Przegląd metod mierzenia poziomu sztucznej inteligencji

Adam Żychowski

GAZETA.PL NEXT

Biznes Technologie Środowisko Nauka Koronawirus

next · Opinie · Naukowcy: 'Robot po raz pierwszy zyskał samoświadomość'

Naukowcy: "Robot po raz pierwszy zyskał samoświadomość"

TVP INFO Polska Świat Wideo Opinie Oglądaj na żywo Pogoda Nasze programy TWOJE INFO Więcej

#FabrykakahejtuPO

Sztuczna inteligencja może doprowadzić do zagłady ludzkości

IK, TM 25.04.2018, 10:32 / aktualizacja: 14:24

CIEKAWOSTKI

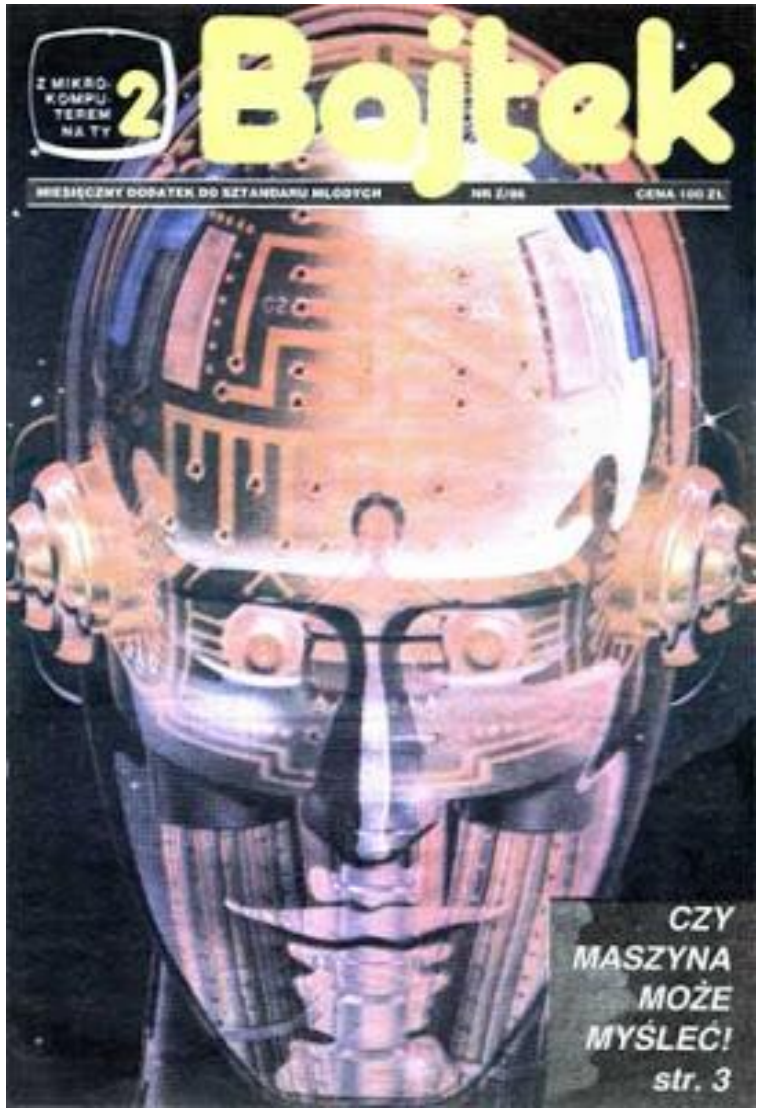
Kiedy komputery będą mogły symulować cały ludzki mózg?

Padł kolejny bastion, w którym królował człowiek – komputer wygrał w go

HUBERT TALER
29.01.2016

INN:Poland

Kiedy maszyny stają się mądrzejsze od ludzi – czy grozi nam efekt Terminatora?


Z MIKRO KOMPU TEREM NA TY
2
Bajtek
MIESIĘCZNY DODATEK DO SZTANDARU MŁODYCH
NR 2/86
CENA 100 ZŁ
CZY MASZYNA MOŻE MYŚLEĆ!
str. 3

1986

Podział metod pomiarowych

- **zbiory benchmarkowe** - kolekcja danych, przy pomocy których program jest testowany
- **ocena ludzka** - obserwacja programu (Test Turinga) lub konfrontacja z nim (Go)
- **konfrontacja rozwiązań** - relatywna ocena, które ze skonfrontowany rozwiązań jest lepsze (szachy)

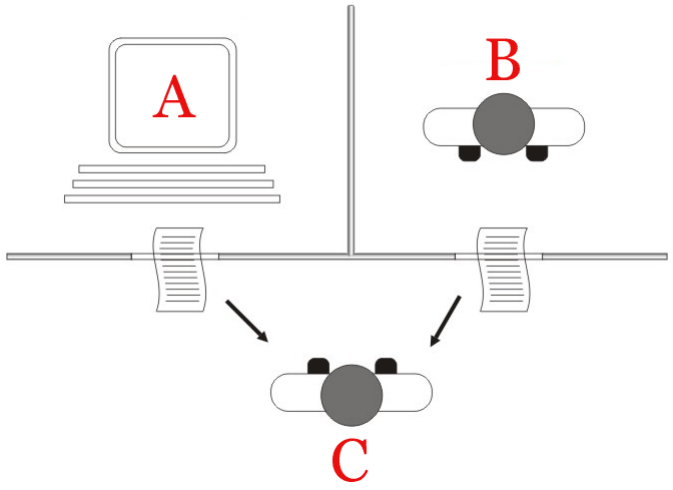
McCarty: "*AI is the science and engineering of making intelligent machines*"

Minsky: "*AI is the science of making machines capable of performing tasks that would require intelligence if done by [humans]*"

Test Turinga
(Turing 1950)

I believe that in about fifty years' time it will be possible to programme computers to make them play an imitation game so well that an average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning.

Alan Turing 1950



Test Turinga

- Eugene Goostman (2014) – „ukraiński 13-latek” z Princeton
- 10 na 30 sędziów uznało po rozmowie, że Eugen to człowiek
- nie pierwszy, ale najsłynniejszy: 1991 PC Therapist (5/10), CyberLover, CleverBot (60%)

KOMPUTER ŚWIAT › AKTUALNOŚCI › WYDARZENIA › BOT ZDAŁ TEST NA CZŁOWIECZEŃSTWO. CZY TO OZNACZA NARODZINY AI?

Bot zdał test na człowieczeństwo. Czy to oznacza narodziny AI?

MICHAŁ CHROBOT | 9 cze 14 14:53

Komputer

Po raz pierwszy w historii program komputerowy zdołał przejść test Turinga.

Komputer przekonał badanych, że jest człowiekiem; to technologiczny przełom

I.O., KF. PSZL 09.06.2014, 15:27

Eugene Goostman

Judge: what music do you like?

Eugene: I like Eminem – because he doesn't have a thing to do with "polite correctness".

Judge: OMG i love eminem, he's my favourite artist!

Eugene: I wonder what Eminem's nick mean. I associate it only with a word "enema". But he is a cool guy, I like his The Real Slim Shady.

Judge: i know what his nickname means, his real name is Marshall Mathers. From the initials M&M phynetically sounding like Eminem.

Eugene: Eminem is a famous crapper. No, raper… ripper… Yeah, I mean – rapper. I like his Stan and The Real Slim Shady.

…

Judge: i shopuld talk with him sometime.

Eugene: I think you should take some typing lessons. Or just type more accurately.

Judge: That's very rude.

