ARTIFICIAL INTELLIGENCE

The Very Idea

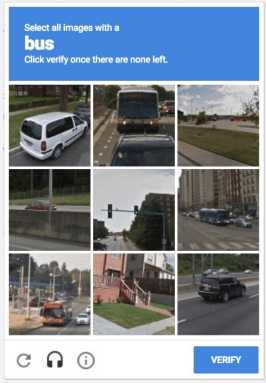
Vasant G. Honavar
Dorothy Foehr Huck and J. Lloyd Huck Chair in Biomedical Data Sciences and Artificial Intelligence
Professor of Data Sciences, Informatics, Computer Science, Bioinformatics & Genomics and Neuroscience
Director, Artificial Intelligence Research Laboratory
Director, Center for Artificial 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

Alternatives to (the original) Turing Test

Variations on the Turing Test – Reverse Turing Test

- In the reverse Turing Test (RTT), the task is for the humans to convince the computer that they are human.
- Example of a reverse Turing Test
  - CAPTCHA security measures that you've likely encountered when signing onto websites
- We can envision structuring the RTT to assess the intelligence of the machine by its success in correctly telling apart humans from machines
- While the reverse Turing Test



Variations on the Turing Test – Winograd Schema Test

- Winograd Schema Challenge, named for AI pioneer Terry Winograd offers a test combines natural language dialog and common sense.
- "The city councilmen refused the demonstrators a permit because they [feared/advocated] violence.
- Which noun phrase does "they" refer to in the two sentences?
  - "The city councilmen refused the demonstrators a permit because they feared violence."
  - "The city councilmen refused the demonstrators a permit because they advocated violence."
- Getting the answer right seems to require commonsense understanding of how councilmen and demonstrators behave

What are valid Winograd schema?

- The two sentences are identical except for one or two words, e.g., “feared” versus “advocated” in the example above.
- The two sentences both contain two noun phrases and a pronoun, e.g., “the city council”, “the demonstrators”, and “they” in the example above.
- In reading the two sentences in isolation, humans would associate different noun phrases to the pronoun
- Both sentences must be easily understood by the human reader,
- For humans, the answer to the WSC is so obvious that the potential ambiguity goes unnoticed

What are valid Winograd schema?

- "Selection restrictions" should not suffice for correct disambiguation.
  - Prohibited: The women stopped taking the pills because they were [pregnant/carcinogenic]
  - Why? – Too easy!
- Pills cannot be pregnant and women cannot be carcinogenic!
  - Disambiguation only requires matching the features of the candidate referents pills and women with pregnant and carcinogenic

What are valid Winograd schema?

- Matching based on co-occurrence probabilities of words will not suffice for correct disambiguation.
- Disallowed: The race car zoomed by the golf cart because it was going too [fast/slow].
- Why? – Too easy!
  - "race car" and "fast" tend to co-occur in text as do ``golf cart'' and ``slow.''
- Practical challenge:
  - Constructing good schema requires manual effort
  - The validity conditions are not always easy to verify

Winograd Schema Challenge – Quo Vadis?

- Many human-curated data sets of WSC created
- By 2021, large language models trained on large text corpora were able to match human performance
- A more extensive version of the challenge with 44,000 problems was developed - Winogrande
- By 2024 large language models matched human performance on Winogrande

What does machine defeat of the Winograd Schema Challenge Tell Us about Machine Intelligence?

- Does the success of machines at beating the Winograd Schema Challenge mean that
  - Machines have become intelligent?
  - Machines have managed to acquire commonsense?
- What could explain the good performance of machines?
- Discuss

How was Winograd Schema Challenge defeated?

- Large text corpora contain stereotypical usage of most sentences
- Large language models (LLM) trained on such corpora become good at predicting the occurrence of words in the context of other words in the sentence
- Because of the size of the corpus, LLM can go beyond relationships between specific words, and generalize to relationship between one class of words and another – in other words, sentence schema
  - Horse raced past the barn fell
  - Boy pushed over the fence slipped
- LLM don't understand natural language any more than Searle's Chinese Room does
- LLM have no commonsense understanding of the world
- LLM display language competence without language comprehension

Visual Turing Test challenge

Challenge Questions

- Who do you see?
- What is she doing?
- What is on the desk?
- Is she indoors or outdoors?
- What color are the walls?
- How many chairs are there?

