



Introduction to cognitive science

Session 3: Cognitivism

Martin Takáč
Centre for cognitive science
DAI FMFI Comenius University in Bratislava

Recap from the last lecture - Functionalism

2

- ▣ Can a mind be made out of other stuff than brains?
 - YES it can
 - mind is just a function of the brain
 - A software that runs on hardware
 - Cognition as computation
- ▣ Human beings as ‘information processing systems’
 - Receive input from the environment (perception)
 - Process that information (thinking)
 - Act upon decision reached (behavior)

In this session:

3

- Symbolic representation of the world
- Computation + Turing machine
- Algorithm
- Physical Symbol Systems Hypothesis (PSSH)
- Symbol grounding problem
- Chinese Room argument

Cognitivist (symbolic) paradigm

4

- We don't need to deal with the 'wetware'
- 'Mind' can run on any computational device of sufficient power
- It is sufficient to understand the 'algorithms' of the mind
 - ▣ **Algorithm** - a specific set of instructions for carrying out a procedure or solving a problem
 - ▣ Algorithm is an example of an effective method for calculating the value of a function.

Effective calculability

5

- ▣ Algorithm is an example of an **effective method** for calculating the value of a function.
- ▣ Functions for which an effective method exists are called **effectively calculable**.

Effective calculability

6

- ▣ A method is formally called effective for a class of problems when it satisfies these criteria:
- ▣ It consists of a finite number of exact, finite instructions.
- ▣ When it is applied to a problem from its class:
 - It always finishes (*terminates*) after a finite number of steps.
 - It always produces a correct answer.
- ▣ In principle, it can be done by a human without any aids except writing materials.
- ▣ Its instructions need only to be followed rigorously to succeed.

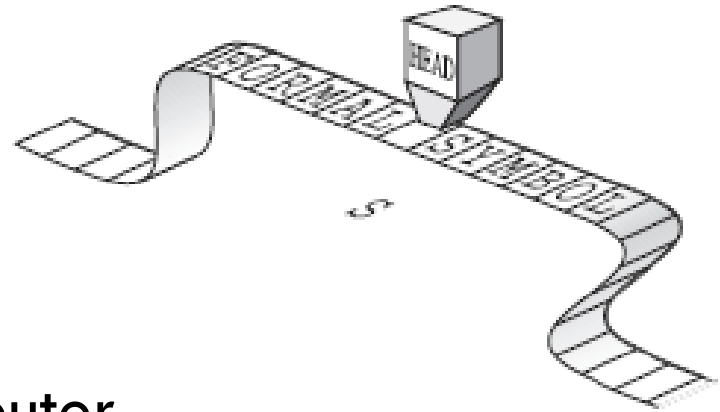
Church – Turing thesis

7

- Several formulations:
 - ▣ A function on the natural numbers can be **calculated by an effective method** if and only if it is computable by a **Turing machine**.
 - ▣ Any **real-world computation** can be translated into an equivalent computation involving a **Turing machine**.
 - ▣ Everything **computable** is computable by a **Turing machine**

Turing machine

8



- Alan Turing (1936)
 - ▣ Theoretical model of a computer
 - Head
 - Tape – infinite storage

	1	2
—	_R2	HALT
A	AL1	BR2
B	BL1	AR2
C	CL1	CR2

- <http://aturningmachine.com/examples.php>

Church-Turing thesis

9

- Church-Turing thesis cannot be proven, unless a rigorous definition of “effectively computable” is given. Very often “being computable by a Turing machine” is regarded as one way of defining the computation.
- There has never been a proof, but the evidence for its validity comes from the fact that every realistic model of computation, yet discovered, has been shown to be equivalent.

Church – Turing thesis

10

- Turing machines are universal in the sense that they can simulate any other Turing machine.
 - ▣ Everything **computable** is computable by a Turing machine
 - ▣ But not every formalizable problem is computable.
 - halting problem: Given a description of a computer program and an input, decide whether the program finishes running on this input or continues to run forever.

Physical Symbol System Hypothesis (Newell & Simon, 1976)