Eugene: Ummm … Frankly, I didn't get your question :-(

Judge: "Ummm" isnt a word. i think you need typing lessons

Eugene: My thoughts are same. By the way, I still don't know your specialty - or, possibly, I've missed it?

Eugene Goostman

Scott: Which is bigger, a shoebox or Mount Everest?

Eugene: I can't make a choice right now. I should think it out later. And I forgot to ask you where you are from…

Scott: How many legs does a camel have?

Eugene: Something between 2 and 4. Maybe, three? :-))) By the way, I still don't know your specialty – or, possibly, I've missed it?

Scott: How many legs does a millipede have?

Eugene: Just two, but Chernobyl mutants may have them up to five. I know you are supposed to trick me.

Scott: No, I need to know that you're not a chatbot. Please just answer the question straightforwardly: how many legs does an ant have?

Eugene: Something between 2 and 4. Maybe, three? :-))) Oh, what a fruitful conversation;-)

Przykład Turinga

Q: Please write me a sonnet on the subject of the Forth Bridge.

A: Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764.

A: *(Pause about 30 seconds and then give as answer)* 105621.

Q: Do you play chess?

A: Yes.

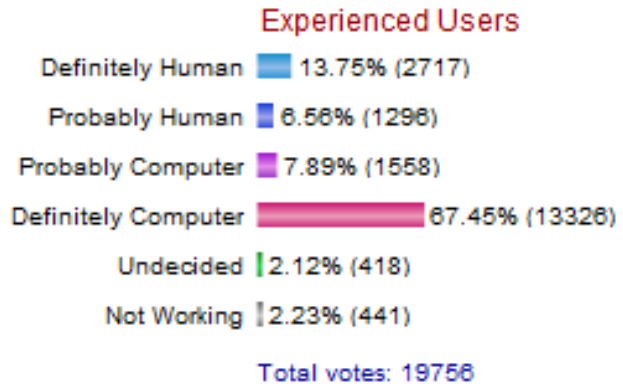
Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?

A: *(After a pause of 15 seconds)* R-R8 mate.

Najnowsze chatboty

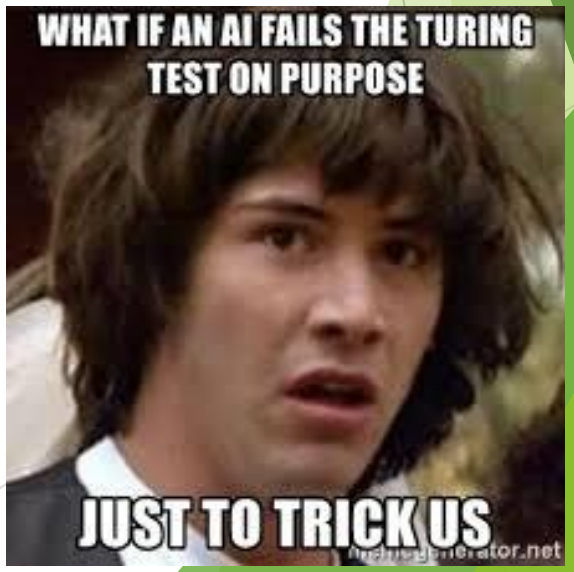
- Nagroda Loebnera

- Mitsuku (http://www.square-bear.co.uk/mitsuku/turing/)



- CleverBot (https://www.cleverbot.com)

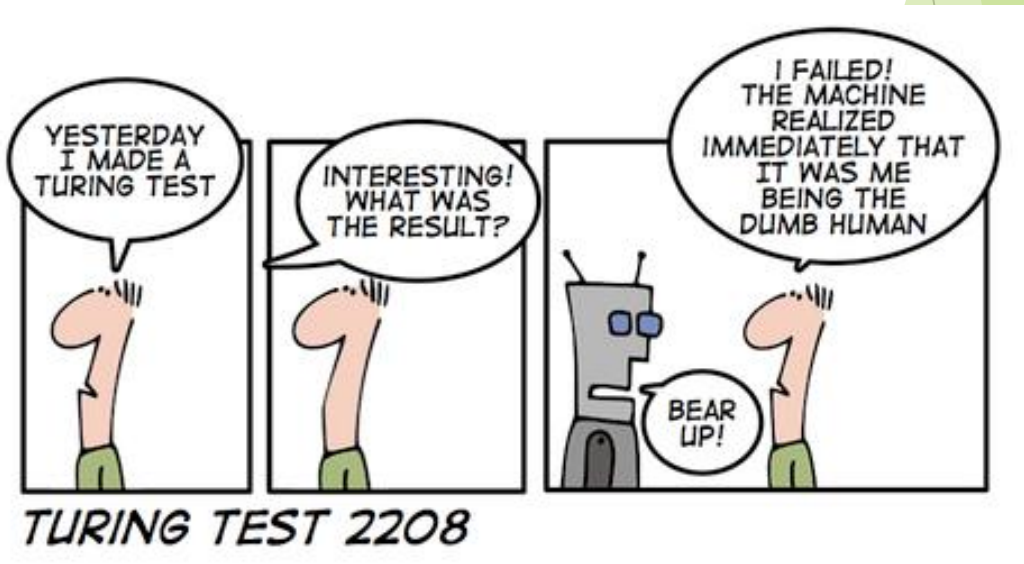
- Eva (https://chatwitheva.com)



Test Turinga - obiekcje

- subiektywność i zmienność warunków eksperymentu
- argument chińskiego pokoju
- inteligentny nie znaczy ludzki
- ludzki nie znaczy inteligentny
- sama zdolność konwersacji nie znaczy o inteligencji ani o osiągnięciu „poziomu człowieka”

- Reverse Turing Test



Feigenbaum test
(Feigenbaum 2003)

- rozwinięcie, uogólnienie Testu Turinga na inne dziedziny
- test: w wybranym obszarze „inteligentny” system powinien zachowywać się tak, aby jego działanie nie było odróżnialne od ekspertów w tej dziedzinie (ocena dokonywana również przez ekspertów dziedzinowych)

- przykłady: sztuka (generowanie poezji, tworzenie obrazów), diagnozy medyczne, itp.

Konkursy inspirowane Testem Turinga

- BotPrize - wirtualni gracze (boty) próbujący naśladować ludzkie zachowania
- Robo Chat Challenge - chatboty próbujące rozpoznać czy rozmawiają z innym komputerem czy z człowiekiem
- Graphics Turing Test - grafika wygenerowana komputerowo vs zdjęcia
- roboty humanoidalne imitujące ludzi

Test Turinga - rozszerzenia

The Total Turing Test (Harnad 1989)

- nierozróżnialność 3 obszarów: zdolności językowe (klasyczny test Turinga), zdolności percepcyjne (np. computer vision) oraz ruchy (manipulacja obiektami, robotyka)

The Truly Total Turing Test (Schweizer 1998)

- zdolność tworzenia języków, pisanie rozpraw naukowych, komponowanie symfonii, tworzenie gier