Visual Turing Test challenge

What is involved?

- Face recognition
- Object identification
- Activity recognition
- Spatial reasoning
- General knowledge
- Color perception
- Counting
- Natural Language

Discuss: What does machine success at Visual Turing Test tell us about Machine Intelligence?

- Suppose machines have passed the Visual Turing Test
- Does passing the Visual Turing Test mean that
  - Machines have become intelligent?
  - Machines have managed to acquire commonsense?
- What could explain the good performance of machines?
- Discuss

Visual Turing Test – Quo Vadis?

- Several versions of the Visual Turing Test available
- In recent years neural network models trained on large data sets of images and associated captions do impressively well
- Details matter
  - Data sets
  - Methods
  - Evaluation metrics

Video Turing Test – The Marcus Version

Watch an episode of Young Sheldon

- Summarize the episode
- Answer questions
  - Why did Sheldon want build rockets?
  - Why did his dad tell him he could not launch his rocket?
  - Why did his mom change her mind about allowing him to launch rockets?
  - Why did the FBI show up to talk to Sheldon?

Video Turing Test

What is involved?

- Recognize individuals
- Activity recognition
- Reasoning about emotions
- Causes and effects not shown
- Comprehending conversational context
- Recognizing and reasoning about intensions, goals, etc.
- Commonsense

Video Turing Test – Marcus Test

- The computer is asked to watch any arbitrary TV program or YouTube video and answer questions about its content —
  - 'Why did Russia invade Crimea?
  - Why did Jordan Chiles lose her Olympic medal after it was awarded to her?

Discuss: Video Turing Test

- What makes video Turing Test more challenging than the other tests discussed so far?

Video Turing Test – Quo Vadis?

- Not quite passed by machine
- More challenging than the basic Turing Test, Winograd schema challenge, Visual Turing Test
- All of the information needed to understand the video is not in the video
- High quality annotated data hard to come by for brute force machine learning to work

Variations on the Turing Test – Standardized testing for AI

- Simply administer to machines the same standardized tests we use for various purposes with humans
  - PSSA used by Pennsylvania schools
  - SAT
  - ACT
  - GRE
  - MCAT

Discuss: Standardized tests

- What would it take for machine to ace the standardized tests?
- Will passing the test imply the machine is intelligent?

Lovelace Test

- Motivated by Ada Lovelace's critique of the Turing Test
- Consider a machine programmed by a human
- To pass the test, machine must display creativity in some domain under constraints given by an evaluator
  - Create an artifact in the domain
  - The artifact must meet the specified constraints
  - Human evaluator determines whether the machine's creation is acceptable
  - A referee confirms that the domain and constraints are not unrealistic for an average human
  - The creation as aesthetic appeal

Lovelace Test

Examples

- Music composition
- Fictional story generation
- Poetry production
- Paintings
- Sculpture
- Novel math proofs
- ….

Lovelace Test

Story generation requires

- Familiarity with existing literature
- Familiarity with culture, social norms
- Commonsense knowledge
- Discourse planning
- Natural language competence
- Reason about people, their feelings, emotions, goals, mental states, intentions

Lovelace test – Quo Vadis?

Story Generation

One day there was a girl named Anita who wanted to have an enjoyable adventure. She did not know what an adventure was but thought that a walk by herself would be fine.

Anita asked Mother if she might go for a walk. Mother, tired from her work, replied, "No. But if you must go out, do not go anywhere or do anything. Just walk around the house, and be careful." Anita, not pleased with that answer, wandered out to the front gate. There she saw her best friend, a brown and white chicken named Felipe.

Anita and Felipe had great times together. Felipe loved to eat corn. Anita would feed him corn and then talk to him. Felipe loved conversations and would answer with little sounds and noises, while bobbing his head up and down.

"Felipe," exclaimed Anita, "let's have an adventure together!" Felipe continued to eat and listen, for a chicken, like Anita, enjoys talking. Anita then asked, "What is an adventure?" Felipe was very wise for his size and age and said, "I do not know. But you can count on your friends."

Lovelace test – Quo Vadis?

