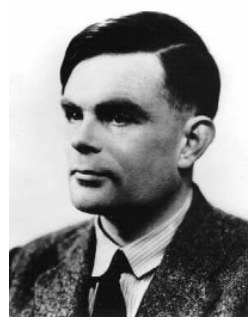
Turing Centenary Celebration

Turing's Test for Artificial Intelligence

Dr. Kevin Korb
Clayton School of Info Tech
Building 63, Rm 205
kbkorb@gmail.com

Can Machines Think? Yes

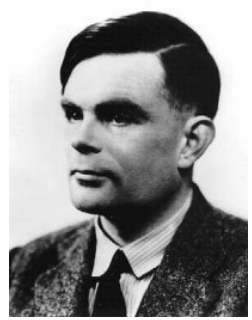
Alan Turing's question (and answer) in his "Computing Machinery and Intelligence" (*Mind*, 1950) — the first paper on Artificial Intelligence (AI).



- Inventor of Computer Science and Turing machines (1936-7)
- Breaker of the Nazi Enigma code (1940-44)
- Designer of the ACE computer (1945-47)
- Inventor of AI (1945-50)
- Convicted of homosexuality; forced estrogen injections; suicided (1954)

Can Machines Think? Yes

Alan Turing's question (and answer) in his "Computing Machinery and Intelligence" (*Mind*, 1950) — the first paper on Artificial Intelligence (AI).



- Inventor of Computer Science and Turing machines (1936-7)
- Breaker of the Nazi Enigma code (1940-44)
- Designer of the ACE computer (1945-47)
- Inventor of AI (1945-50)
- Convicted of homosexuality; forced estrogen injections; suicided (1954)

Machine Intelligence

At Bletchley Park, in his spare time, Turing wrote the first "AI program"

1. A chess program. . .
2. Testing it with the only computer available in his hut

Machine Intelligence

At Bletchley Park, in his spare time, Turing wrote the first "AI program"

1. A chess program. . .
2. Testing it with the only computer available in his hut

Machine Intelligence

At Bletchley Park, in his spare time, Turing wrote the first "AI program"

1. A chess program...
2. Testing it with the only computer available in his hut

Machine Intelligence

During the war, and immediately after, Turing began testing the claim that a machine intelligence is possible.

1. Arguing with psychologists, philosophers, etc.
2. Without resolution, because “intelligence” couldn’t be pinned down

Leading him to formulate his (in)famous test for intelligence

Machine Intelligence

During the war, and immediately after, Turing began testing the claim that a machine intelligence is possible.

1. Arguing with psychologists, philosophers, etc.
2. Without resolution, because “intelligence” couldn’t be pinned down

Leading him to formulate his (in)famous test for intelligence

Machine Intelligence

During the war, and immediately after, Turing began testing the claim that a machine intelligence is possible.

1. Arguing with psychologists, philosophers, etc.
2. Without resolution, because “intelligence” couldn’t be pinned down

Leading him to formulate his (in)famous test for intelligence

Machine Intelligence

During the war, and immediately after, Turing began testing the claim that a machine intelligence is possible.

1. Arguing with psychologists, philosophers, etc.
2. Without resolution, because “intelligence” couldn’t be pinned down

Leading him to formulate his (in)famous test for intelligence

The Turing Test

To avoid endless argument, Turing proposed

The Imitation Game:

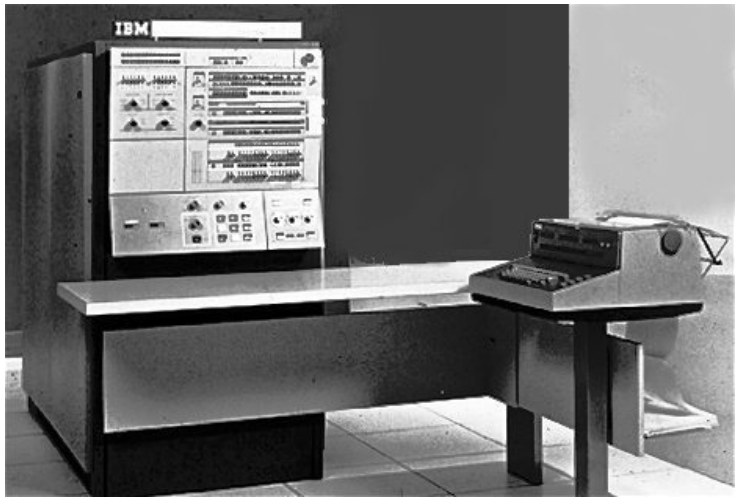
- Behind one curtain is a woman.
- Behind another curtain is a computer.