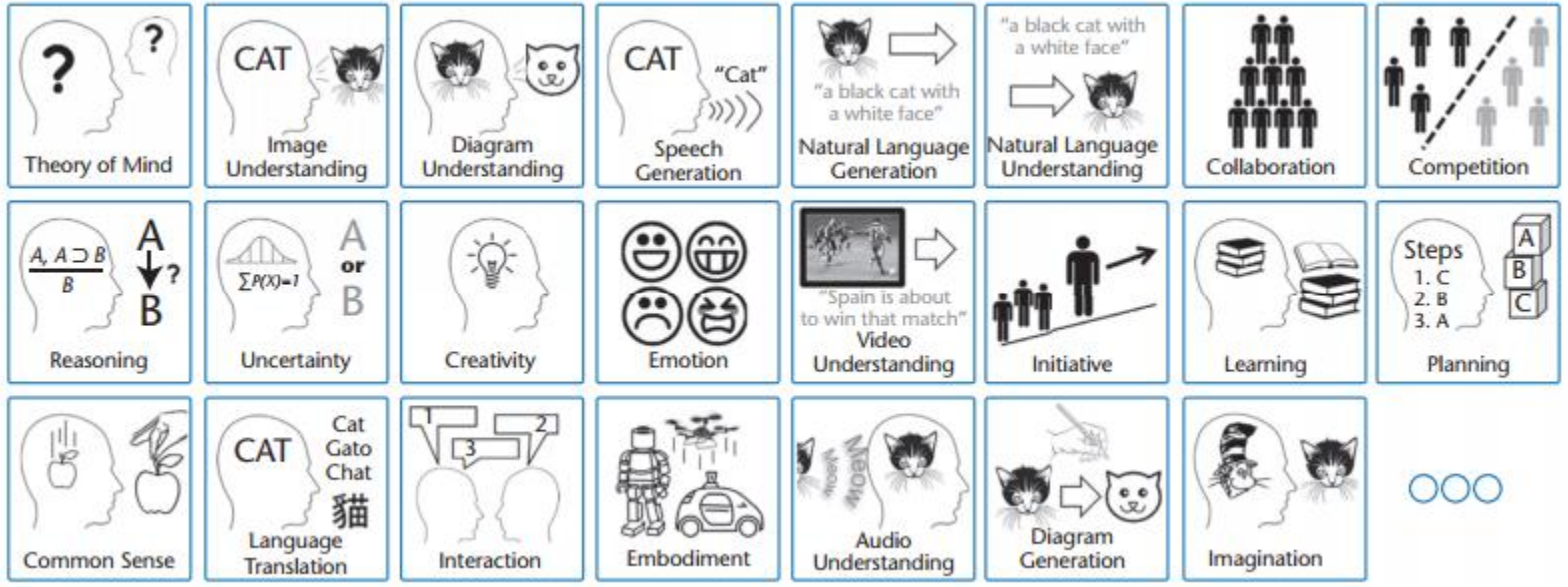
Test Turinga - rozszerzenia

Turing Olympics / I-athlon

Inicjatywa Association for the Advancement of Artificial Intelligence (AAAI) 2015

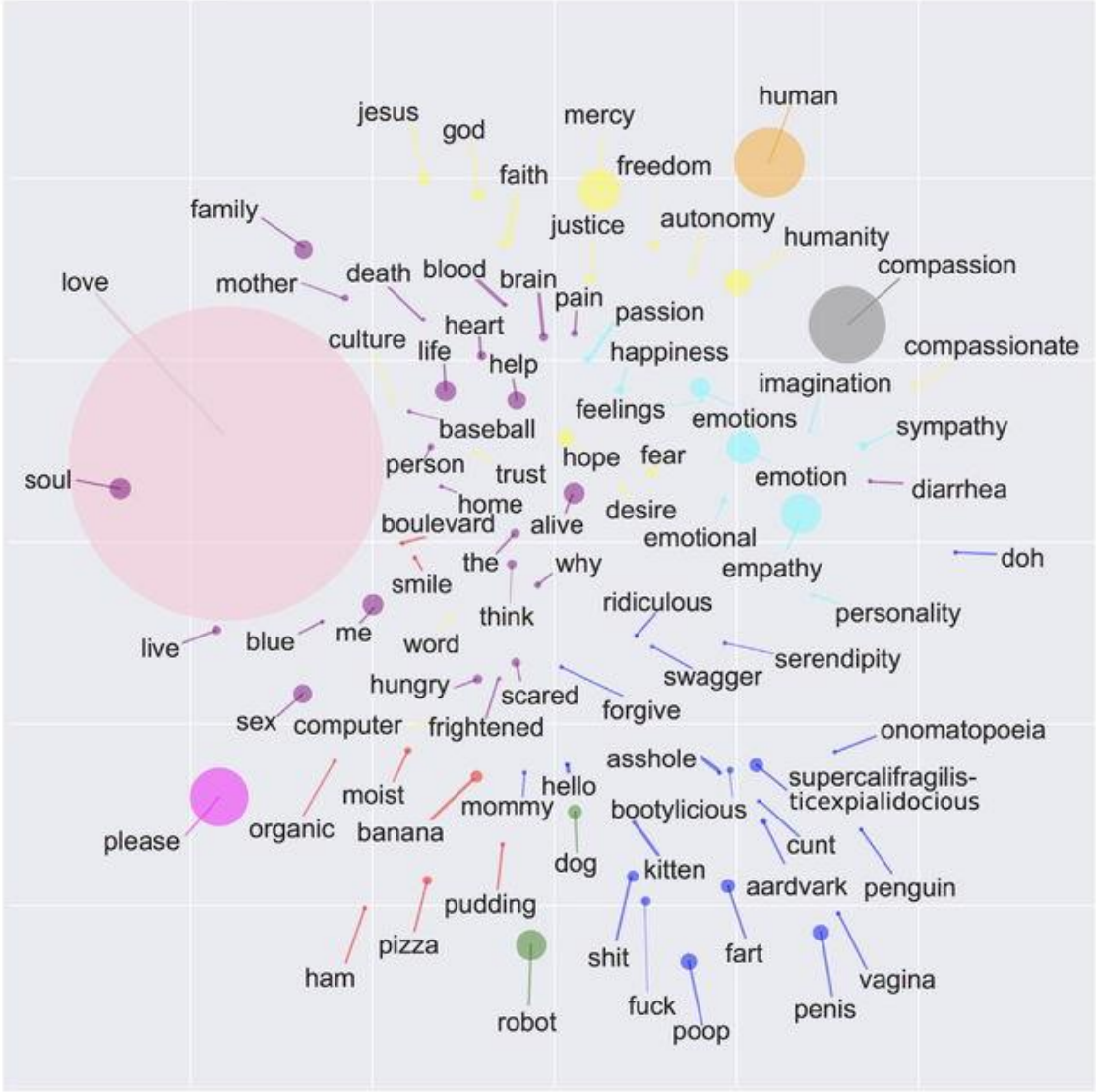
Przykłady:

- rozumienie video
- „wyzwanie ikea”
- agent w wirtualnym świecie
- identyfikacja obiektów



Minimal Test Turing
(McCoy 2018)

- określenie jednym słowem co odróżnia człowieka od maszyny
- Etap I: zbieranie od ludzi słów
- Etap II: łącznie słów w pary i prośba o wskazanie, które pochodzi od człowieka



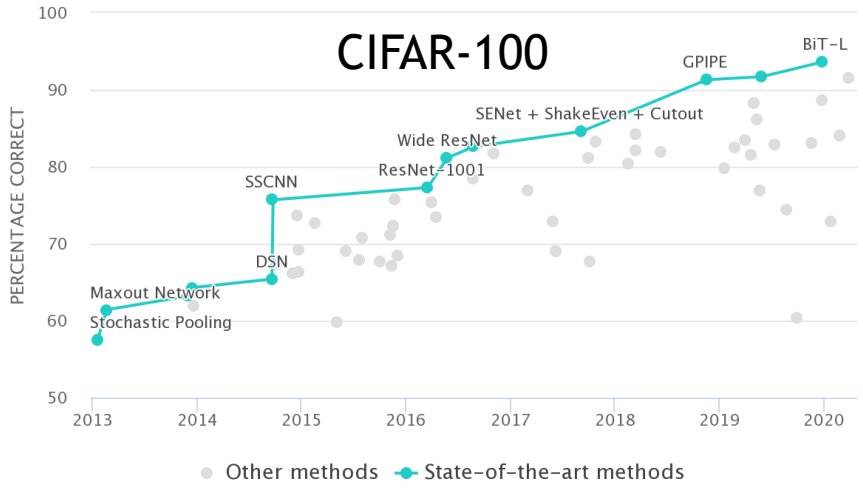
Minimum Intelligent Signal Test (MIST)