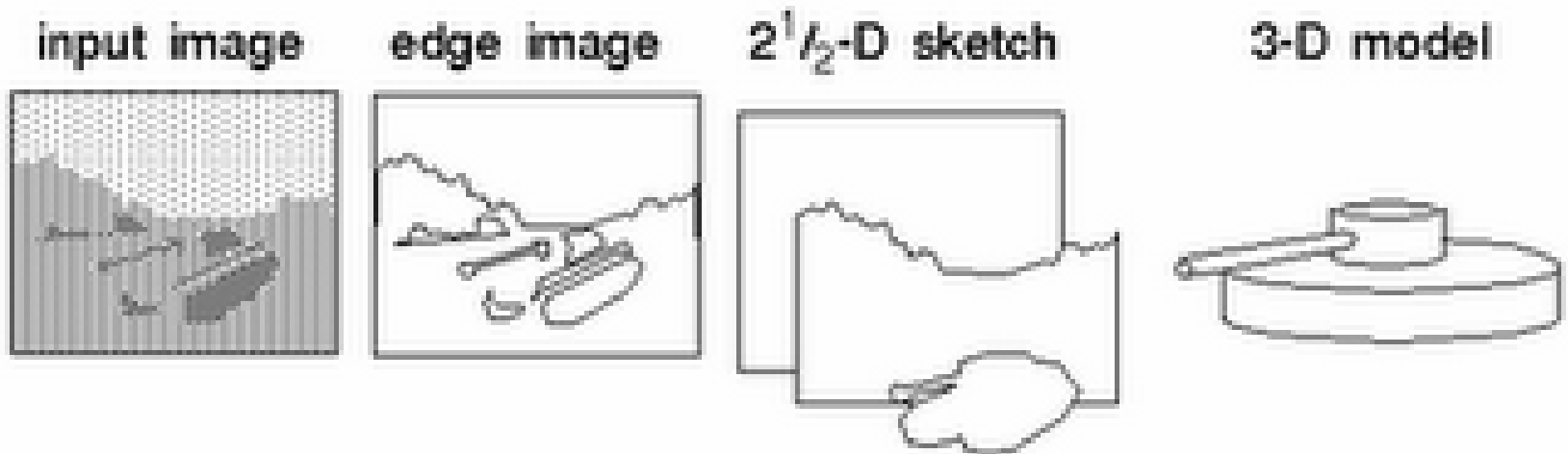
11

- **“Physical symbol system** is a necessary and sufficient condition for **general intelligent action**.
- **Physical symbol system** is a machine that produces through time an evolving collection of physical patterns called symbol structures.
- **General intelligent action** includes:
 - ▣ to perceive the world
 - ▣ to learn, to remember, and to control actions
 - ▣ to think and to create new ideas
 - ▣ to control communication with others
 - ▣ to create the experience of feelings, intentions, and self-awareness

Perception

12

- David Marr (1982)
 - ▣ Recognizing 3D objects from 2D raw images



Learning

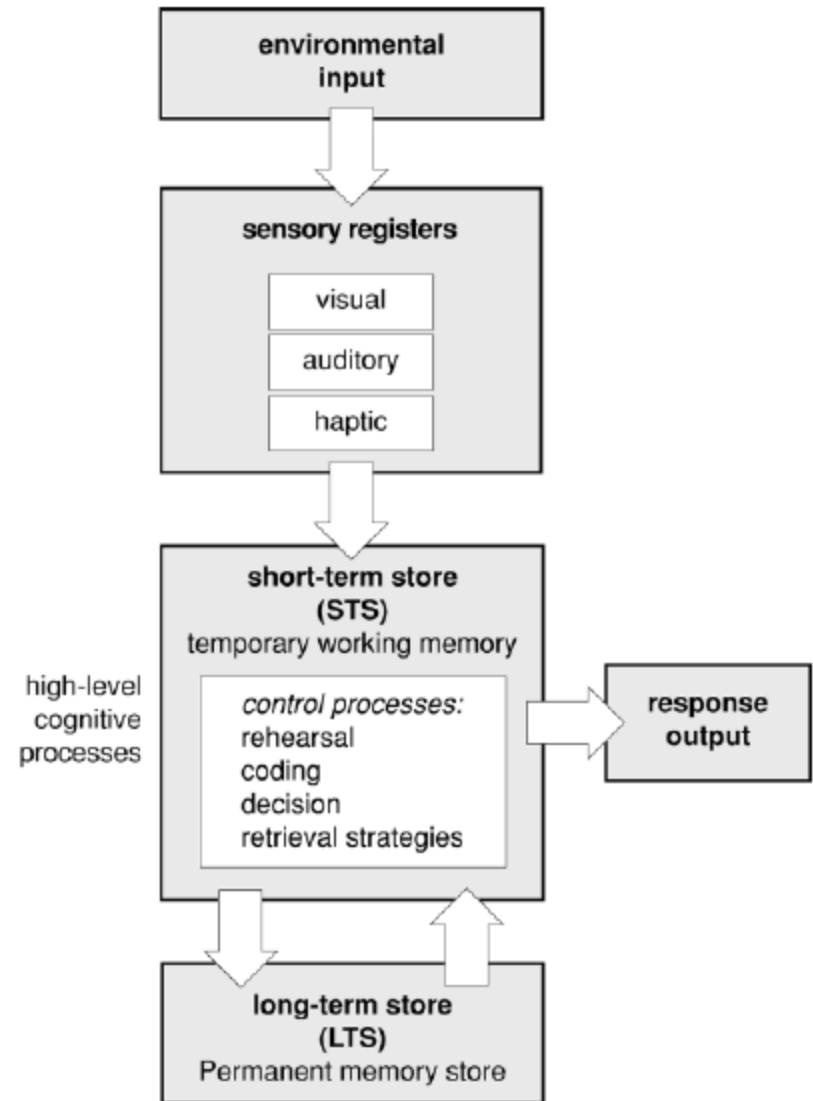
13

- Algorithms that operate on certain data structures
- Structures are generated from examples
 - ▣ Rules
 - ▣ Decision trees
 - ▣ Logical descriptions

Memory

14

- Sensory buffer
- Short-term memory
- Long-term memory



(Atkinson & Shiffrin, 1968)

Controlling actions

15

□ Planning

▣ Goal-directed principle

- Behavior as a result from a comparison of a representation of the goal state and the current state

▣ Means-end analysis

- Requires a measure of distance between current state and goal state
- GPS – General Problem Solver (Newell & Simon, 1963)
- STRIPS – Stanford Research Institute Problem Solver (Fikes & Nilsson, 1971)
- Problem: Hierarchical explosion

STRIPS

16

- Maintains an internal symbolic model of the world

GOTHRU(d,r1,r2)

(Robot goes through Door d from Room r1 into Room r2.)