The Turing Test

To avoid endless argument, Turing proposed

The Imitation Game:

- Behind one curtain is a woman.
- Behind another curtain is a computer.



The Turing Test

Criterion: If, after five minutes, the interrogator has no better than a 50-50 chance of distinguishing woman and computer, the computer is intelligent.

> *Turing's Prediction: by 2000 AD, computers will be confused with humans 30% of the time.*

The Strong AI Thesis

— named by John Searle (1980)

The set of all possible Turing machines (computer programs) can be enumerated, simpler to more complex:

$$T_1, T_2, \ldots, T_k, \ldots$$

If AI is possible, at least one of these is an AI.
(Note that the complexity is unbounded. . .)

The Strong AI Thesis

— named by John Searle (1980)

The set of all possible Turing machines (computer programs) can be enumerated, simpler to more complex:

$$T_1, T_2, \ldots, T_k, \ldots$$

If AI is possible, at least one of these is an AI.
(Note that the complexity is unbounded. . .)

The Strong AI Thesis

— named by John Searle (1980)

The set of all possible Turing machines (computer programs) can be enumerated, simpler to more complex:

$$T_1, T_2, \ldots, T_k, \ldots$$

If AI is possible, at least one of these is an AI.
(Note that the complexity is unbounded. . .)

The Strong AI Thesis

Strong AI: Amongst the list of all possible Turing machines are all possible *intelligences*:

The Strong AI Thesis

Strong AI: Amongst the list of all possible Turing machines are all possible *intelligences*:

$$T_1, T_2, \ldots,$$

$$T_{korb}, \ldots, T$$  $$, \ldots, T$$  $$, \ldots$$

Can Machines Think? Yes

Turing's defence of AI:

- *Ada Lovelace's Objection:*

"The Analytical Engine has no pretensions to *originate* anything. It can do *whatever we know how to order it* to perform."

Likely translation: it can do all and only what we tell it to do. I.e., it's stupid.

Can Machines Think? Yes

Turing's defence of AI:

- *Ada Lovelace's Objection:*

"The Analytical Engine has no pretensions to *originate* anything. It can do *whatever we know how to order it* to perform."

Likely translation: it can do all and only what we tell it to do. I.e., it's stupid.

Can Machines Think? Yes

Turing's defence of AI:

- *Ada Lovelace's Objection:*

"The Analytical Engine has no pretensions to *originate* anything. It can do *whatever we know how to order it* to perform."

Likely translation: it can do all and only what we tell it to do. I.e., it's stupid.

Can Machines Think? Yes

Turing's defence of AI:

- *Ada Lovelace's Objection:*

"The Analytical Engine has no pretensions to *originate* anything. It can do *whatever we know how to order it* to perform."

Likely translation: it can do all and only what we tell it to do. I.e., it's stupid.

$\Rightarrow$ *Turing:* The Analytical Engine was the first universal machine, fully capable of being programmed to learn [Machine Learning]; hence, it was fully capable of doing things we do not know how to order.

Can Machines Think? Yes

- *The Mathematical Objection:* Gödel, Turing, etc. have shown that no purely formal system (e.g., computers) can prove (know) every mathematical truth, including truths *that humans know* about those systems. Therefore, humans are superior to machines. [cf. Roger Penrose]

Can Machines Think? Yes

- *The Mathematical Objection:* Gödel, Turing, etc. have shown that no purely formal system (e.g., computers) can prove (know) every mathematical truth, including truths *that humans know* about those systems. Therefore, humans are superior to machines. [cf. Roger Penrose]

$\Rightarrow$ *Turing:* "Although it is established that there are limitations to the powers of any particular machine, it has only been stated, without any sort of proof, that no such limitations apply to the human intellect."