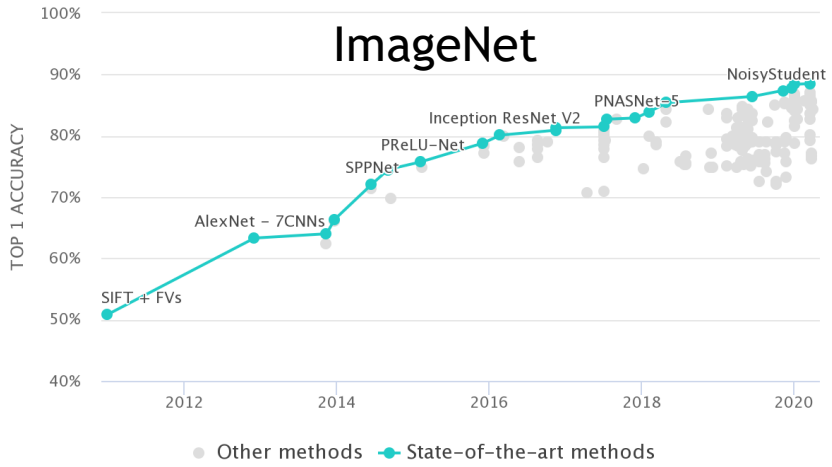
- Chris McKinstry (1997)
- Baza pytań (Mindpixel) z odpowiedziami tak/nie
- Ponad 1,4 mln zweryfikowanych pytań
- Przykłady:
  - Czy Ziemia jest planetą?
  - Czy Abraham Lincoln był prezydentem USA?
  - Czy Słońce jest większe od psa?
  - Czy ludzie czasami kłamią?
  - Czy Honduras leży w Azji?
  - Czy drewno jest twardsze od diamentu?
- NELL: Never-Ending Language Learning (50 mln)

  playsInstrument(George_Harrison, guitar)

Computer vision
klasyfikacja obrazów

- MNIST (99,84%)
- CIFAR-10 (99,3%), CIFAR-100 (93,6%)
- ImageNet (88,5%)



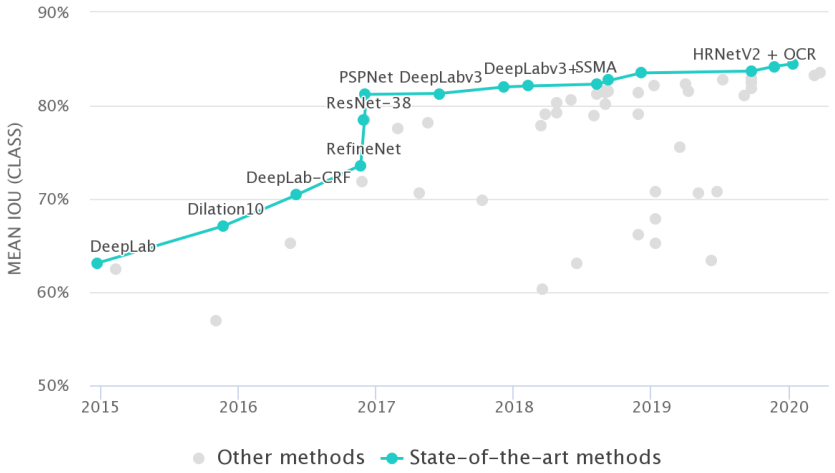


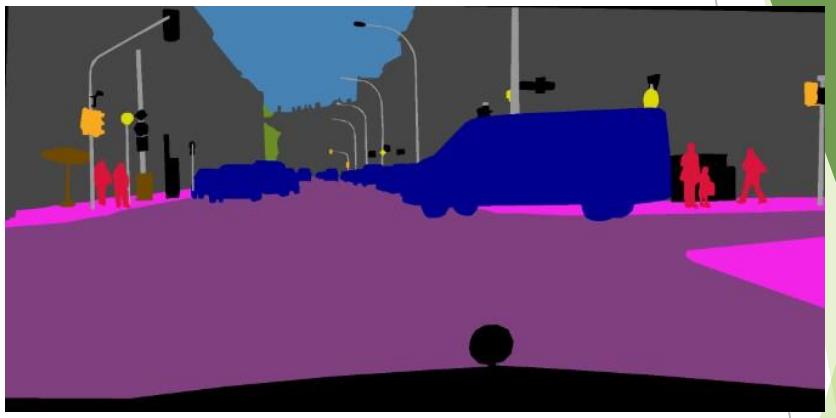
- powiązane zadania: few-shot image classification, semi-supervised image classification, fine-grained image classification

Computer vision

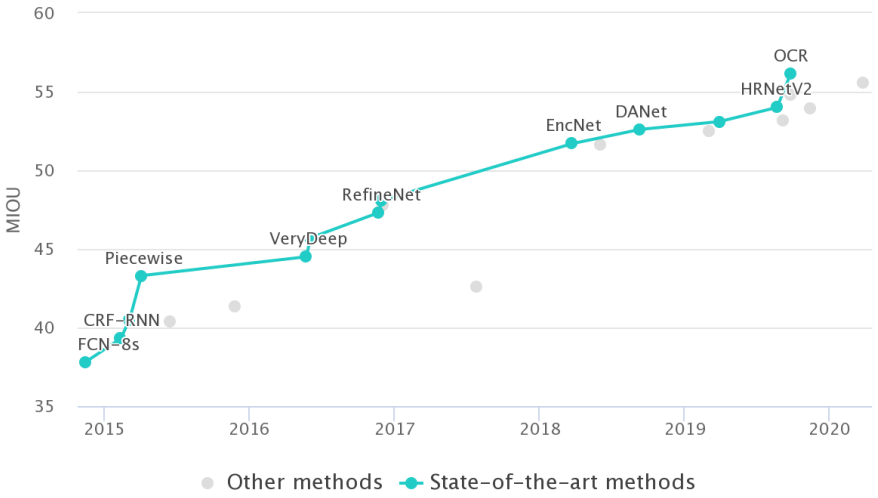
segmentacja

Cityscapes (84,5%)