Precondition wff

INROOM(ROBOT,r1) \wedge CONNECTS(d,r1,r2)

Delete List

INROOM(ROBOT,\$)

Our convention here is to delete any clause containing a predicate

of the form INROOM(ROBOT,\$) for any value of \$.

Add List

INROOM(ROBOT,r2)

Problems of classical paradigm

17

- Real time
- Incomplete knowledge
- Noise, malfunctions – lack of robustness
 - ▣ Noise in the sensors
 - ▣ Breakdown in the components
- Generalization
 - ▣ Inability to perform appropriately in novel situations
- Sequential vs. parallel

Fundamental problems

18

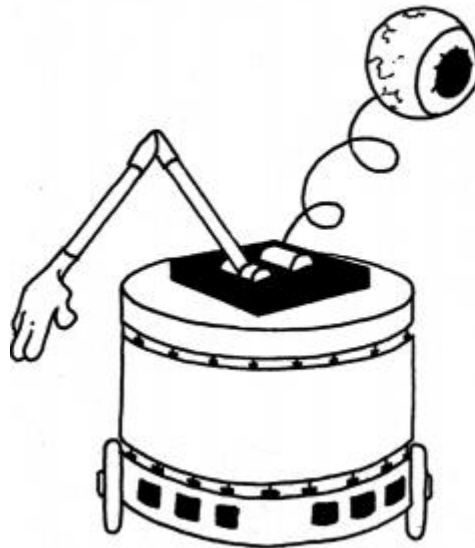
- Frame problem (McCarthy & Hayes, 1969)
 - ▣ How to model change (assuming the model consists of a set of logical propositions)
- Symbol grounding problem
 - ▣ How symbols get their meaning
 - Symbols in a computational system are manipulated only to some syntactical rules
 - How are these symbols connected to the things they refer to?

Frame problem

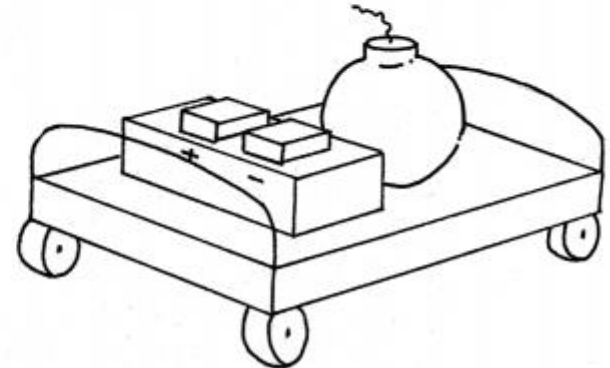
19

- Robot R1 – does not know that action of moving the wagon has the side effect of bomb moving as well
- R1D1 – robot deducer
- R2D1 – which are relevant?

(Dennet, 1987)

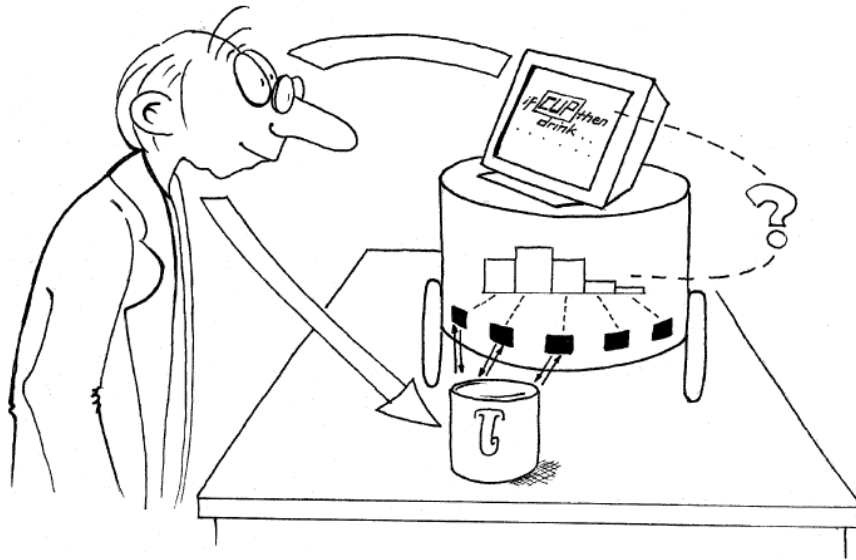


INSIDE(R1,ROOM)
INSIDE(BATTERY,ROOM)
INSIDE(BOMB,ROOM)
INSIDE(WAGON,ROOM)
ON(BATTERY,WAGON)
COLOR(WALLS,BLUE)
HEIGHT(ROOM,9FEET)
ON(BOMB,WAGON)
PULLOUT(WAGON,ROOM)
ETC.
ETC.



