyfriend made me come here.
  Chatbot: YOUR BOYFRIEND MADE YOU COME HERE?
  Person: He says I'm depressed much of the time.
  Chatbot: I AM SORRY TO HEAR YOU ARE DEPRESSED

- https://eliza.botlibre.com

How ELIZA works

- ELIZA simulates a conversation between psychiatrist and patient
  - It mirrors the patient's feelings by recasting the patient's statements as questions
- Looks for patterns of words in the input
- Applies a series of pattern matching rules
- Replies with output based on a predefined set of rules
- Uses a very small database of words and phrases
- ELIZA is as effective as it is –
  - not because it understands English
  - or because it knows psychotherapy
  - but because humans read so much meaning into what is said
- ELIZA fools human into interpreting nonsense as conversation
- Modern SIRI is not that much better!

ELIZA

- To Weizenbaum's horror
  - some users were convinced that they were really talking to a person
  - Some psychiatrists were willing to try ELIZA in actual therapy setting
- Weizenbaum was horrified and became concerned about the ethics of AI and eventually one of the most fervent critics of AI
- Nevertheless many variants of ELIZA-like chatbots followed
- Chatbots have their uses, but they must be designed and used under ethics oversight

The Loebner prize competition

- The first formal instantiation of a Turing test
- Started in 1991 by Dr. Hugh Loebner
  - Loebner pledged $100,000 to whoever programmed a computer that could pass the test
- The competition ran every year from 1991 to 2020
- Because passing the Turing test was considered hard, to encourage progress and participation in the competition,
  - Each year an annual prize of $2000 and a bronze medal was awarded to the best performer irrespective of how good it is in an absolute sense
- https://en.wikipedia.org/wiki/Loebner_Prize

Sample questions from the Loebner competition

- What is your most memorable childhood event and how has that impacted you today?
- Describe yourself using only colors and shapes.
- Describe why time flies like an arrow but fruit flies like a banana?
- How do you feel when you think about your upbringing and what makes you feel that way?
- What historical event changed you the most and where were you when it happened?
- Which of the previous questions was the most difficult to answer and why?

The Loebner competition

- The competition ran every year from 1991 to 2020
- The competition was controversial in the AI community
- Many criticized it as a distraction – Minsky jokingly offered a prize to anyone who would stop the competition

On Turing Test - Discussion

- Exercise
  - Is the Turing test a good test for determining whether the computer is intelligent? Or that the computer thinks?
  - Why or why not?
  - Reflect on this for 2-3 minutes, discuss with those at your table, and summarize your conclusions

Some remarks about the validity of the test

- Supporters
  - Human-like interaction is essential to human-like intelligence
- Opponents
  - Passing the Turing Test is neither necessary nor sufficient for intelligence
  - It is not necessary for program to talk to us in our language for it to be intelligent
  - There are humans that would fail the Turing test
  - Searle's Chinese Room Argument

Searle's Chinese Room Argument

Image source: Alan Tan

Searle's Chinese Room Argument

- You see a room.
- Sometimes you see some Chinese friends of yours come to the room with a piece of paper which they slip into the room through a slot.
- Then they wait a while until the same piece of paper comes back out of a room through a second slot.
- You soon learn that the questions slipped into the room were written in Chinese.
- Suppose the question was

什麼是生命？

What is life?

Searle's Chinese Room Argument

- When the paper is later passed out of the room, your Chinese friends find that an answer to their question has been written on the slip.

  生命是一系列自然而自發的改變。不要抗拒它們——那隻會帶來悲傷。
  讓現實成為現實。讓事情以任何他們喜歡的方式自然地向前發展。

  Life is a series of natural and spontaneous changes. Don't resist them — that only creates sorrow. Let reality be reality. Let things flow naturally forward in whatever way they like.
- The answer is also in Chinese and your friends, all fluent in Chinese agree that it is, in fact, a wise answer to their question – a direct quote from the writings of the great Lao-Tzu himself.
- Your Chinese friends continue to pass questions
- They receive what seem to be intelligent answers to their questions
- They conclude that the person inside the room who is answering their questions must be an intelligent person who understands Chinese

Searle's Chinese Room Argument

- Suppose then you find out that the only person inside the room was your not-so-smart brother who understands only English and has absolutely no knowledge of Chinese.
- You ask your brother how he managed to answer the questions so wisely.
- He tells you that he was given a set of books by his boss and instructed to look through the books until he found the string of symbols that look exactly like the ones written on the piece of paper.
- And when he finds that string of symbols, the book will tell him (in English) what new string of symbols he is to write in response.