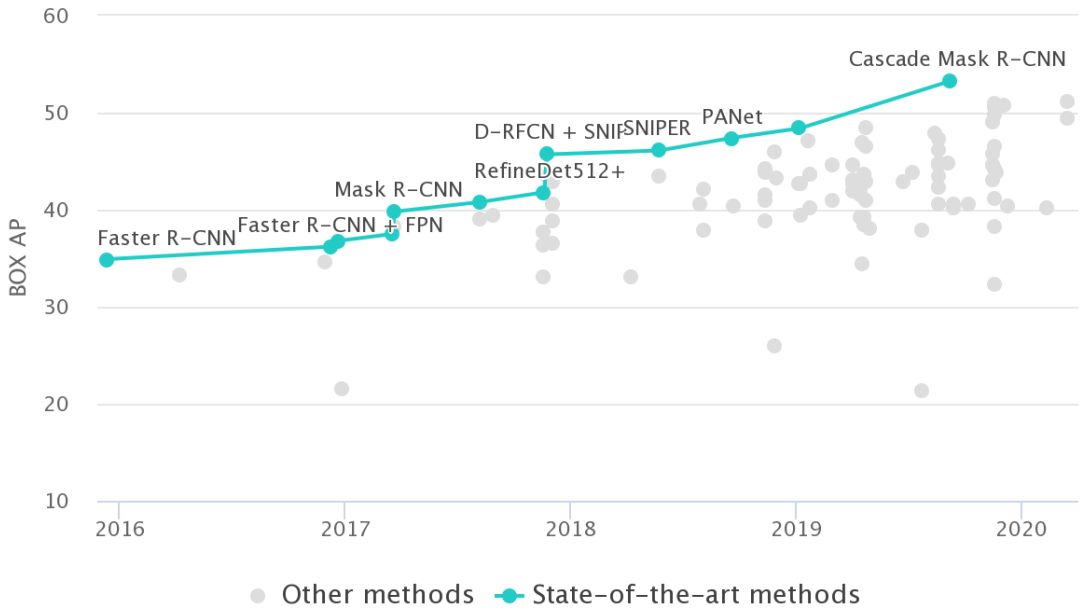


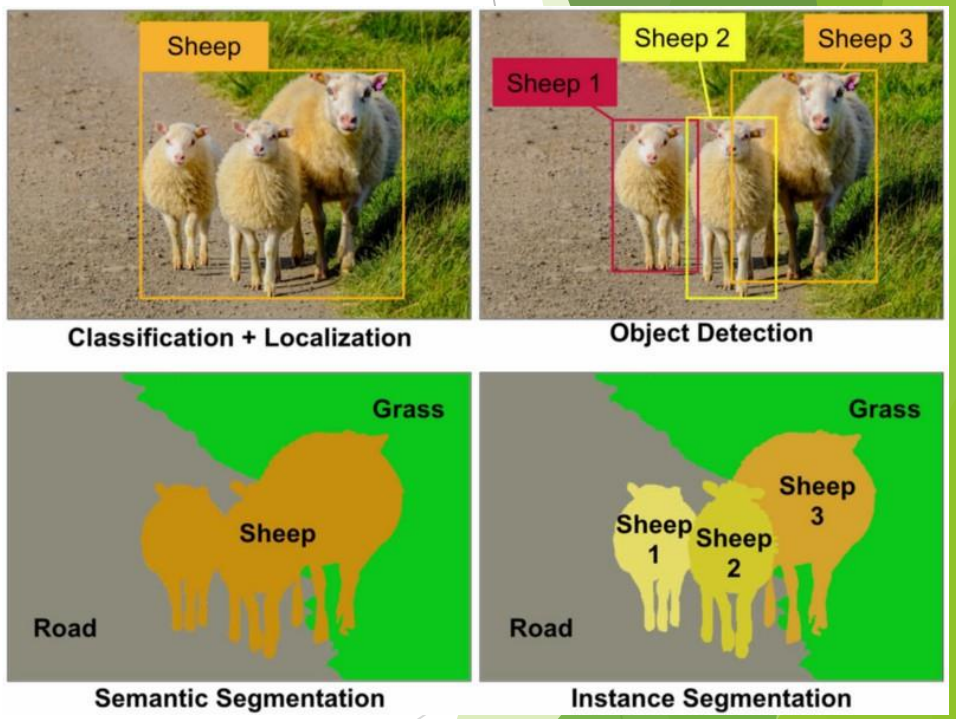
Pascal Context (56,2%)



Computer vision
detekcja i segmentacja obiektów

COCO (53,3%)





Visual Turing Test
(Geman 2015)

- Is there an instance of an object of type t with attributes A partially visible in region r?
- Does object o have attribute a?
- Does object o_1 have relationship r with object o_2?

1. Q: Is there a person in the blue region?	A: yes
2. Q: Is there a unique person in the blue region? (Label this person 1)	A: yes
3. Q: Is person 1 carrying something?	A: yes
4. Q: Is person 1 female?	A: yes
5. Q: Is person 1 walking on a sidewalk?	A: yes
6. Q: Is person 1 interacting with any other object?	A: no
⋮	
9. Q: Is there a unique vehicle in the yellow region? (Label this vehicle 1)	A: yes
10. Q: Is vehicle 1 light-colored?	A: yes
11. Q: Is vehicle 1 moving?	A: no
12. Q: Is vehicle 1 parked and a car?	A: yes
⋮	
14. Q: Does vehicle 1 have exactly one visible tire?	A: no
15. Q: Is vehicle 1 interacting with any other object?	A: no
17. Q: Is there a unique person in the red region?	A: no
18. Q: Is there a unique person that is female in the red region?	A: no
19. Q: Is there a person that is standing still in the red region?	A: yes
20. Q: Is there a unique person standing still in the red region? (Label this person 2)	A: yes
⋮	
23. Q: Is person 2 interacting with any other object?	A: yes
24. Q: Is person 1 taller than person 2?	A: amb.
25. Q: Is person 1 closer (to the camera) than person 2?	A: no
26. Q: Is there a person in the red region?	A: yes
27. Q: Is there a unique person in the red region? (Label this person 3)	A: yes
⋮	
36. Q: Is there an interaction between person 2 and person 3?	A: yes
37. Q: Are person 2 and person 3 talking?	A: yes

Postrzeganie relacyjne
(Santoro 2017)

- badanie relacji między obiektami

Image

Non-relational question

Q: What is the shape of the gray object?
A: circle

Relational question

Q: What is the shape of the object
that is furthest from the gray object?
A: square

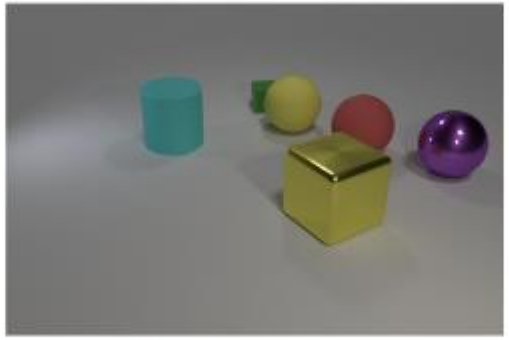
Non-relational question

Q: Is the green object on the left or on the right?
A: right

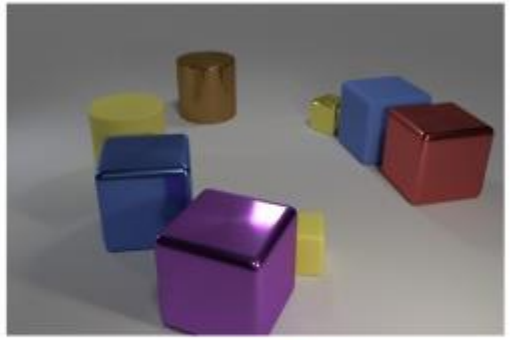
Relational question

Q: How many objects have the shape of the orange object?
A: 3

What shape is the small object that is in front of the yellow matte thing?