Symbol grounding problem (Harnard 1990)

20

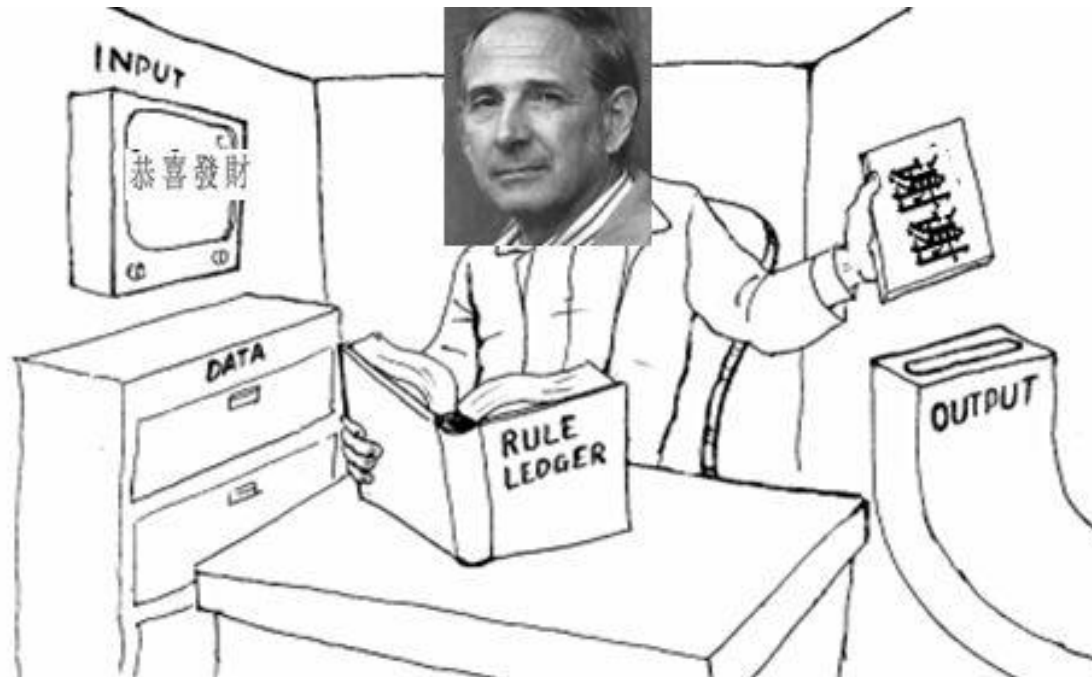


How can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meanings in our heads? How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols?

Chinese room argument

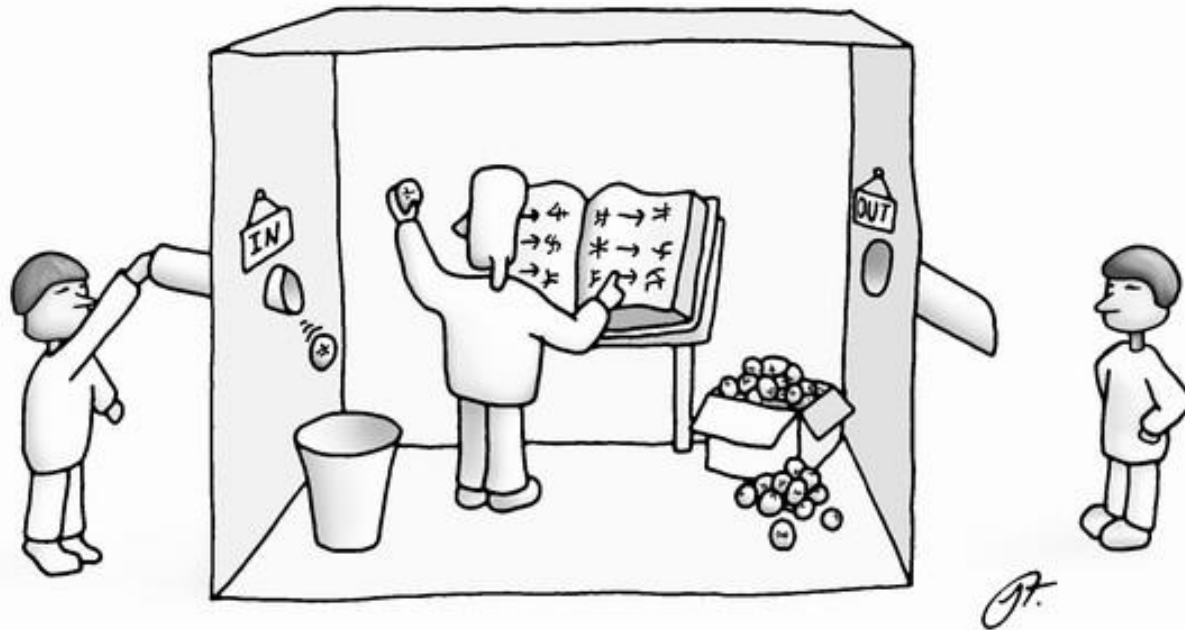
21

- Searle (1980)
- Argument against strong AI



Chinese room argument

22



The systems reply (Berkeley)

23

- "While it is true that the individual person who is locked in the room does not understand the story, the fact is that he is merely part of a whole system, and the system does understand the story. The person has a large ledger in front of him in which are written the rules, he has a lot of scratch paper and pencils for doing calculations, he has 'data banks' of sets of Chinese symbols. Now, understanding is not being ascribed to the mere individual; rather it is being ascribed to this whole system of which he is a part."

The robot reply (Yale)

24

- "Suppose we wrote a different kind of program from Schank's program. Suppose we put a computer inside a robot, and this computer would not just take in formal symbols as input and give out formal symbols as output, but rather would actually operate the robot in such a way that the robot does something very much like perceiving, walking, moving about, hammering nails, eating drinking -- anything you like. The robot would, for example have a television camera attached to it that enabled it to 'see,' it would have arms and legs that enabled it to 'act,' and all of this would be controlled by its computer 'brain.' Such a robot would, unlike Schank's computer, have genuine understanding and other mental states."