If we develop an AI, it may very well know truths that humans cannot.

The Turing Test

ELIZA – Joe Weizenbaum

S: Are you working at home today?
E: What do you think?
S: Well, I suppose so. When will you be in next?
E: I'll ask the questions, if you don't mind!
S: Huh???
E: Can you elaborate on that?
S: I have some things for you to sign. When can you sign them?
E: I could ask the same thing myself.
S: Yea, but what would your answer be?
E: Is it because what would my answer be that you came to me?
S: Perhaps you should have a lie down.
E: Why do you say perhaps I should have a lie down?
S: You really are not yourself today.
E: Earlier you said what would my answer be?

The Turing Test

Unfortunately, the Turing Test is neither necessary nor sufficient as a test for intelligence:

- **Not necessary.** Presumably, Andromedans can be intelligent while incapable of mimicking women.
  - NB: Turing never suggested necessity!
- **Not sufficient.**
  - Imbecilic programs regularly come close to passing for imbecilic humans (ELIZA, PARRY)
  - Monkeys *may* type Shakespeare's complete works — there's more to it than random search.
    I.e., if you pass the test without *any* identifiable brains inside, something's fishy. (Cf. the chess "proof" of AI)

The Turing Test

Unfortunately, the Turing Test is neither necessary nor sufficient as a test for intelligence:

- **Not necessary.** Presumably, Andromedans can be intelligent while incapable of mimicking women.
  - NB: Turing never suggested necessity!
- **Not sufficient.**
  - Imbecilic programs regularly come close to passing for imbecilic humans (ELIZA, PARRY)
  - Monkeys *may* type Shakespeare's complete works — there's more to it than random search.
    I.e., if you pass the test without *any* identifiable brains inside, something's fishy. (Cf. the chess "proof" of AI)

The Turing Test

Unfortunately, the Turing Test is neither necessary nor sufficient as a test for intelligence:

- **Not necessary.** Presumably, Andromedans can be intelligent while incapable of mimicking women.
  - NB: Turing never suggested necessity!
- **Not sufficient.**
  - Imbecilic programs regularly come close to passing for imbecilic humans (ELIZA, PARRY)
  - Monkeys *may* type Shakespeare's complete works — there's more to it than random search.
    I.e., if you pass the test without *any* identifiable brains inside, something's fishy. (Cf. the chess "proof" of AI)

The Turing Test

Unfortunately, the Turing Test is neither necessary nor sufficient as a test for intelligence:

- **Not necessary.** Presumably, Andromedans can be intelligent while incapable of mimicking women.
  - NB: Turing never suggested necessity!
- **Not sufficient.**
  - Imbecilic programs regularly come close to passing for imbecilic humans (ELIZA, PARRY)
  - Monkeys *may* type Shakespeare's complete works — there's more to it than random search.
    I.e., if you pass the test without *any* identifiable brains inside, something's fishy. (Cf. the chess "proof" of AI)

The Turing Test

Unfortunately, the Turing Test is neither necessary nor sufficient as a test for intelligence:

- **Not necessary.** Presumably, Andromedans can be intelligent while incapable of mimicking women.
  - NB: Turing never suggested necessity!
- **Not sufficient.**
  - Imbecilic programs regularly come close to passing for imbecilic humans (ELIZA, PARRY)
  - Monkeys *may* type Shakespeare's complete works — there's more to it than random search.
    I.e., if you pass the test without *any* identifiable brains inside, something's fishy. (Cf. the chess "proof" of AI)

The Total Turing Test

- **Total Turing Test** (Harnad, 1989): drop the screen. Tests ability of robot to interact with the world (embodiment).
- **Even More Total Turing Test:** Require isomorphic internal information processing.
- **Totally Complete Turing Test:** Require isomorphic internal processing of all types, down to subatomic level.