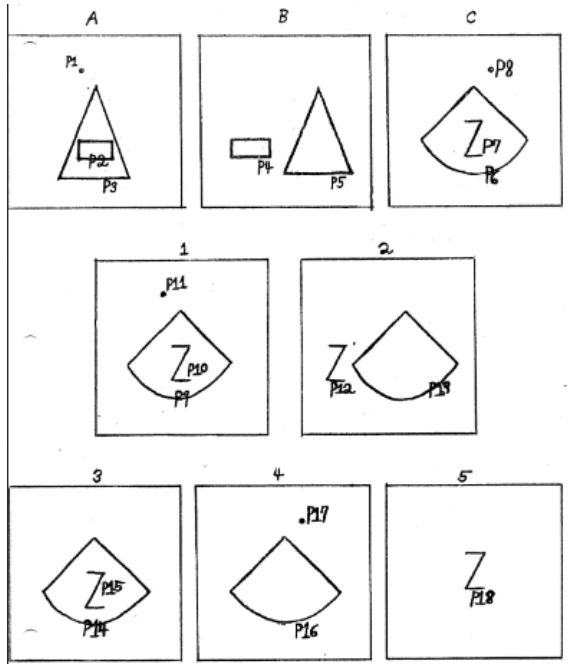
What number of things are either tiny green rubber objects or shiny things that are behind the big metal block?



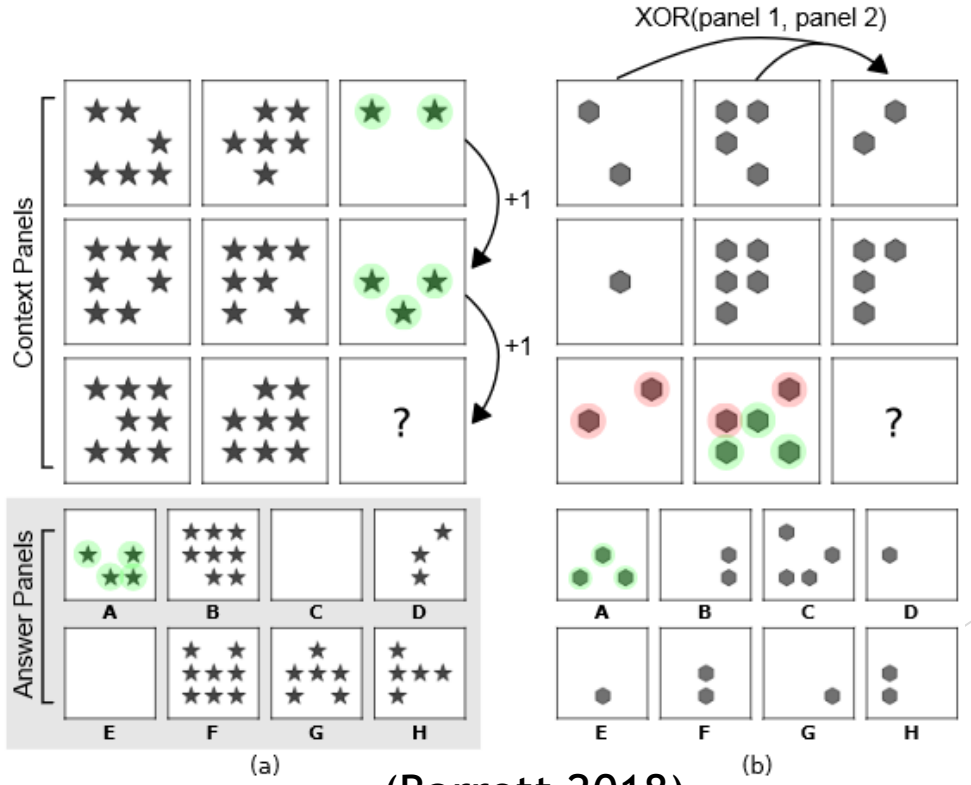
What number of objects are blocks that are in front of the large red cube or green balls?

Myślenie abstrakcyjne – testy IQ

- stworzone dla ludzi, niekoniecznie odpowiednie dla programów/maszyn
- testy mierzące inteligencję zwierząt (zbyt trudne lub zbyt łatwe)



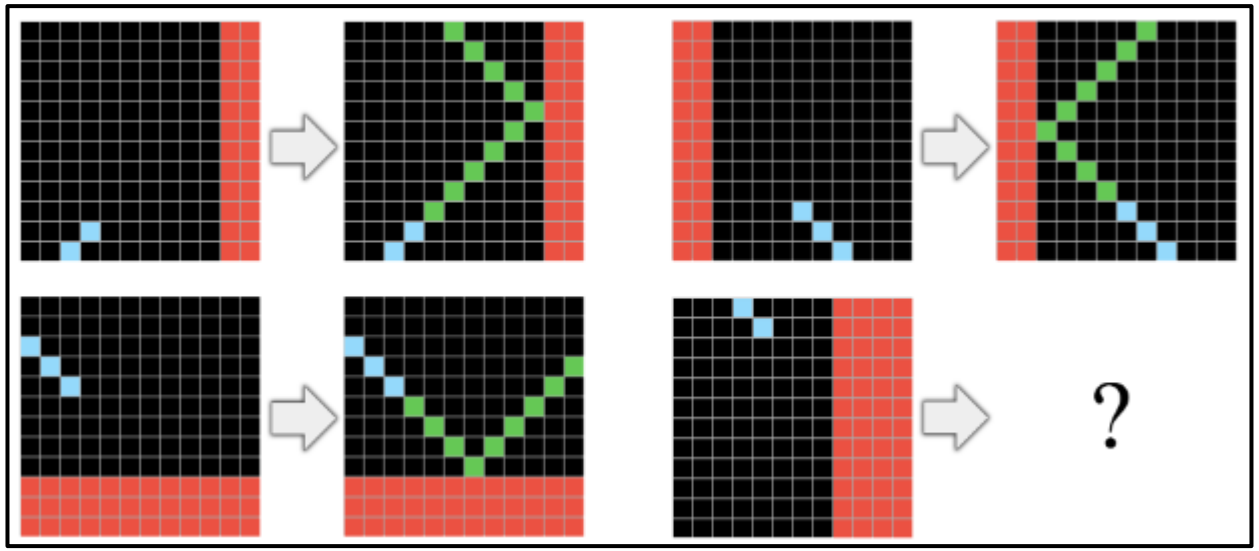
(Evans 1962)