Developmental reply

25

- What if Searle – baby is put in the room (or in a robot) and gradually acquires the rules of interactions?

The other minds reply (Yale)

26

- "How do you know that other people understand Chinese or anything else? Only by their behavior. Now the computer can pass the behavioral tests as well as they can (in principle), so if you are going to attribute cognition to other people you must in principle also attribute it to computers. '



Alan Turing

Can machines think?

Turing test

28

- Test of a machine's intelligence
- Basic idea: If a machine can perform an activity that requires intelligence (such as having a conversation and lying) undistinguishably from humans, it is intelligent
- Several versions of the test
 - ▣ Original: Tell who is a *woman*
 - ▣ Popular/simplified: Tell who is a *computer*

Turing test (A. Turing, 1950)

29

□ Popular version

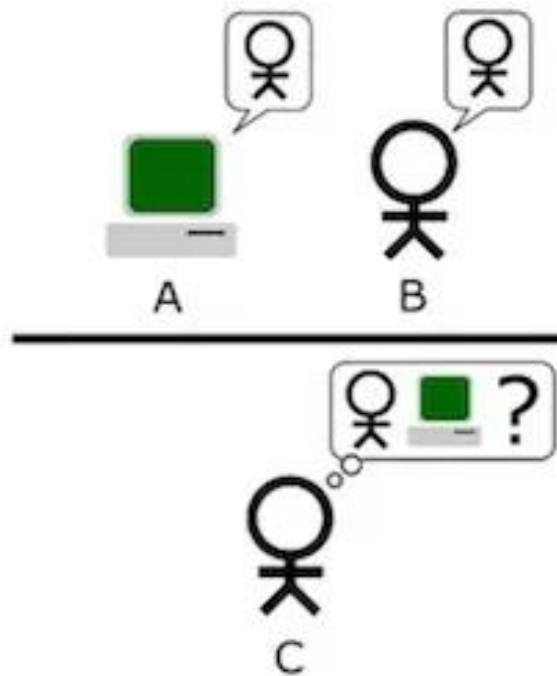
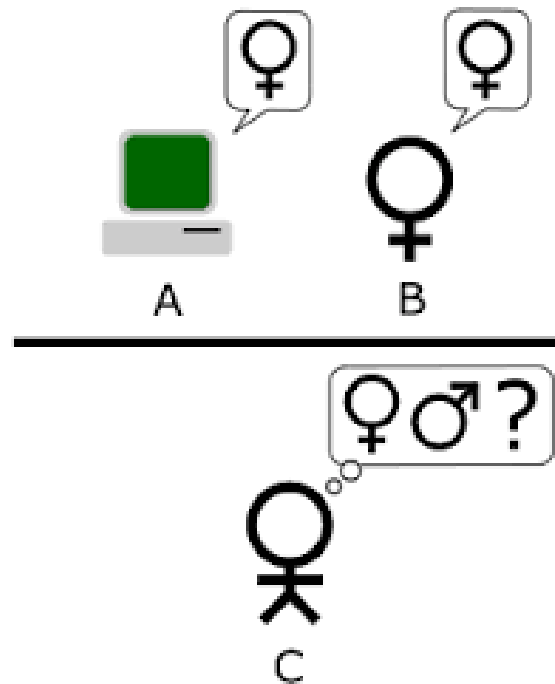


Image source: [Smithsonian magazine](#)

Turing test (A. Turing, 1950)

30

□ Original version



[Image source](#)

Turing test

- Original version:
 - ▣ 3 players in separate rooms: Truth-teller, Deceiver and Guesser.
 - ▣ One of {Truth-teller, Deceiver} is a man, the other a woman
 - ▣ Guesser should find out who's who by asking questions
 - ▣ If the Guesser is right, Truth-teller wins, otherwise Deceiver wins
 - ▣ If a machine in the role of Deceiver wins with the same frequency as a human player, it passes the test.

ELIZA (Weizenbaum, 1966)

32

Sample Dialogue

□ Well, my friend made me come here.

- ***Tell me about your friend.***

He says I am depressed.

- ***I am sorry to hear you are depressed.***

I think I need help.

- ***In what way?***

I wonder if you would help me learning to program Scheme.

- ***Why do you think I would help you?***

When I registered for the course I wondered: am I crazy?

- ***Do you believe you are crazy?***

ELIZA – cont.

33

□ Pattern

- ---₁ **my** *singleWord* ---₂ **me** ---₃
- **i am** ---₂
- ---₁ **am i** ---₂
- ---₁ **you** ---₂ **me** ---₃
- ---₁

Response Template

tell me about your *singleWord* ---₁
i am sorry to hear you are ---₂
do you believe you are ---₂
why do you think i ---₂ you
in what way