Which test(s) is/are sufficient for intelligence?

REGARDLESS *if we get a program that can regularly fool many humans, we surely will have something impressive. . . yes?*

The Total Turing Test

- **Total Turing Test** (Harnad, 1989): drop the screen. Tests ability of robot to interact with the world (embodiment).
- **Even More Total Turing Test:** Require isomorphic internal information processing.
- **Totally Complete Turing Test:** Require isomorphic internal processing of all types, down to subatomic level.

Which test(s) is/are sufficient for intelligence?

REGARDLESS *if we get a program that can regularly fool many humans, we surely will have something impressive. . . yes?*

The Total Turing Test

- **Total Turing Test** (Harnad, 1989): drop the screen. Tests ability of robot to interact with the world (embodiment).
- **Even More Total Turing Test:** Require isomorphic internal information processing.
- **Totally Complete Turing Test:** Require isomorphic internal processing of all types, down to subatomic level.

Which test(s) is/are sufficient for intelligence?

REGARDLESS *if we get a program that can regularly fool many humans, we surely will have something impressive. . . yes?*

The Total Turing Test

- **Total Turing Test** (Harnad, 1989): drop the screen. Tests ability of robot to interact with the world (embodiment).
- **Even More Total Turing Test:** Require isomorphic internal information processing.
- **Totally Complete Turing Test:** Require isomorphic internal processing of all types, down to subatomic level.

Which test(s) is/are sufficient for intelligence?

REGARDLESS *if we get a program that can regularly fool many humans, we surely will have something impressive. . . yes?*

The Total Turing Test

- **Total Turing Test** (Harnad, 1989): drop the screen. Tests ability of robot to interact with the world (embodiment).
- **Even More Total Turing Test:** Require isomorphic internal information processing.
- **Totally Complete Turing Test:** Require isomorphic internal processing of all types, down to subatomic level.

Which test(s) is/are sufficient for intelligence?

REGARDLESS *if we get a program that can regularly fool many humans, we surely will have something impressive. . . yes?*

The Total Turing Test

- **Total Turing Test** (Harnad, 1989): drop the screen. Tests ability of robot to interact with the world (embodiment).
- **Even More Total Turing Test:** Require isomorphic internal information processing.
- **Totally Complete Turing Test:** Require isomorphic internal processing of all types, down to subatomic level.

Which test(s) is/are sufficient for intelligence?

REGARDLESS *if we get a program that can regularly fool many humans, we surely will have something impressive. . . yes?*

Can Machines Think? No Way

Hubert Dreyfus *What Computes Can't Do* (1965, 1993):

- Human intellect is an inarticulable *skill*; computer "intellect" involves rule-following.
- Human thinking has a context; computer rules are context-free.
- Human thinking is embodied; the essence of computation is universality — i.e., disembodiment.

Can Machines Think? No Way

Hubert Dreyfus *What Computes Can't Do* (1965, 1993):

- Human intellect is an inarticulable *skill*; computer "intellect" involves rule-following.
- Human thinking has a context; computer rules are context-free.
- Human thinking is embodied; the essence of computation is universality — i.e., disembodiment.

Can Machines Think? No Way

Hubert Dreyfus *What Computes Can't Do* (1965, 1993):

- Human intellect is an inarticulable *skill*; computer "intellect" involves rule-following.
- Human thinking has a context; computer rules are context-free.
- Human thinking is embodied; the essence of computation is universality — i.e., disembodiment.

Can Machines Think? No Way

Hubert Dreyfus *What Computes Can't Do* (1965, 1993):

- Human intellect is an inarticulable *skill*; computer "intellect" involves rule-following.
- Human thinking has a context; computer rules are context-free.
- Human thinking is embodied; the essence of computation is universality — i.e., disembodiment.

Rejoinders:

- Rule-following need only be at the implementation level.
- Neither rules nor emergent behavior need be context-free.
- Robots are not without bodies.