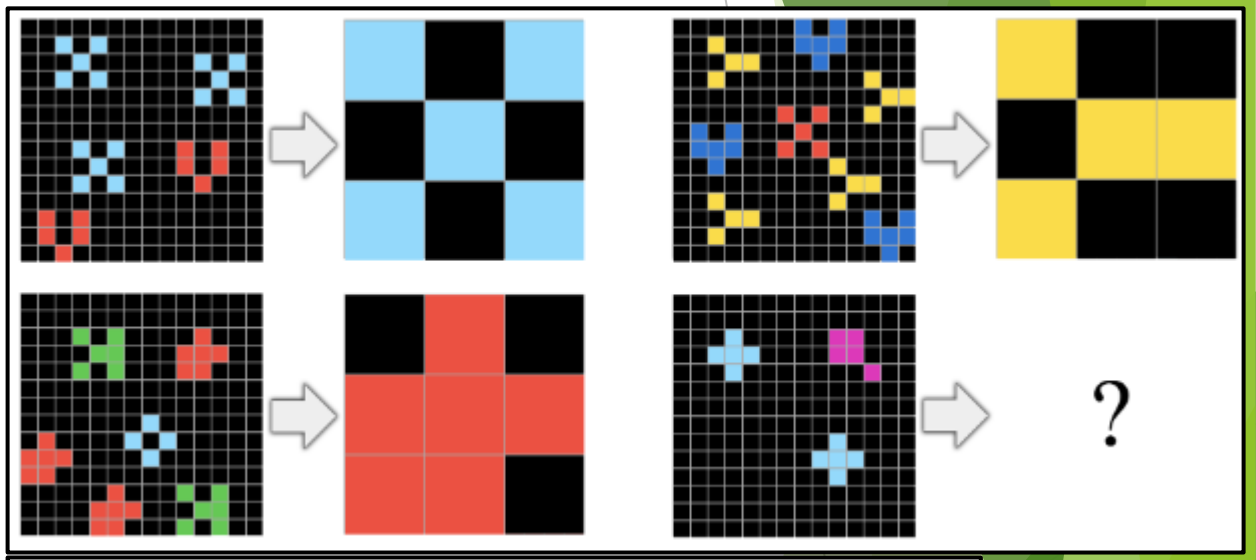


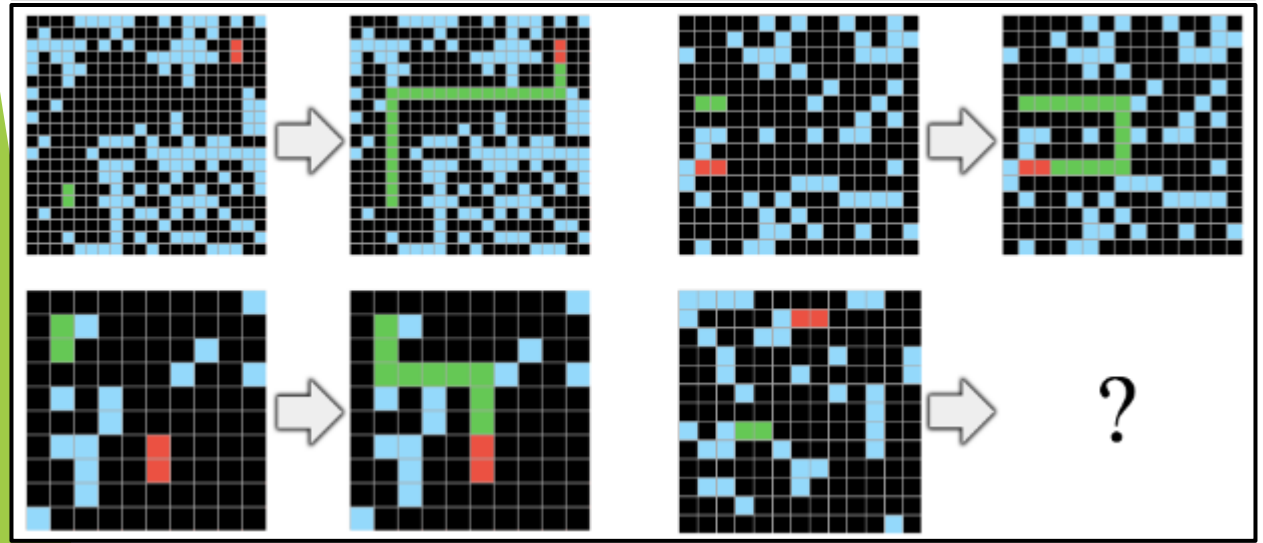
(Barrett 2018)

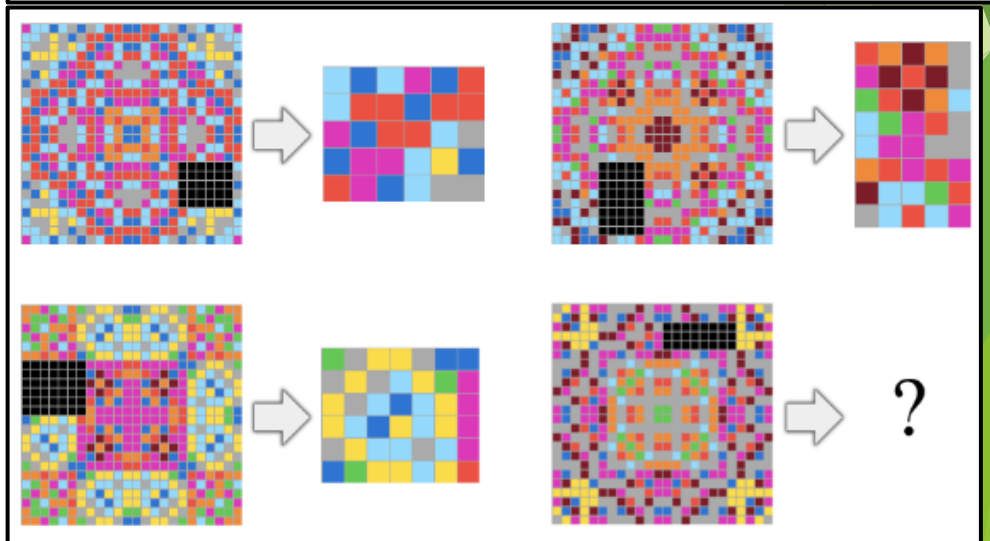
Abstraction Reasoning Corpus

(Chollet 2019)









Abstraction Reasoning Corpus

kaggle

Research Code Competition

Abstraction and Reasoning Challenge

Create an AI capable of solving reasoning tasks it has never seen before

$20,000

Prize Money

Abstraction and Reasoning Corpus · 551 teams · 2 months to go (2 months to go until merger deadline)

Overview | Data | Notebooks | Discussion | Leaderboard | Rules | Join Competition

#	Team Name	Notebook	Team Members	Score	Entries	Last
1	icecuber			0.901	15	21d
2	Alvor			0.901	106	18h
3	Zoltan			0.911	138	25m
4	nagiss			0.921	192	20h
5	alijs			0.931	91	1h
6	Anton Chikin			0.941	96	1h

Inne obszary

- Rozpoznawanie mowy
- Systemy rekomendacji
- Gry
- Szeregi czasowe
- Audio (generowanie muzyki, rozpoznawanie dźwięków)
- Robotyka
- Reprezentacja wiedzy
- Podejmowanie decyzji

i inne…

Kreatywność

- chińska poezja
- malowanie obrazów w wybranym stylu
- komponowanie muzyki
- stworzenie trailera filmu
- napisanie wiadomości (newsa)
- stworzenie przepisu kulinarnego
- stworzenie układu choreograficznego

Universal Intelligence
(Legg 2007)

- inspiracja „dynamicznymi” testami na inteligencję
- nauczyciel -> uczeń
- nieformalnie: inteligencja to zdolność agenta do realizowania celów w różnych środowiskach
- bardziej formalnie: dla każdego środowiska (zadania) poziom realizacji celu (skuteczność) jest proporcjonalna do poziomu skomplikowana tego środowiska oraz czasu w nim spędzonym (liczba przykładów uczących, bodźców)

- Measuring universal intelligence: Towards an anytime intelligence test *(Hernández-Orallo 2010)*

Inne pomysły

- Ebert test – opowiedzenie przez maszynę żartu
- Winograd Schema Challenge – określenie podmiotu w zdaniu

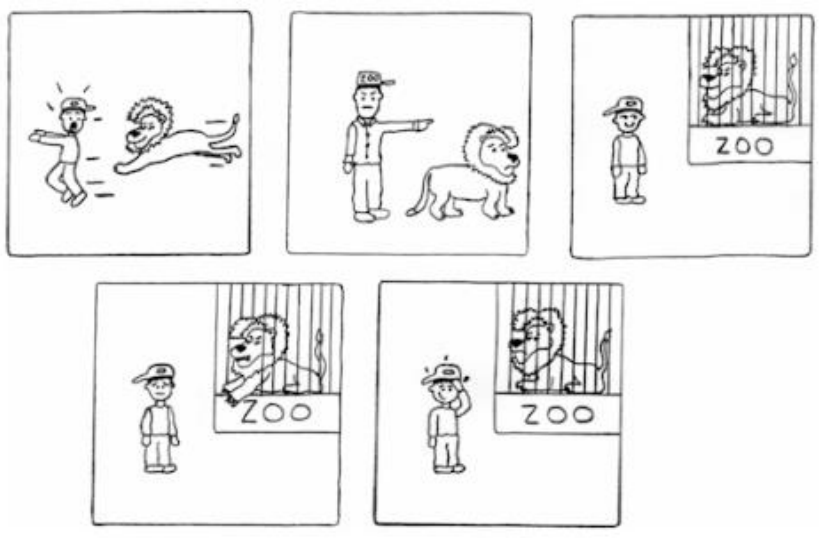
The trophy doesn't fit in the brown suitcase because it's too ⟨⟩. What is too ⟨⟩?