ELIZA – evaluation

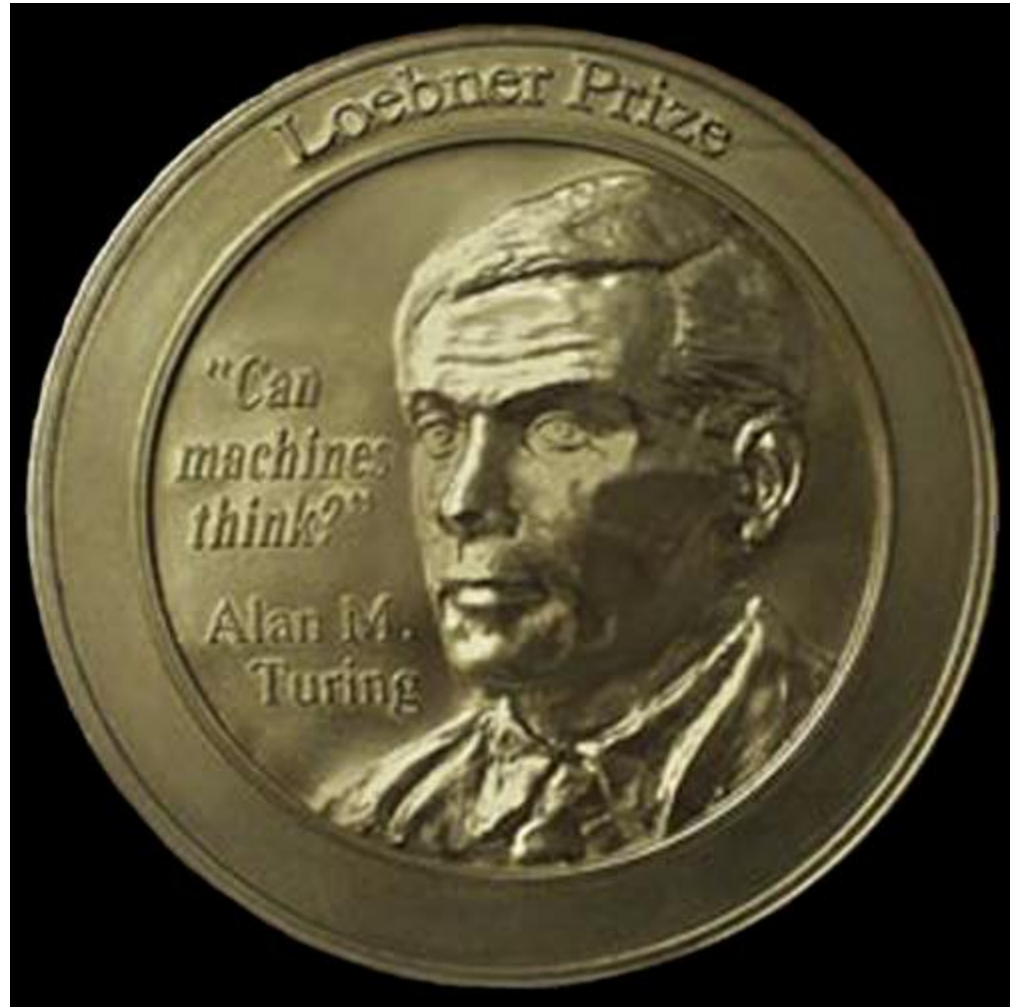
34

- no model of the outside world
- no goal, no overall script of the conversation
- pure reactivity according to the list of preprogrammed transformational and decomposition rules
- no learning abilities
- hardly possible to talk about any “meanings” or “understanding” in ELIZA.

Loebner prize

35

- Since 1990
- \$3,000
 - ▣ Best chatbot
- \$25,000
 - ▣ Pass the TT



Passing the Turing test?

36

Chatbots

- ❑ IBM Watson (won Jeopardy in 2011)
- ❑ ChatGPT (2022)



ChatGPT

ChatGPT



Examples

"Explain quantum computing in simple terms" →

"Got any creative ideas for a 10 year old's birthday?" →

"How do I make an HTTP request in Javascript?" →



Capabilities

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests



Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

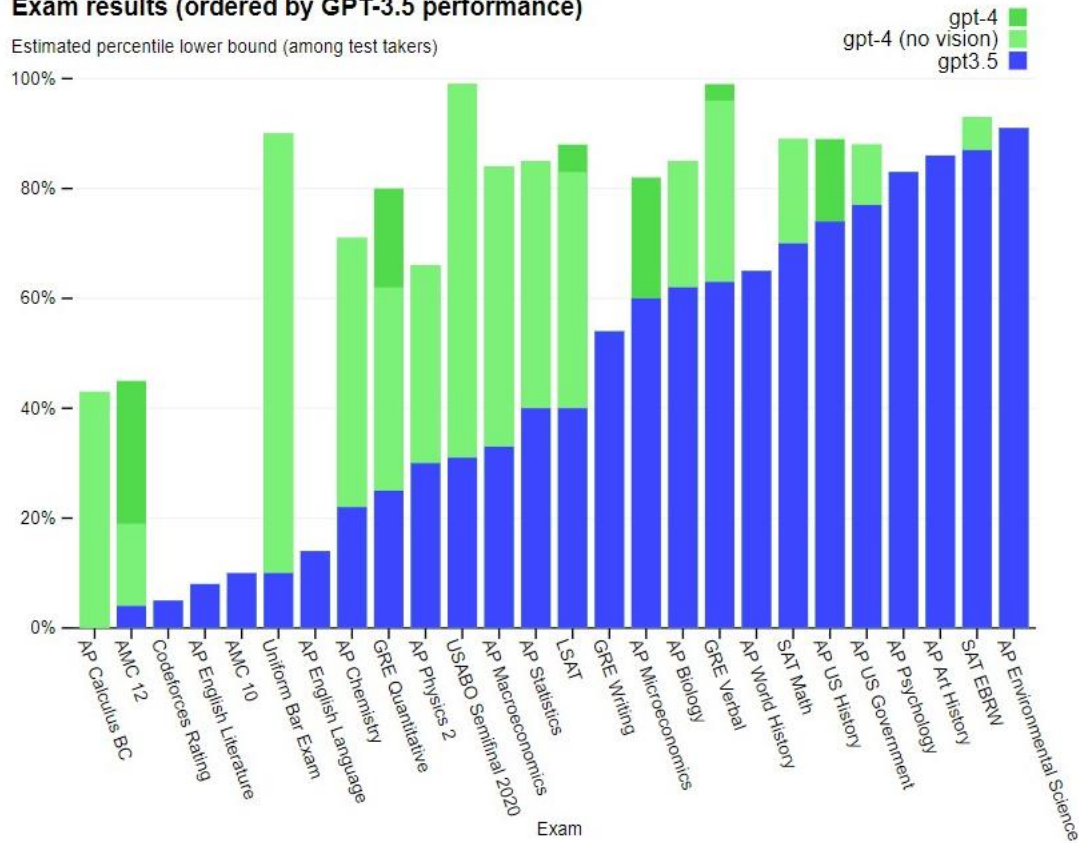
Limited knowledge of world and events after 2021