Can Machines Think? No Way

Hubert Dreyfus *What Computes Can't Do* (1965, 1993):

- Human intellect is an inarticulable *skill*; computer "intellect" involves rule-following.
- Human thinking has a context; computer rules are context-free.
- Human thinking is embodied; the essence of computation is universality — i.e., disembodiment.

Rejoinders:

- Rule-following need only be at the implementation level.
- Neither rules nor emergent behavior need be context-free.
- Robots are not without bodies.

Can Machines Think? No Way

Hubert Dreyfus *What Computes Can't Do* (1965, 1993):

- Human intellect is an inarticulable *skill*; computer "intellect" involves rule-following.
- Human thinking has a context; computer rules are context-free.
- Human thinking is embodied; the essence of computation is universality — i.e., disembodiment.

Rejoinders:

- Rule-following need only be at the implementation level.
- Neither rules nor emergent behavior need be context-free.
- Robots are not without bodies.

Can Machines Think? No Way

John Searle's Chinese Room Argument (1980):

- Suppose there is a computer program for understanding (written) Chinese.
- Then: Put the instructions in a book inside a room that has only an input and an output slot.
- Put John Searle inside the room.
- Whenever a squiggle comes in, Searle looks in the book what to do; eventually out comes a squoggle.

Can Machines Think? No Way

John Searle's Chinese Room Argument (1980):

- Suppose there is a computer program for understanding (written) Chinese.
- Then: Put the instructions in a book inside a room that has only an input and an output slot.
- Put John Searle inside the room.
- Whenever a squiggle comes in, Searle looks in the book what to do; eventually out comes a squoggle.

Can Machines Think? No Way

John Searle's Chinese Room Argument (1980):

- Suppose there is a computer program for understanding (written) Chinese.
- Then: Put the instructions in a book inside a room that has only an input and an output slot.
- Put John Searle inside the room.
- Whenever a squiggle comes in, Searle looks in the book what to do; eventually out comes a squoggle.

Can Machines Think? No Way

John Searle's Chinese Room Argument (1980):

- Suppose there is a computer program for understanding (written) Chinese.
- Then: Put the instructions in a book inside a room that has only an input and an output slot.
- Put John Searle inside the room.
- Whenever a squiggle comes in, Searle looks in the book what to do; eventually out comes a squoggle.

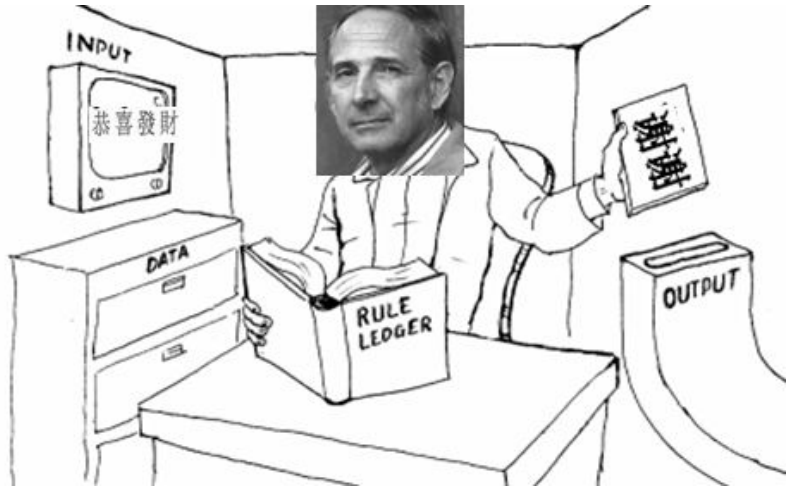
Can Machines Think? No Way

John Searle's Chinese Room Argument (1980):

- Suppose there is a computer program for understanding (written) Chinese.
- Then: Put the instructions in a book inside a room that has only an input and an output slot.
- Put John Searle inside the room.
- Whenever a squiggle comes in, Searle looks in the book what to do; eventually out comes a squoggle.