If you see this shape, "什麼" followed by this shape, "帶來" followed by this shape, "快樂"	then produce this shape, "爲天" followed by this shape, "下式".

Searle's Chinese Room Argument

- Note that the only sentences on the page that your brother, as an English speaker, can understand are in English.
- This book is one of countless volumes that together tell your brother what output (in the form of Chinese symbols) should be given in response to virtually any input (of Chinese symbols) that comes through the slot into the room.
- This particular volume tells what output to give in response to virtually any input of Chinese symbols that *begins* with the first two Chinese symbols written on the piece of paper.

If you see this shape, "什麼" followed by this shape, "帶來" followed by this shape, "快樂"	then produce this shape, "爲天" followed by this shape, "下式".

Searle's Chinese Room Argument

- Your brother tells you he does not recognize any of the symbols.
- You know that your brother is truthful and never lies
- They are simply meaningless shapes to him.
- For all he knows, they may be nothing more than patterns for making wall-paper and not a language at all.
- As it turns out, the symbols do have meanings. They are Chinese symbols.
- More than that, they describe questions in Chinese being asked by your Chinese friends.

Searle's Chinese Room Argument

- The books function as a computer program. Each page gives instructions for how to manipulate symbols.
- The instructions at no point make any reference to the *meaning* of the symbols.
- That is, nowhere will you find a sentence that gives the English translation of any of the Chinese symbols.
- None of these books is anything like a Chinese-English dictionary.
- Instead, like a computer program itself, it instructs the reader how to manipulate the symbols based on their properties (their shape and position) not their meaning.
- If you see symbol "X" here, then write symbol "Y" there.

Searle's Chinese Room Argument

- Let's imagine that instead of your brother in the room, we have an AI program that passes the Turing Test in Chinese, just like your brother did
- Should we say that the AI program understands Chinese?
- No, says Searle.
- What do you think?

Searle's Chinese Room Argument

- Searle's claim is that he has demonstrated that no computer program that manipulates symbols based solely on their formal "syntactic" properties (e.g., their shape and their position) can ever be said to understand a language . . even if it *does* pass The Turing Test.
- Searle is *not* saying that no machine can understand a language.
- He is not even saying that no computer could understand a language.
- Searle is willing to concede that certain operations of the human brain can properly be described as "computation."
- Searle is not insisting that humans do *not* compute and do not implement functions.

Searle's Chinese Room Argument

- Searle is insisting, however, that genuine thought and understand require something *more* than mere computation.
- He is saying that understanding a language we do not *merely* manipulate symbols based on their formal properties.
- That is, we do something (he doesn't say what) in addition to manipulating symbols by virtue of which we actually understand the *meaning* of the symbols – which your brother or the computer did not.

Searle's Chinese Room Argument

- Searle is insisting, however, that genuine thought and understand require something *more* than mere computation.
- He is saying that understanding a language we do not *merely* manipulate symbols based on their formal properties.
- That is, we do something (he doesn't say what) in addition to manipulating symbols in virtue of which we actually understand the *meaning* of the symbols – which your brother or the computer did not.

Discuss: Searle's Chinese Room Argument

- Is Searle right in that a process that simply syntactically manipulates its inputs to produce its outputs cannot be intelligent?
- What does this say about the working hypothesis of AI?
  - Thinking can be modeled by computation (weak version)
  - Thinking is computation (strong version)
- Does the Chinese Room Argument suggest a more effective test of machine intelligence?
- Discuss