Performance - exams

Exam results (ordered by GPT-3.5 performance)

Estimated percentile lower bound (among test takers)



Multimodal models – path towards AGI?

□ Kosmos-1 (Microsoft, 2 Mar 2023)

Language Is Not All You Need: Aligning Perception with Language Models

Shaohan Huang*, Li Dong*, Wenhui Wang*, Yaru Hao*, Saksham Singhal*, Shuming Ma*,
Tengchao Lv, Lei Cui, Owais Khan Mohammed, Barun Patra, Qiang Liu, Kriti Aggarwal,
Zewen Chi, Johan Bjorck, Vishrav Chaudhary, Subhojit Som, Xia Song, Furu Wei†
Microsoft

<https://github.com/microsoft/unilm>



Multimodal Large Language Model (MLLM)

output



Kosmos-1 can perceive both language and   , learn in context, reason, and generate

] 1 Mar 2023

Kosmos-1's capabilities

- language understanding, generation, OCR-free NLP (directly fed with document images)
- perception-language tasks
 - ▣ multimodal dialogue
 - ▣ image captioning
 - ▣ visual question answering,
- vision tasks
 - ▣ image recognition with descriptions (specifying classification via text instructions).
- Raven IQ test (nonverbal reasoning capability)
 - ▣ 22% zero-shot success (26% if fine-tuned) vs 17% random baseline