Can Machines Think? No Way

John Searle's Chinese Room Argument (1980):

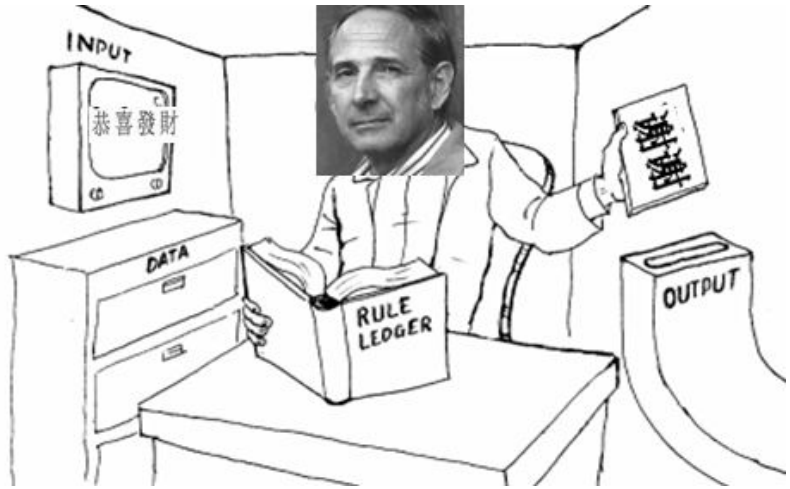


This room (over time!) has exactly the same (written) language behavior of some native Chinese speaker. . . But . . . but . . .

There's no UNDERSTANDING *going on in there!*

Can Machines Think? No Way

John Searle's Chinese Room Argument (1980):



This room (over time!) has exactly the same (written) language behavior of some native Chinese speaker. . . But . . . but . . .

There's no UNDERSTANDING *going on in there!*

Can Machines Think? Way

Responses to Searle:

- *The System Response:* There's no understanding *inside* the room, because it's the room *as a whole* that understands.
- *The Embodiment Response:* Understanding requires causal embedding; teletype communication with the world isn't enough. ("Symbol-grounding" is required.) What would the Searle-Room reply if asked: "Do you like my new shirt?"
- *The Consciousness Response:* Searle's intuition (no understanding) is based on a deeper and different intuition — there's no consciousness going on in there (distinct from Searle's)! But this replies to a mystery with an enigma. . .

Can Machines Think? Way

Responses to Searle:

- *The System Response:* There's no understanding *inside* the room, because it's the room *as a whole* that understands.
- *The Embodiment Response:* Understanding requires causal embedding; teletype communication with the world isn't enough. ("Symbol-grounding" is required.) What would the Searle-Room reply if asked: "Do you like my new shirt?"
- *The Consciousness Response:* Searle's intuition (no understanding) is based on a deeper and different intuition — there's no consciousness going on in there (distinct from Searle's)! But this replies to a mystery with an enigma. . .

Can Machines Think? Way

Responses to Searle:

- *The System Response:* There's no understanding *inside* the room, because it's the room *as a whole* that understands.
- *The Embodiment Response:* Understanding requires causal embedding; teletype communication with the world isn't enough. ("Symbol-grounding" is required.) What would the Searle-Room reply if asked: "Do you like my new shirt?"
- *The Consciousness Response:* Searle's intuition (no understanding) is based on a deeper and different intuition — there's no consciousness going on in there (distinct from Searle's)! But this replies to a mystery with an enigma. . .

References

Drefyus, H. (1965) Alchemy and Artificial Intelligence. Rand Corp.

Drefyus, H. (1993) What Computers Still Can't Do: A Critique of Artificial Reason, 3rd ed. MIT Press.

Harnad, S. (1990) The Symbol Grounding Problem. Physica D 42: 335-346.

Penrose, R. (1989) The Emperor's New Mind. Oxford.

Searle, J. R. (1980) Minds, brains and programs. Behavioral and Brain Sciences 3: 417-457.

Turing, A (1950) Computing machinery and intelligence. *Mind, 59,* 433-460. Reprinted many times (e.g., M. Boden (ed) *Philosophy of AI*, Oxford, 1990).

Weizenbaum, J. (1976) Computer Power and Human Reason. W.H. Freeman.