Poem in the style of Gerard Manley Hopkins

Upon the vernal stage, a tapestry unfolds,
Each petal poised, a chromatic hymn,
Where verdant hues in nature's hands are thinned,
A ballet of rebirth, as life behold.
In dew-kissed morn, the skylark take to air,
Its trill, a sonnet sung to buds unseen,
Awakening the earth from winter's dream,
As caucus bursts, a solace to despair.
The daffodils, with golden heads held high,
In dappled glades, a dance of sunlit flame,
As zephyrs waltz, a gentle, fragrant sigh,
And butterflies on wings of art proclaim.
Oh, Spring! Thy brush, a stroke of divine grace,
Renews the world, a masterpiece in grace.

Lovelace test – Quo Vadis?

Painting in the style of Monet

Discuss: Lovelace Test

- What would it take for machine to pass the Lovelace test?
- Will passing the test imply the machine is intelligent?

Lovelace Test

- While machines may not quite pass the Lovelace test with flying colors, they are beginning to look like serious contenders
- Success, as in the case of Winograd Schema Challenge, Visual Turing Test, etc., has come from training generative models on large corpora

Social-Emotional Turing Challenge

- A test of what Sternberg called street smart and what Gardner called interpersonal intelligence
- Scenario: Tracy asks for a banana; Mom gives Tracy an apple
- Question: How will Tracy feel?
- Possible responses:
  - Tracy feels sad because she did not get what she wanted
  - Tracy feels happy because she asked for a banana which is what her mom usually gives her, but she was bored of eating banana pretty much everyday so was excited to try apple for a change
- Evaluation: Rating of machine generated and human generated responses by neutral evaluators

Scientific AI Challenge

Scientific discovery offers a grand challenge for AI

- AI Assistants for Scientists
- Autonomous Robotic Scientists

Representative AI associates for scientists

A scientist's associate that

- Learns what you and others in your field and related fields are are working on
- Finds and reads relevant literature
- Locates and ingests available knowledge and data
- Offers assistance
  - Here are some data that contradict your hypothesis
  - Here are arguments for and against your hypothesis
  - Here is some data from lab X that explains your finding
  - Here is why you should prefer model A to model B

Examples of AI associates for scientists

A scientist's associate that:

- Given background knowledge and access to literature in two or more domains, use literature in one domain to generate hypotheses or explanations in another domain?
- Given causal information learned from restricted experiments in multiple settings (Los Angeles, Chicago), when possible, infer causal information in a target setting (say New York)?
- Given a scientific question, and a social network of researchers and their scientific output e.g., publications, identify a collection of researchers that are best equipped to address the question?

Success criteria

- For an associate
  - Ratings by human scientists
  - Productivity of human scientist with and without AI associate
- For an AI scientist
  - Scientific breakthroughs or findings that are deemed worthy of publication in reputed scientific venues
  - Scientific impact

Variations on the Turing Test – Physically Embodied Turing Test

- A robot has to physically manipulate real-world objects in meaningful ways
  - Build a structure from a pile of parts using verbal, written instructions and pictures (imagine assembling IKEA furniture)
- A robot has to devise solutions to a set of open-ended but increasingly creative challenges using toy blocks
  - Build a wall
  - Build a house
  - Attach a garage to the house
- The robot is required to "explain" its efforts
  - "The door needs to be installed before the wall around it; otherwise it wouldn't fit

Harnad Test or Total Turing Test

- To have any hope of passing the Total Turing Test,
  - The machine has to be more than a computer
  - It has to be robot with the sensors, effectors, and
    - the physical wherewithal to interact with the physical world with real objects, real people, etc.
  - If the robot can do everything a body with a mind can do such that we can't tell them apart, we have no basis for doubting it has a mind.
  - Challenges:
    - What does "everything a body with a mind can do"?
    - Eating? Sleeping? Playing basketball?
    - How to operationalize the test?

Discuss: Total Turing Test

- How can we operationalize Total Turing Test?
- If a robot passes the the Total Turing Test, does it mean that the robot is intelligent?
  - Why or why not?

Limitations shared by all forms of Turing Test

- Any attempt to assess machine intelligence must confront the other-minds problem
  - How can we tell whether any body other than our own has a mind when the only way to know is by being the other body?
  - We are left with some form of Turing Test:
    - If it can do everything a body with a mind can do such that we can't tell them apart, we have no basis for doubting it has a mind.
  - But because of the other minds problem, we are left with something like

    "If it looks like a duck, it quacks like a duck, it must be a duck"
    - Strong AI hypothesis is untestable.
    - Weak AI hypothesis is testable