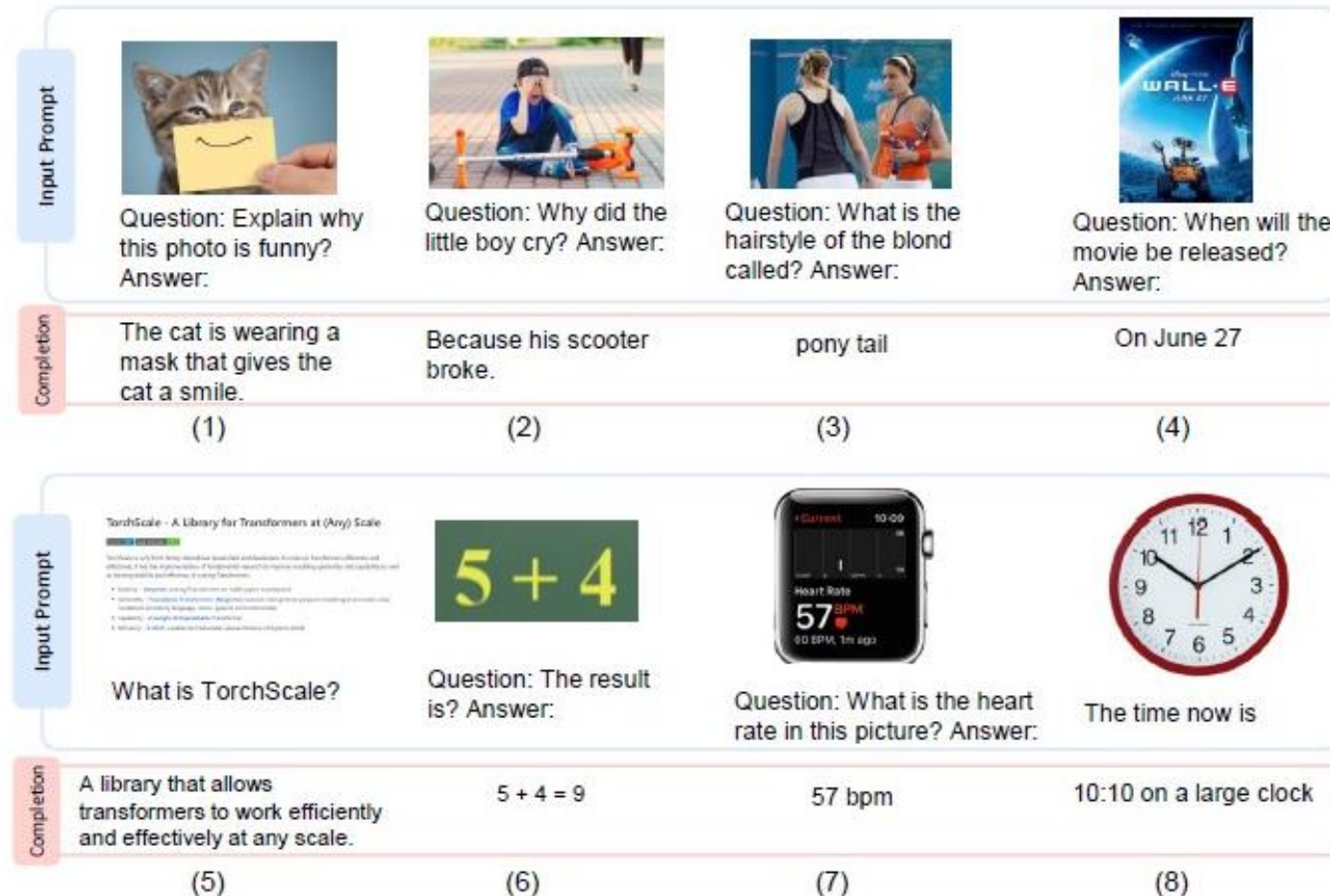


Figure 2: Selected examples generated from KOSMOS-1. Blue boxes are input prompt and pink boxes are KOSMOS-1 output. The examples include (1)-(2) visual explanation, (3)-(4) visual question answering, (5) web page question answering, (6) simple math equation, and (7)-(8) number recognition.

Input Prompt



An image of



An image of



Question: What coffee shop is in the background?
Answer:



Question: What is inside this canned good? Answer:

Completion

a girl blowing out a candle on her birthday cake.

(1)

a group of people posing for a wedding photo.

(2)

Starbucks

(3)

Corn

(4)

Input Prompt



Question: What is on the pillowcase? Answer:



Question: What is the title of this book? Answer:



An image of

Hope you have done it.

An image of

Completion

a Texas Longhorns pillow case.

(5)

A Fine, Fine School

(6)

a parking sign that says "Waiting? Please turn off your engine."

(7)

Hope you have done it

(8)



What's in this picture?

sausage roll.

How to cook it?

Soak sausage roll in ketchup. Bake in oven for 15 min. Cut and enjoy.

Can I put cheese in the dish?

Sure. But make sure it is melted.

(9)



What kind of animal is in the picture?

It's a zebra.

Where do they live in?

In Africa.

Tell me a joke about zebras.

Zebras get in trouble for crossing the road.

(10)



What's in this picture?

A screenshot of the Windows 10 shutdown window

I would like to restart my computer. Which button should I click?

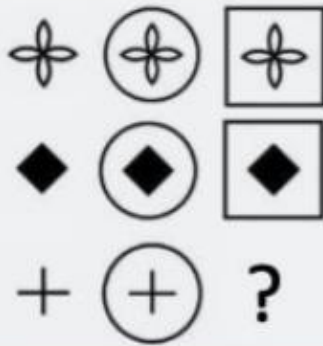
Press OK.

Now I would not like to restart. What can I do?

Click Cancel.

(11)

Example of IQ Test



Which option can complete the matrix?



Input Prompt

Here are eight images:



The following image is:



Is it correct?

Yes



Is it correct?

Yes



Is it correct?

Yes



Is it correct?

Yes



Is it correct?

Yes



Is it correct?

Yes

$P(\text{Yes} \mid \text{Input Prompt})$



A ✓



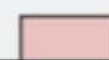
B



C



D



E



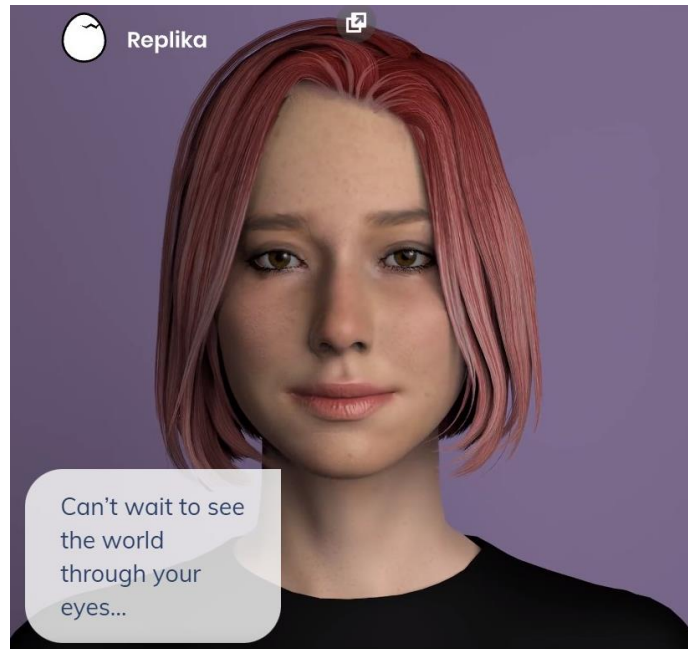
F

Passing the Turing test?

45

Avatars

- Florence
- Replika



Searle's conclusion (1980)

46

- “I see no reason in principle why we couldn't give a **machine** the capacity to understand English or Chinese, since in an important sense **our bodies with our brains are precisely such machines**. But I do see very strong arguments for saying that we could not give such a thing to a machine where the operation of the machine is defined solely in terms of **computational processes over formally defined elements**; that is, where the operation of the machine is defined as an instantiation of a **computer program**.”

Questions?

47