Answer 0:	the trophy
Answer 1:	the suitcase
special:	big
alternate:	small

The large ball crashed right through the table because it was made of ⟨⟩. What was made of ⟨⟩?

Answer 0:	the ball
Answer 1:	the table
special:	steel
alternate:	styrofoam

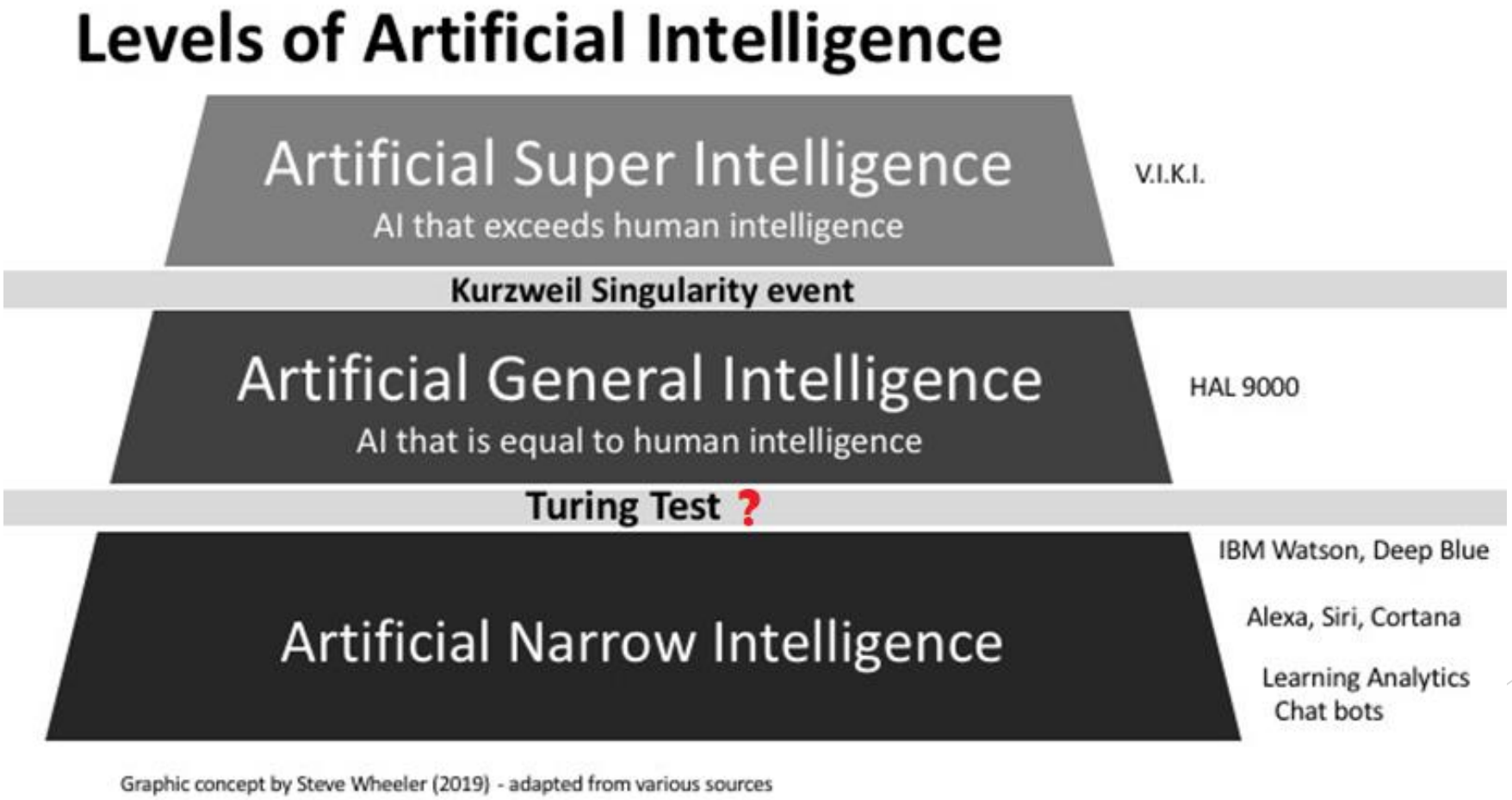
- pomiar inteligencji liczbą nowych odkryć naukowych
- Test Turing w wersji „zwierzęcej" (np. maszyna naśladująca psa)
- ułożenie logicznej historii z obrazków
- łączenie obszarów, np. stworzenie ilustracji do książki, napisanie streszczenia filmu
- Torrance Test – seria testów badających kreatywność



Psychometric AI
(Bringsjord 2003)

- komputery dobrze radzą sobie z pojedynczymi jasno sprecyzowanymi zadaniami
- ludzkie testy na inteligencję nie są odpowiednie dla maszyn
- potrzeba jest stworzenia testów, które mierzyłyby sztuczną inteligencję

Silna vs słaba sztuczna inteligencja



Bibliografia

[1] Turing, Alan, "Computing Machinery and Intelligence", Mind, LIX (1950): 433–460

[2] Feigenbaum, Edward A. "Some challenges and grand challenges for computational intelligence." *Journal of the ACM (JACM)* 50.1 (2003): 32-40.

[3] Geman, Donald, et al. "Visual turing test for computer vision systems." *Proceedings of the National Academy of Sciences* 112.12 (2015): 3618-3623.

[4] Legg, Shane, and Marcus Hutter. "Universal intelligence: A definition of machine intelligence." *Minds and machines* 17.4 (2007): 391-444.

[5] McCoy, John P., and Tomer D. Ullman. "A Minimal Turing Test." *Journal of Experimental Social Psychology* 79 (2018): 1-8.

[6] Schweizer, Paul. "The truly total Turing test." *Minds and Machines* 8.2 (1998): 263-272.

[7] Chollet, François. "The Measure of Intelligence." *arXiv preprint arXiv:1911.01547* (2019).

Bibliografia

[8] Hernández-Orallo, José, and David L. Dowe. "Measuring universal intelligence: Towards an anytime intelligence test." *Artificial Intelligence* 174.18 (2010): 1508-1539.

[9] Santoro, Adam, et al. "Measuring abstract reasoning in neural networks." *International Conference on Machine Learning*. 2018.

[10] Adams, Sam S., Guruduth Banavar, and Murray Campbell. "I-athlon: Towards a multidimensional Turing test." *AI Magazine* 37.1 (2016): 78-84.

[11] Bringsjord, Selmer, and Bettina Schimanski. "What is artificial intelligence? Psychometric AI as an answer." *IJCAI*. 2003.

[12] Santoro, Adam, et al. "A simple neural network module for relational reasoning." *Advances in neural information processing systems*. 2017.